

# Good Clinical Data Management Practices



October 2013 Edition



Recipient, **2007 Clinical Research Excellence Award**  
"Most Successful Company or Programme of the Year in  
Raising GCP Standards" category

Society for Clinical Data Management





**GOOD**  
Clinical Data Management  
**PRACTICES**

# Good Clinical Data Management Practices

---

Published October 2013

“The need for Good Clinical Data Management Practices is not new. In the early 1970s, the Public Health Service recognized this need through a contract to a major research university for training of research data managers. However, the need continues, the need changes over time, and the need for good clinical data management practices has become even more important as biopharmaceutical and medical device industry and regulatory bodies rely more and more heavily on the evaluation of electronically transmitted clinical trials data for critical data-based decision making.”

Thus, the Society for Clinical Data Management provides the *Good Clinical Data Management Practices* to the SCDM membership.

This document constitutes neither consensus nor endorsement by regulatory agencies, pharmaceutical or biotech companies, contract research organizations or the academic community, but rather reflects the current views of SCDM membership. Additionally, none of the recommendations contained herein supersede regulations or regulatory guidelines, which should always be consulted prospectively to assure compliance. The document should not be considered an exhaustive list of topics.



**This page is intentionally blank.**

## GCDMP Revision History

Publication Date	Comments
September 2000	Initial publication of the GCDMP with the following chapters: Assuring Data Quality; Data Acquisition; Data Entry and Data Processing; Data Storage; Database Closure; Database Validation, Programming and Standards; Laboratory and Other External Data; Measuring Data Quality; Safety Data Management and Reporting; Vendor Management; Glossary.
January 2002	The following chapters added to the GCDMP: CDM Presentation at Investigator Meetings; CRF Printing and Vendor Selection; Preparation and Preservation of CRF Completion Guidelines; Serious Adverse Event Data Reconciliation; Training.  Data Entry and Data Processing chapter revised.
September 2003	The following chapters added to the GCDMP: Clinical Data Archiving; Data Privacy; Dictionary Management; Electronic Data Capture Principles.
October 2005	Metrics chapter revised.
May 2007	All chapters revised for consistency of style, grammar, and clarity. Substance of chapter content unchanged.
July 2008	All chapters revised with new headers, footers and pagination. The following chapters were revised for content, style, grammar and clarity: Serious Adverse Event Data Reconciliation; CRF Completion Guidelines; Clinical Data Archiving, CDM Presentation at Investigator Meetings, Vendor Management..
September 2008	The following chapters added to the GCDMP: Electronic Data Capture—Concepts and Study Start-up; Electronic Data Capture—Study Conduct; Electronic Data Capture—Study Closeout.  Measuring Data Quality chapter revised for content, style, grammar and clarity.
December 2008	The following chapter added to the GCDMP: Data Management Plan.
March 2009	Database Validation, Programming and Standards chapter revised for content, style, grammar and clarity.
April 2009	Data Privacy chapter revised for content, style, grammar and clarity.
May 2009	Dictionary Management chapter revised for content, style, grammar and clarity and renamed Medical Coding Dictionary Management and Maintenance.
July 2009	The following chapters added to the GCDMP: Patient-Reported Outcomes; Data Management Standards in Clinical Research.
October 2009	The following chapter added to the GCDMP: Laboratory Data Handling.  Data Entry and Data Processing chapter revised for content, style, grammar and clarity and renamed Data Entry Processes  Laboratory and Other External Data chapter renamed External Data Transfers
December 2009	The following chapter added to the GCDMP: Edit Check Design Principles.
March 2010	Vendor Management chapter revised for content, style, grammar and clarity and renamed Vendor Selection and Management.

June 2010	The following chapter added to the GCDMP: Project Management for the Clinical Data Manager.
October 2010	Data Acquisition chapter revised for content, style, grammar and clarity and renamed Design and Development of Data Collection Instruments.
April 2011	Metrics for Clinical Trials revised for content, style, grammar and clarity and renamed Metrics in Clinical Data Management.
October 2013	Assuring Data Quality revised for content, style, grammar, and clarity. Added more explicit description of quality management system components important in clinical research data management  Database Closure revised for content, style, grammar, and clarity with database closure sample flowchart and sample checklist added.  Glossary revised with the addition of approximately seventy-five (75) terms.

## Contents

Executive Summary .....	2 pages
Acknowledgements.....	2 pages
Introduction .....	2 pages
Data Privacy.....	Revised April 2009 .....14 pages
Data Management Plan.....	Added Dec 2008.....16 pages
Project Management for the Clinical Data Manager .....	Added June 2010.....24 pages
Vendor Selection and Management .....	Revised March 2010.....24 pages
Data Management Standards in Clinical Research.....	Added July 2009.....20 pages
Design and Development of Data Collection Instruments .....	Revised Oct 2010 .....18 pages
Edit Check Design Principles .....	Added Dec 2009.....18 pages
Electronic Data Capture—Concepts and Study Start-up.....	Added Sep 2008.....54 pages
Electronic Data Capture—Study Conduct.....	Added Sep 2008.....24 pages
Electronic Data Capture—Study Closeout .....	Added Sep 2008.....14 pages
CRF Completion Guidelines.....	Revised June 2008.....8 pages
CRF Printing and Vendor Selection .....	Revised May 2007 .....8 pages
Database Validation, Programming, and Standards.....	Revised March 2009.....20 pages
Laboratory Data Handling .....	Added Oct 2009 .....22 pages
External Data Transfers .....	Revised May 2007 .....14 pages
Patient-Reported Outcomes .....	Added July 2009.....14 pages
CDM Presentation at Investigator Meetings.....	Revised July 2008.....6 pages
Training.....	Revised May 2007 .....14 pages
Metrics in Clinical Data Management.....	Revised April 2011 .....20 pages
Assuring Data Quality.....	Revised Oct 2013 .....20 pages
Measuring Data Quality .....	Revised Sep 2008.....12 pages
Data Storage.....	Revised May 2007 .....6 pages
Data Entry Processes.....	Revised Oct 2009 .....20 pages
Medical Coding Dictionary Management and Maintenance.....	Revised May 2009 .....16 pages
Safety Data Management and Reporting .....	Revised May 2007 .....22 pages
Serious Adverse Event Data Reconciliation.....	Revised Jan 2008.....8 pages
Database Closure.....	Revised Oct 2013.....12 pages
Clinical Data Archiving .....	Revised June 2008.....10 pages
Glossary.....	Revised October 2013 .....32 pages

**This page is intentionally blank.**



---

## Executive Summary

The Society for Clinical Data Management (SCDM) is a non-profit professional organization founded to advance the discipline of clinical data management (CDM). The SCDM is organized exclusively for educational and scientific purposes. The mission of the SCDM, promoting clinical data management excellence, includes promotion of standards of good practice within clinical data management. In alignment with this part of the mission, the SCDM Board of Trustees established a committee to determine standards for Good Clinical Data Management Practices (GCDMP) in 1998. The committee charter reads as follows:

The review and approval of new pharmaceuticals by federal regulatory agencies is contingent upon a trust that the clinical trials data presented are of sufficient integrity to ensure confidence in the results and conclusions presented by the sponsor company. Important to obtaining that trust is adherence to quality standards and practices. To this same goal, companies must assure that all staff involved in the clinical development program are trained and qualified to perform those tasks for which they are responsible.

The discipline of Clinical Data Management includes paper and electronic case report form (CRF) design, clinical trials database design and programming, data standards, system implementation, data acquisition, data integration, into the clinical trials database, data review, validation, coding and database finalization. Independent of how individual companies perform these tasks within their company each company is obligated to ensure that the individuals performing these tasks follow Good Clinical Practices. However, currently prior to SCDM and this committee, there were no published good clinical practice guidelines specific to the discipline of Clinical Data Management. As the organization representing Clinical Data Management professionals in North America, SCDM is in a position to develop, maintain and publish GCDMP guidelines

that define and promote current industry procedures and best practices.

One of the objectives of the committee is to develop, publish, and recommend use of guidelines for Good Clinical Data Management Practices. In addition to this stated objective of the GCDMP committee, it has been our continuing goal to obtain as much input and participation as possible from the SCDM members and other users to further develop GCDMP guidelines.

Over three years have passed since the September 2003 edition of the GCDMP was completed. During that time, the GCDMP Committee focused on the stability and future of the GCDMP and established a lifetime maintenance plan (LMP) to document the processes that guide changes. In an effort to keep the GCDMP current in a changing industry, this plan defines a formal process and timeline for review by the committee; the SCDM Board of Trustees; the international community, which is currently represented by the International Network of Clinical Data Management Associations (INCDMA); and the users. Four working subcommittees are defined in the LMP to assist in the maintenance of the GCDMP and the LMP itself.

In addition to planning for, writing, and putting in place the LMP, the GCDMP committee finalized a new chapter (“Metrics for Clinical Trials”) and revised five chapters. These updated chapters will be released when the review process has been completed.

The GCDMP is provided as a special service to the SCDM membership. The primary recipients include professionals involved in the pharmaceutical, biotechnology, and medical device clinical data management. It will provide assistance to data managers in their implementation of high quality data management processes and in their quest to become Certified Clinical Data Managers (CCDM). It will also provide management with a guide for planning training and education for new clinical data management staff.

---

## Acknowledgements

As the committee chairperson, I would like to acknowledge the expertise, dedication and hard work of the document authors. The following individuals have contributed to one or more versions of the GCDMP: Susan Bornstein, Letitia Bowen, Sally Cassells, Anthony J. Costello, Wendy Cuthbert, Bernadette Farrell, Kaye Fendt, Lisa Freeman, Volker Freiman, Imogene Grimes, Marysasser Hedrick Holloway, Susan Howard, Becky Kush, Angel Lazarov, Terrence Loding, Meredith Nahm, Armelde Pitre, Don Rosen, Barbara Tardiff, Lisa Taylor, and Beth Wilson. In addition, Sasha Zucker provided his knowledge and skills as technical editor, for which we are most grateful. While I spearheaded the effort to update the Lifetime Maintenance Plan, Susan Howard led the Review and Update subcommittee, which dedicated its efforts to reviewing existing chapters and incorporating feedback from users. I would also like to acknowledge the GCDMP Full Committee, which has provided insight and expertise during the review of the new and revised chapters. Kaye Fendt—who initially took the idea of this committee to the Board of Trustees and to interested members of SCDM and who served as Board and FDA Liaison in its early years—has continued to lend her expertise to this committee as an innovator, an author, an editor, a supporter, and a motivator. Susan Bornstein led the committee during its formation and coordinated the creation of the CDM Task List, which served as the basis for the organization of this document. Meredith Nahm chaired the committee through 2001, served as Board Liaison through 2004, and has continued to contribute to the review process. Anthony Costello, who is currently Chair of the Board of Trustees and served as Board Liaison through 2006, continues to bring driven energy and focus on exposure and training of the document to the committee.

Special acknowledgements are extended to the users who offered helpful comments and feedback, the SCDM Board of Trustees, and the INCDMA members who participated in the review process. Without their continued interest and support, the GCDMP would not exist or be current.

Administrative help (which includes providing the technical expertise needed

to post the document and the Lifetime Maintenance Plan) was provided by SCDM's management organization, including Kim Breitbach and Monica Drake.

We are most grateful to all of you for your contributions and dedication.

Carol Garvey, GCDMP Committee Chair

Linda Talley, Board of Trustees Liaison

---

## Introduction

The purpose of this document is to provide guidance on accepted practices for the many areas of CDM that are not covered by existing regulations and guidance documents. The intent is to remain consistent with regulatory practices in related areas of clinical research and to apply the concepts contained in those regulations and associated guidance documents to CDM. It is also the intent of the GCDMP to provide practical suggestions and proven means of meeting the guidelines recommended in the GCDMP. The GCDMP is written to serve the needs of multiple audiences including: data managers, data processors, statisticians, site personnel, clinical professionals, compliance auditors, regulatory affairs personnel, and all clinical research professionals making decisions regarding or using clinical trial data.

The GCDMP addresses the CDM areas of responsibility in 20 chapters. Each chapter includes two sections titled Minimum Standards and Best Practices respectively. These sections summarize the main recommendations of the chapter in bulleted form. For an executive summary or an overview of a chapter, read the chapter's abstract, Minimum Standards, and Best Practices. The Minimum Standards ensure that data are complete, reliable, and processed correctly, otherwise known as data integrity. The Best Practices offer higher efficiency, quality, and function and lower risk in addition to assuring data integrity. The body of each chapter provides the rationale, technical details, and, often, discussion of alternate or common practices. References are provided at the end of each chapter to guide the reader to additional resources. Each chapter also contains recommended standard operating procedures (SOPs). Whether the SOPs are departmental or institutional in nature, it is the data manager's responsibility to ensure that all data management concerns are addressed.

In the absence of CDM regulatory standards, it is important for experienced, professional data managers to provide thought leadership on accepted data quality levels, on practical methods of achieving them, and on the implications of new technology on the CDM tasks. Data management tasks

are often technical and specialized. As the industry utilizes new technologies, it is therefore crucial that data management professionals take an active and forward-thinking role in setting appropriate expectations and standards for data quality, methodology for quantifying data quality, and auditing practices to ensure data quality.

The presence of acceptable quality standards becomes even more important as the industry undertakes larger trials where manual processes are no longer effective. New technologies often require not only retooling the data management process but also reforming the data management process to take advantage of the efficiencies offered by new technologies.

---

# Data Privacy

April 2009

## *Abstract*

The privacy of any subject who participates in a clinical study must be protected for ethical and legal reasons. Clinical data management professionals must be familiar with privacy laws that exist for the regions in which clinical studies are occurring and ensure all reasonable and appropriate precautions are taken. This chapter discusses strategies and considerations that data managers must understand and follow, including the varying types of personal data in clinical studies, best practices for securing and protecting data (both paper and electronic), methods of data collection, and strategies for ensuring that personnel, both internal and external (e.g., vendors), follow applicable data privacy standards.

## *Introduction*

Data privacy refers to the standards surrounding protection of personal data. Personal data can be defined as any information that can lead to identification, either directly or indirectly, of a research subject. Some examples of personal data are subject names, initials, addresses, and genetic information.

The *ICH Guideline for Good Clinical Practice (GCP)* states “The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with applicable regulatory requirement(s).”<sup>1</sup>

Privacy protection afforded to research subjects includes:

- Protocol review and approval by an institutional review board (IRB)
- Right to informed consent
- Right of the subject to withdraw consent and have no further data collected

- Right to notice of disclosure
- Confidential collection and submission of data

Although the majority of data privacy responsibilities rest with site management or clinical monitoring, data management professionals should be familiar with basic data privacy issues and follow regulatory and organizational guidelines to ensure the privacy of research subjects.

Having complete anonymity may not always be practical for the design of a study, however, personal information should always be safeguarded to the greatest extent possible.

## **Scope**

This chapter focuses on considerations needed to maintain a high degree of privacy protection (or security) for research subjects during data collection and management. Since significant regulatory guidance exists on data privacy, all applicable regulations should be considered in the creation of company policy or standard operating procedures (SOPs) to ensure full compliance with regulations governing the jurisdictions in which business is conducted. References for various regulatory documents can be found in the Further Reading section of this chapter.

Many of the tasks described in this chapter may be joint responsibilities between different groups, just as there may be many different groups involved in the implementation of various tasks. However, clinical data managers need to be conscious of whether or not these tasks have in fact been performed in a satisfactory manner.

## **Minimum Standards**

- Ensure all personnel (including vendors) who directly or indirectly handle identifiable personal data are properly trained on data privacy issues. Training sessions should cover data privacy concepts; company policy; regulatory agency policy and applicable local, state, federal, and international laws.
- Design data-collection instruments with the minimum subject identifiers needed, including the design of case report forms (CRFs), clinical and



laboratory databases, data transfer specifications, and any other area of data collection that may contain personal information.

- Ensure personal data is not identifiable, other than subject identifiers used to link documentation to a database record, from documentation (e.g., CRFs, lab reports, images associated with the clinical study) submitted to data management.
- Review and update data management processes regularly to ensure consistency with current company privacy policies and government regulations.

### ***Best Practices***

- Develop and maintain an environment that respects the privacy of research subjects. Consider employee education programs that highlight the potential impact of lapses in data privacy, the benefits of applying strict criteria when handling personal information, and verification that procedures are in compliance with regulations.
- Implement procedures prior to data transfer between sites, departments, subsidiaries, and countries to ensure all privacy considerations have been considered, addressed, and documented.
- Promote internal and external accountability through company policies and regulations governing the use of personal information.
- Implement procedures for using data for an alternate or new purpose other than what was originally intended by the informed consent. Ensure all privacy considerations have been considered, addressed, and documented.
- Enforce a baseline policy of denying access to personal data. Evaluate any request for this information. If information is determined to be required for specific scientific reasons, ensure all privacy considerations have been considered, addressed, and documented.
- Put stringent procedures in place to securely transfer, store, access, and report on extremely sensitive data (e.g., genetic information).

- Work with those responsible for quality assurance to ensure compliance with data privacy regulations. This assurance of regulatory compliance should be a central focus of audits and a contract contingency when using external service providers.
- Maintain proper physical and electronic security measures. Data should be stored in protective environments relevant to the type of media being stored. Paper CRFs should be stored in an environment with regulated access. Proper precautions should be taken to prevent external access to electronic data, such as password authentication and firewall security.

### ***Importance of Data Privacy***

Revealing a subject's personal medical information could potentially lead to embarrassment, denial of insurance coverage, or discrimination in the workplace. For these and other reasons, most countries have passed stringent laws that mandate the protection of research subjects' privacy.

Every organization with access to subjects' personal data should have SOPs addressing data privacy. At a minimum these SOPs should comply with all regulations of the study locale, although many organizations put SOPs in place that are stricter than required by local regulations.

All personnel with access to personal data must be adequately educated in data privacy related SOPs. The reasons for data privacy, what constitutes personal data, and how to handle various situations that may arise in the course of the study should be explained.

The data manager's role has a narrower focus than an investigator site in regards to data privacy. Nonetheless, the data manager needs to ensure data privacy is maintained throughout all aspects of data management.

### **Legislation and Regulatory Guidance**

Legislation and guidance documents from the EU and US have a greater impact on clinical research than laws in other countries, because the EU and US are involved with a higher volume of clinical research. In Europe, EU Data Protection Directive 95/46/EC, which became mandatory in October 1998, covers privacy of all types of personal data including data from clinical studies.<sup>2</sup> Directive 2001/20/EC subsequently became mandatory in May 2004

and expanded upon the previous directive in relation to data privacy and informed consent in clinical studies.<sup>3</sup> One of the stipulations of these directives is that members of the EU are not allowed to transfer personal data to countries that the EU Commission has determined lack adequate subject privacy standards. Countries that are found to have adequate privacy standards are given an “adequacy determination” by the EU Commission. In regards to the US, the EU has agreed to give *individual* US companies an adequacy determination if they meet the privacy standards of the EU.<sup>4</sup> As a result, many US companies have adopted the stricter privacy requirements of the EU.

The processes for US companies to acquire an adequacy determination are known as Safe Harbor Principles, and were developed by the US Department of Commerce in collaboration with the EU. Once a company receives an adequacy determination through adherence to these principles, they must recertify every 12 months. According to these principles, companies must provide the following:

- Notice—Subjects must be informed of how their data will be collected and used.
- Choice—Subjects must be able to opt out of collection of their data and its transfer to third parties.
- Data transfers—Any transfers of data to third parties must only be to other organizations that have rigorous data-protection policies.
- Security—All reasonable efforts must be made to prevent the loss of any data collected.
- Data integrity—Data must be reliable and relevant to the purpose for which it was collected.
- Access—Subjects must be able to access information about them that is collected, and have an opportunity to have this data corrected or deleted if necessary.
- Enforcement—A mechanism must be in place to effectively and consistently enforce these rules.

It is recognized that laws dealing with medical data privacy in the US are more fragmented than those of the EU. One example of this fragmentation is the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, which went into effect in April 2003.<sup>5</sup> Although HIPAA covers a wide range of organizations possessing health data, research recruitment organizations, clinical research organizations and pharmaceutical companies fall outside HIPAA's purview.<sup>4</sup> Other US privacy laws include Section 5 of the Federal Trade Commission Act (15 United States Code § 45(a)(1)), the Gramm-Leach Bliley Act (15 United States Code, Subchapter 1, § 6801–6809), several parts of Code of Federal Regulations Titles 21 and 45, and numerous state laws regarding data privacy. *ICH Guideline for Good Clinical Practice* and various FDA guidance documents give additional advice and directives for privacy issues in clinical studies, but are not legally binding documents.

## **What Constitutes Private or Personal Information?**

According to EU Directive 95/46/EC, personal data “shall mean any information relating to an identified or identifiable natural person (‘data subject’); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.”<sup>2</sup>

Similarly, 45 CFR Section 164.501 (HIPAA) defines individually identifiable health information as “...information that is a subset of health information, including demographic information collected from an individual and:

(1) Is created or received by a health care provider, health plan, employer, or health care clearinghouse; and

(2) Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual; and

(i) That identifies the individual; or

(ii) With respect to which there is a reasonable basis to believe the information can be used to identify the individual.”<sup>5</sup>

## **Data Privacy Focus Areas**

Clinical data managers should make every effort to ensure access to data is restricted to qualified and approved personnel. In particular, the following areas should be examined to ensure appropriate data privacy is maintained.

### **Vendors with Access to Data**

Different standards may need to be employed for vendors who only have access to vendor-specific data versus those who have access to the study database and all subject-associated data. For those vendors having access to the database, the data manager should ensure that the vendors subscribe to standards that meet or surpass internal standards. As an overall strategy, ensure your company is performing external audits of vendors that include investigations into their compliance with regulations concerning the protection of personal data.

### **Lab Data**

Reports generated from all types of labs should not contain any subject-specific information. This information should be built into data-transfer and reporting specifications.

If source documents are to be collected (e.g., radiology, MRI, or ECG reports), the sites should be instructed that all documentation should be stripped of personal identifiers, and appropriate subject identifiers should be assigned prior to submission to data management. If that direction is not followed, data management should follow up with the appropriate internal or external clinical site management to ensure that follow-up and further direction is recommended for specific site violators.

### **Central Committees**

Reports to and meetings with various committees may necessitate presentation of some study data. Different types of committees may require different data points and data sources, according to the committee's function. A committee may require reports based on the database, data from the database, original source data or copies of source data. In all cases, personal subject identifiers should be removed prior to presentation of data to the committee, and in some

cases, study identifiers may need to be added. The parties responsible for anonymity of the data may vary depending on the type and source of the data. Someone independent of the study may be utilized when necessary to ensure data anonymity, such as a liaison between the company and the committee.

### **Data transfers**

Prior to any data transfer, a data transfer specification document should be produced to identify the secure method of transfer and fields to be transferred, including the data keys and structure. Before any data is transferred, the transfer process should be thoroughly tested to ensure no extraneous information is transferred that could jeopardize data privacy. Once the planned data transfer is performed, the transfer should be reviewed to ensure all transferred data matches the database.

### **Computer and network security**

Computer and network security are typically developed and maintained by an organization's information technology personnel. However, data managers do have a responsibility to ensure that the systems are used appropriately and responsibly. Any lapses in computer or network security may jeopardize the integrity of the database, and therefore, data privacy.

### **Appropriate Redaction of Personal Data**

Redaction is the act of obscuring or removing text from a document before releasing the document to other personnel or departments. An example of clinical data needing to be redacted could include a situation where a comments field was completed with personal identifiers. If for example a comments field had the text "Mr. Jones showed improvements," the data manager should obscure or remove "Mr. Jones" from this text. Organizations should have SOPs to determine when redaction of personal data is needed. This should preferably be performed by the site or monitor, but if not handled at the site, data managers should be mindful of when redaction of personal data is required as well as knowledgeable on the process.

## **Data Collection**

To ensure proper assignment of data into a clinical database, data collection instruments should be designed with some type of research subject identifiers. The use of these identifiers should be taken into consideration not only in CRF design, but also in scenarios in which the processing, transfer, reporting, or analysis of data will be completed. These scenarios include the design of clinical databases, laboratory databases, and data transfer specifications. In general, a random subject number can be used to resolve any discrepancies that might arise from transcription errors.

Recent scientific advances in genetics have made it possible to capture the ultimate identifier, subject DNA. Utmost care should be taken to protect this data. Strict standards should be adopted, including storage in completely independent data servers and physical locations, independent resources to manage genomic data, and specific SOPs dedicated to the processing and use of this data.

## **Variance Between Data Collection Methods**

Different data collection methodologies may necessitate different considerations to maintain privacy of data. The following are common considerations for different collection methodologies.

- Paper-based studies—Follow organization SOPs for appropriate redaction of personal identifiers as well as appropriate study procedures for handling, transfer and storage of documents containing privacy data.
- EDC studies—Follow organization SOPs to ensure appropriate network security, including password security and automatic user logout after a determined period of time.
- ePRO—Follow organization SOPs to ensure appropriate network security, as well as training of subjects on use of devices and protection of data by use of assigned passwords and user identification or pin numbers.

## ***International Studies and Data Privacy***

International studies should adhere to the most restrictive regulations of the countries involved. However, ensuring data privacy also needs to be balanced with the need for collecting all data pertinent to the study. Some questions to ask in this regard may include:

- Is the data really needed?
- Does collection of needed data compromise privacy?
- Is collection of the data acceptable in all countries with study sites?

## ***Policy Definition and Training***

Corporate policy definition and training should be based on relevant company policy; regulatory agency policy; and applicable local, state, federal, and international law. Policy training sessions should address the implementation and maintenance of standards and potential harm to subjects that may occur when basic principles are not followed.

## ***Potential Future Concerns for Data Privacy***

Electronic health records and their potential integration with EDC systems are expected to garner more attention in the future. Although there is currently no mandate to use electronic health records, the topic has been discussed frequently not only by those involved with health care or clinical studies, but also within political circles. If health records do become exclusively electronic, new safeguards will be needed to ensure privacy of these records.

## ***Recommended Standard Operating Procedures***

- Organization Procedures for Data Privacy Protection
- Vendor Management



## References

1. International Conference on Harmonisation. *Harmonised Tripartite Guideline for Good Clinical Practice*. 2<sup>nd</sup> ed. London: Brookwood Medical Publications; 1996.
2. European Parliament and Council of Europe. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Strasbourg, France: European Parliament and Council of Europe; 1995. Available at: [http://ec.europa.eu/justice\\_home/fsj/privacy/law/index\\_en.htm](http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm). Accessed November 10, 2008.
3. European Parliament and Council of Europe. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Strasbourg, France: European Parliament and Council of Europe; 2001. Available at: [http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol1\\_en.htm](http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol1_en.htm). Accessed November 10, 2008.
4. Antokol J. Protecting Personal Data in Global Clinical Research. *The Monitor*.2008;22:57–60.
5. Code of Federal Regulations, Title 45, Part 164.501, Uses and disclosures for which consent, an authorization, or opportunity to agree or object is not required. Washington DC. US Government Printing Office; 2002. Available at: [http://www.access.gpo.gov/nara/cfr/waisidx\\_02/45cfr164\\_02.html](http://www.access.gpo.gov/nara/cfr/waisidx_02/45cfr164_02.html). Accessed November 10, 2008.

## Further Reading

Love CB, Thomson EJ, Royal CD. Current Bibliographies in Medicine 99-3: Ethical issues in research involving human participants Web page. Available

from: [http://www.nlm.nih.gov/archive//20061214/pubs/cbm/hum\\_exp.html](http://www.nlm.nih.gov/archive//20061214/pubs/cbm/hum_exp.html).  
Accessed November 10, 2008.

## **United States of America**

US Department of Commerce. Safe harbor documents Web page. Available at: [http://www.export.gov/safeharbor/SH\\_Documents.asp](http://www.export.gov/safeharbor/SH_Documents.asp). Accessed on November 10, 2008.

US Department of Health and Human Services. *Health Insurance Portability and Accountability Act of 1996*, (HIPAA) Public Law 104-191, as amended, 42 United States Code 1320-d. Washington, DC: US Government Printing Office; 1996. Available at: <http://aspe.hhs.gov/admsimp/pvcrec0.htm>. Accessed on November 10, 2008.

US Department of Health and Human Services Data Council. Office of Data Policy Web page. Available at: <http://aspe.hhs.gov/datacncl/>. Accessed November 10, 2008.

US Department of Health and Human Services. *Code of Federal Regulations*, Title 45, Volume 1, Parts 160 and 164. Washington, DC: US Government Printing Office; 1998.

US Department of Justice. *Federal Privacy Act*, 1974. PL 93-579,5 USC 552a, as amended. Washington, DC: US Government Printing Office; 1974. Available at: <http://www.usdoj.gov/oip/privstat.htm>. Accessed November 10, 2008.

## **European Union**

The European Group on Ethics in Science and New Technologies. *Opinion No. 13: Ethical Issues of Healthcare in the Information Society*. Brussels, Belgium: The European Commission; 1999. Available at: [http://europa.eu.int/comm/european\\_group\\_ethics/docs/avis13\\_en.pdf](http://europa.eu.int/comm/european_group_ethics/docs/avis13_en.pdf). Accessed November 10, 2008.

European Parliament and Council of Europe. *Directive 95/46/EC of the European Parliament and of the Council 24 October 1995 on the protection of individuals with regard to the processing of personal data and the free movement of such data*. Strasbourg, France: European Parliament and Council

of Europe; 1995. Available at:  
[http://europa.eu.int/eur-lex/en/lif/dat/1995/en\\_395L0045.html](http://europa.eu.int/eur-lex/en/lif/dat/1995/en_395L0045.html). Accessed  
November 10, 2008.

## **Australia**

Attorney-General's Department. *Privacy Act 1988. Act No. 119 of 1988, as amended*. Canberra, Australia: Office of Legislative Drafting and Publishing; 2006. Available at:  
[http://www.privacy.gov.au/publications/Privacy88\\_100107.pdf](http://www.privacy.gov.au/publications/Privacy88_100107.pdf). Accessed  
November 10, 2008.

National Health and Medical Research Council. *National Statement on Ethical Conduct in Research Involving Humans*. Canberra, Australia: Commonwealth of Australia; 1999.

## **Canada**

Kosseim P, ed. *A Compendium of Canadian Legislation Respecting the Protection of Personal Information in Health Research*. Ottawa, Canada: Public Works and Government Services; 2005. Available at:  
[http://www.cihr-irsc.gc.ca/e/documents/ethics\\_privacy\\_compendium\\_june2005\\_e.pdf](http://www.cihr-irsc.gc.ca/e/documents/ethics_privacy_compendium_june2005_e.pdf).  
Accessed on November 10, 2008.

Office of the Privacy Commissioner of Canada. *Personal Information Protection and Electronic Documents Act. S.C. 2000. c.5*. Ottawa, Canada: Office of the Privacy Commissioner of Canada; 2000. Available at:  
[http://www.privcom.gc.ca/legislation/02\\_06\\_01\\_01\\_e.asp](http://www.privcom.gc.ca/legislation/02_06_01_01_e.asp). Accessed  
November 10, 2008.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2003	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
April 2009	Revised for content, style, grammar, and clarity.

---

# Data Management Plan

December 2008

## *Abstract*

Every clinical study should have a data management plan to ensure and document adherence to good clinical data management practices for all phases of a study. This chapter identifies data management plan components and provides information on acceptable criteria for various sections of the plan. Although the clinical data manager will not personally perform all the tasks or prepare all the sections of the data management plan described in this chapter, the data manager should ensure all of these tasks and sections are completed according to good clinical data management practices.

## *Introduction*

Although a study protocol contains the overall clinical plan for a study, separate plans, such as a data management plan (DMP) or statistical analysis plan, should be created for other key areas of emphasis within a study. Before data collection begins, all clinical studies should have a DMP in place to document the relevant conventions for that particular study. A well-designed DMP provides a road map of how to handle data under any foreseeable circumstances and establishes processes for how to deal with unforeseen issues.

The optimal end result for a clinical data manager is to provide a study database that is accurate, secure, reliable, and ready for analysis. Many people will be involved in handling data throughout the course of a clinical study, so it is imperative that all parties refer to the DMP for a consistent approach to the processes and guidelines for conducting data management activities.

The DMP is an auditable document often asked for by regulatory inspectors and should be written in a manner that is professional and of high quality.

During an audit, the inspectors may also seek to ascertain the degree to which the project team adheres to the processes described in the DMP.

## **Scope**

Although style and format may differ from one organization to the next, this chapter gives a broad overview of the components and processes that make up a DMP. Whether the DMP document itself contains all of the elements or refers the reader to other study documents for further detail, this chapter provides the data manager with the minimal components that should be addressed within the overall study documentation.

## **Minimum Standards**

- Complete a draft of the DMP prior to enrollment of the first subject.
- Ensure the DMP supports compliance with applicable regulations and oversight agencies.
- Identify and define the personnel and roles involved with decision making, data collection, data handling and data quality control.
- Ensure data management processes are described and defined from study initiation until database closeout.

## **Best Practices**

- Develop the DMP in collaboration with all stakeholders to ensure that all responsible parties understand and will follow the processes and guidelines put forth in the DMP from study initiation to database closeout.
- Develop and maintain a DMP template for the organization that ensures consistency and standardization across all projects.
- Ensure the DMP for each study is kept current, including proper versioning, and that all responsible parties are aware of and agree to the current content.
- Ensure that an approved, signed version of the DMP is completed prior to starting on the work it describes. The job functions or titles that must

approve and sign the DMP may vary between organizations and depending on the type of study.

### ***Purpose of the DMP***

The DMP documents the processes and procedures employed by organizations to promote consistent, efficient and effective data management practices for each individual study. A primary goal of the DMP is to communicate to all stakeholders the necessary knowledge to create and maintain a high-quality database ready for analysis. The DMP serves as the authoritative resource, documenting data management practices and decisions that are agreed to at study initiation. The DMP should comply with all applicable regulatory guidelines (e.g., FDA, ICH, GCP) or local laws of the country; as well as the standard operating procedures (SOPs) of the organization. The DMP should also address any procedural or protocol updates that are made during conduct of the study.

### ***Creation and Maintenance***

For each new study, clinical data management (CDM) personnel should compose a detailed DMP based on the protocol, work scope, contract, analysis plans, dataflows, case report forms (CRFs), other supporting documents, and data management standards and practices. The entire DMP should be drafted and approved by all responsible parties prior to commencement of the work it describes. The clinical data manager should ensure the DMP is kept current, including proper version control, and that all parties involved agree with the content. Upon conclusion of the study, the DMP should be archived with all other pertinent study documentation.

The DMP should be created during the setup phase of each study and should contain information relating to all aspects of data management activities to be performed. The DMP should be considered a living document throughout the life cycle of a study, capturing any changes impacting data management made to the protocol or processes being used. The DMP must be uniquely identifiable, carry such identification on each page (e.g., study code/title) and be subject to version control. Each version should be documented and include date, author, reason for version change and an individual version identifier.

## **Organization of a DMP**

The organization, structure and order of topics presented in a DMP may differ between organizations. The following sections of this chapter cover the components that typically make up a DMP. Some of these components may be contained in documents referenced by the DMP rather than being detailed within the DMP itself. In either case, these components should be addressed within the overall study documentation.

### **Approval Page**

The approval page should detail the study identifiers and primary reviewers or signatories. The signature line(s) should include dates of approval. For companies allowing e-signatures, company requirements for e-signatures must be followed. The work detailed in the DMP should not begin until signatures are present from all relevant stakeholders.

### **Protocol Summary**

Many organizations may include a short synopsis of the study protocol, visit schedule, or critical data analysis variables within the DMP. This summary or synopsis gives a broad overview of the protocol and should refer the reader to the full protocol for more detailed information. Just as a DMP typically omits a full version of the study protocol, the DMP also typically omits a record of each protocol change or amendment. However, in some organizations the DMP may maintain a list of major protocol revisions and associated version numbers.

### **Dictionary and Coding Management**

The DMP should indicate which medical coding dictionaries (e.g., MedDRA, WHO Drug, SNOMED) and versions of the dictionaries will be used for the study. The DMP should reference documents providing instructions for how to handle dictionary updates or changes and define all quality control measures, validation methods, and user acceptance testing (UAT) for the dictionary. The DMP should also describe any auto-encoding or study-specific conventions used, as well as listing appropriate SOPs. Some examples of different types of coding include medication coding (prior/concurrent), adverse event (AE) coding, medical history coding, non-AE medical event coding (primarily for observational studies), and physical exam coding.



Please refer to the “Dictionary Management” and “Safety Data Management and Reporting” chapters of *Good Clinical Data Management Practices* for more information, including recommendations, minimum standards and best practices.

## **Definitions and Acronyms**

The DMP should include a list of acronyms that are specific to the protocol and DMP. Acronyms can be very helpful, but if their meaning is obscure they can become a hindrance. The DMP should also provide definitions of terms that may be misinterpreted or misunderstood.

## **Personnel/Role Identification/Training**

The DMP should specify key personnel with roles and responsibilities for the associated protocol and study activities, or the DMP may refer to external documents or related SOPs containing this information. The DMP should also refer to documents related to project-specific training requirements for various roles and functions.

## **Timelines**

The timeline included in the DMP or document referenced by the DMP lists expected completion targets for all deliverables. For example, database validation could be targeted for completion a specified number of weeks from the time the protocol is finalized.

Some organizations may have more detailed timelines, including more interim, internal activities; other organizations may have less detail, only tracking critical path activities. Timelines may also vary based on parameters of the study, such as between paper-based studies and those utilizing electronic data capture (EDC). Following are examples of milestones that may appear on a study timeline and be detailed in a DMP or associated documentation.

- Protocol finalization
- CRF development
- Database design and UAT

- Data validation, programming and UAT
- First patient first visit
- Last patient last visit
- Last CRF/data element received/entered
- Last query/discrepancy form received/completed
- Final SAE reconciliation completed
- Medical coding completed and approved
- Interim analysis, when applicable
- Database audit
- Database lock
- Study data and documentation archiving

### **Case Report Forms**

According to ICH E6, a CRF is defined as “A printed, optical, or electronic document designed to record all of the protocol-required information to be reported to the sponsor on each trial subject.”<sup>1</sup> The following are specific areas that should be elucidated within the DMP or other documents referenced by the DMP.

- CRF design—Provide a detailed description of the CRF design process or refer to the organization’s SOPs relating to CRF design and development.
- CRF instructions—Include general guidelines for CRF completion as well as protocol-specific guidelines.
- CRF changes—Describe the process for managing changes to the CRF design or reference the organization’s appropriate SOP. Changes to CRFs may also involve metadata changes, which should be governed by the same SOPs or one SOP designed specifically for the description of that process.

## **Database Design, Creation and Maintenance**

The DMP should refer to an in-depth study-specific database validation plan and include a brief description of how the database is created and maintained, a description of the system that is holding the data and table naming conventions. Title 21, Code of Federal Regulations Part 11 (21 CFR Part 11) mandates that procedures and controls be in place to ensure appropriate control of and access to documentation as well as revision and change control procedures to maintain an audit trail of modifications to documentation.<sup>2</sup>

## **Database Archive**

The DMP should outline specific information regarding the organization's procedures for archiving the electronic records.

## **Database Roles and Privileges**

The DMP should include profiles for available database roles within the system being used to support the study. Assign privileges to roles based upon the duties performed in the study. At a minimum, the roles should be listed or a reference should be made to a document where the roles are described. A detailed description of each role and the associated privileges is optimal.

## **Database Security**

The DMP should describe or refer to documents that describe the security of networked equipment and servers as well as security features of the electronic records within the clinical data management system (CDMS). The database security section of the DMP should also address:

- Maintenance of user roles and access—Describe the procedure(s) or refer to the organization's SOPs for defining, creating and maintaining system user roles and access. This description should include the process for revoking access.
- Database backup—Outline database backup procedures, frequency and routines. The disaster recovery plan and database backup SOPs should also be referenced in this section.

## **Data Entry and Processing**

The DMP or referenced documents should define data entry and processing plans. Data handling guidelines provide details of general study rules, which may cover acceptable abbreviations, symbol conversions, incomplete dates, illegible text, allowed changes and self-evident corrections. Ensure the DMP or DMP-referenced documents provide clear guidance for all of the following areas where applicable:

- Data entry guidelines—Describe proper entry of various data elements, proper handling of data anomalies, proper handling of missing data, and proper notation of self-evident changes. A comprehensive list of accepted abbreviations as well as symbols and their translations should be included in the guidelines. This list may be presented using a table within the DMP or by referring to a separate document.
- Data discrepancy conventions—Develop guidelines to provide consistency in classifying and processing data discrepancies.
- Data receipt—Specify the type of receipt (paper CRF or EDC), the expected frequency of data receipt, and how data receipt will be tracked. This also refers to data transfers from any third-party vendors.
- Data processing—Describe how data will be processed upon receipt at the organization (either electronic or paper-based data).
- Data entry—Indicate who will perform data entry and whether single or double entry will be used.
- Self-evident corrections—Specify the criteria for self-evident corrections and identify authorized data management personnel who will make these corrections to the data as necessary. A self-evident correction is a change to data or resolution of a query that can easily and obviously be made on the basis of other existing information on the CRF without sending a query to the investigative site. The most common self-evident corrections are obvious spelling errors. Self-evident corrections, like all other data changes, must be clearly documented and audited via the audit trail within the organization's database system. A list of approved self-evident corrections must be included in the DMP or exist in a separate document to be attached or referenced. Ensure the investigators associated with the study are in agreement with the self-evident correction process and that

the method of additional documentation (e.g., generation of reports for sign off) is thoroughly described. Self-evident corrections might not be applicable to all data management systems and types of data (e.g., source records).

- Data reconciliation—Provide details about the data fields and external databases requiring reconciliation per the study protocol.
- Database lock—Provide details defining the criteria for database lock, who will be responsible for database lock, and processes that will be employed in locking the database. Refer to the organization’s SOPs on study closeout as well. The DMP may also contain or refer to other SOPs for the unlocking and relocking processes if required.

Please refer to the “Data Entry and Data Processing” chapter of *Good Clinical Data Management Practices* for more information, including recommendations, minimum standards and best practices.

## **Data Validation and UAT**

The DMP should define validation test procedures to ensure integrity of study-specific components such as programming/algorithms, data entry/EDC screens, online logic/data-checking routines, security, backups, and archiving. If the DMP does not contain this information, it should reference a separate validation plan and/or validation and UAT SOPs. Please refer to the “Database Validation, Programming, and Standards” chapter of *Good Clinical Data Management Practices* for more information, including recommendations, minimum standards and best practices.

In addition to ensuring data entered into the database are complete, correct, allowable, valid, and consistent, other types of data quality checks may be applied. Once these checks have been identified, appropriate and verified programs are created to help identify discrepancies. All derivation and validation procedures may be fully tested and documented in the DMP or a referenced validation plan.

Data quality checks include:

- Manual review specifications—Describe all types of manual review specifications. Some aspects of these checks may be identified electronically depending on the features of the CDMS utilized. Other manual reviews (e.g., medical history, adverse events, concomitant medications reports, header information) may be generated via the CDMS; however, reviews of these data are usually accomplished through visual inspection.
- Discrepancy management—Describe the query process in detail, including how data clarification forms for paper studies or electronic queries for EDC studies are to be raised, tracked and handled when resolved, the annotation of any working copy CRFs and the documentation to be filed or retained. If different statuses are used for discrepancies, they should be defined.
- Electronic data discrepancy management—Define and describe processes to resolve electronic data discrepancies for the dataset or module being checked. These processes should include presentation of information which may include the CRF module, variable description, name of the edit check, processes for the use of test cases, a description of the edit check, an output message that would translate to a data query, other associated variables in the case of cross-checking data, and processes for documentation of these testing and validation activities.

### **SAE Data Reconciliation**

The DMP should describe or refer to documents that describe the protocol specific SAE reconciliation plan.

### **Quality Assurance/Control Processes**

The DMP should define quality assurance (QA) plans and quality control (QC) process steps. As defined by ICH E6, quality control is “the operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the study-related activities have been fulfilled.”<sup>1</sup>

Because studies of differing levels of regulatory importance are undertaken, occasionally a study will not be carried out within the established quality system. If this is the case, the study may not follow any SOPs in place or may only follow some of them. Complete an SOP compliance checklist indicating which SOPs are applicable to the study. Document in the comments section of the SOP compliance checklist any justification for opting out of all or part of the SOPs.

The DMP should address:

- Level of checks—Decide on and specify the required level of checking to be performed before data collection begins. Depending on the type and regulatory importance of a study, different levels of checking may be implemented. For example, an observational study may need only a minimal level of checking, whereas a highly regulated drug or device study requires a much more stringent level of QC checking.
- Frequency of quality control checks—Specify the frequency of QC checks in the DMP. According to ICH E6, “Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly.”<sup>1</sup>
- QC check documentation processes—Define the means by which QC checks are documented and how this documentation is maintained throughout the course of the study.

For more information about quality assurance and quality control, please refer to the chapters entitled “Assuring Data Quality” and “Measuring Data Quality.”

## **External Data Transfers**

For external data transfers, the DMP should describe the data type (e.g., safety lab data), the entity providing or receiving the data and any applicable agreements, the format, the frequency of transfers, and contact information for all those involved with the data transfer. Good practice is to have an established data transfer plan and to conduct a test data transfer prior to the need for a live transfer.

Specific data transfer details may include, but are not limited to, the following:

- Variable/element specifications
- Format of transfer (SAS<sup>®</sup> datasets, ASCII files, XML files, etc.)
- Method of transfer (encrypted e-mail, FTP, CD, DVD, etc.)
- Recipient of data (site, sponsor, data safety monitoring board (DSMB), statisticians, etc.)
- Frequency of transfer
- Quality control/validation steps performed to maintain integrity

The DMP should describe procedures used for collecting and handling laboratory data. If data comes from any combination of central labs, core labs, local labs, or specialty labs, there should be a short section differentiating between procedures for collecting and handling different types of lab data. Include or reference guidelines on how to transport, track, clean and report upon the various types of laboratory data.

Please refer to the “External Data Transfers” chapter of *Good Clinical Data Management Practices* for more information, including recommendations, minimum standards and best practices.

## **Audit Plans**

The DMP should either define the on-site audit and corrective action plans, or refer to those documents that do cover these processes. All interim and final study database audits should also be defined. As defined by ICH E6, quality assurance is “all those planned and systematic actions that are established to ensure that the study is performed and the data are generated, documented (recorded), and reported in compliance with GCP and the applicable regulatory requirements(s).”<sup>1</sup>

The DMP should also define how often during the course of a study QA will take place. Please refer to the “Assuring Data Quality” chapter of *Good Clinical Data Management Practices* for more information, including recommendations, minimum standards and best practices.



## Metrics

The DMP should include the metrics that will be used for the study. Please refer to the “Metrics for Clinical Trials” chapter of *Good Clinical Data Management Practices* for a list of commonly used metrics.

## Reports

The DMP should include a list of available reports for dissemination throughout the life of the study. For each report, specify the target audience, content of the report, level of detail provided, date of data extraction, frequency of generation and the mechanism used for distribution (e.g., e-mail, posting electronically). Additions and deletions to the report listing may occur throughout the life of the study and should be updated in subsequent versions of the DMP.

## Communications

The DMP should describe the types of communications or correspondence used in the study. Detail where records of these communications (whether paper or electronic) will reside, as well as any associated archiving requirements. Document how communications will be conducted and outline regularly scheduled communications. Indicate where to find communications after the fact. For example, if there is a particular form that must be signed and faxed, an auditor could see this in the DMP and not waste time searching through e-mails.

The DMP should include information on:

- Frequency of communication—Describe how frequency of communication may vary throughout the course of a study. For example, the communication may be more frequent in the setup and early stages of the study, then become less frequent as the study progresses. During the study conduct, many communications may be limited to study maintenance issues. During the closeout and lock portion of a study, communication frequency may increase again. Although most studies will have communication variability of this nature, specify any regularly scheduled communications in the DMP.

- Medium (e.g., face-to-face vs. conference call vs. Web conference)—Describe the estimated amount and timing of meetings, as well as which medium will be used. Try to schedule one or two face-to-face meetings (or more depending on length) during the course of a study. Web conferences are a good medium to share information in real time, such as when collaborating with the study team to edit a document or modify a process.
- Escalation process—Determine if issues need to be moved up the chain of command, when is it appropriate, and which parties should be involved.

### **Other Processes**

Every study is unique to some degree, and there may be processes within a particular study that have not been covered within this chapter. If a study involves other processes, they should always be described in detail somewhere within the protocol or DMP. Some additional processes that may need to be examined include the following:

- DSMB requirements—Describe any requirements pertaining to DSMB meetings that may occur during the course of the study. What preparation is expected to be performed prior to these meetings? Will this preparation be treated as a lock in regards to having all data clean and reported upon prior to the meeting? Will the DSMB be focusing on a sample of the data or the complete data set?
- Business rules—Specify business rules that may have an impact on data handling or data integrity in the DMP. For example, regularly scheduled IT maintenance that limits server access, organization-wide observed holidays or an anticipated change of address during the course of the study may affect data handling.
- Flowcharts and forms (e.g., CRFs, source documents, adjudication and query forms)—Include applicable flowcharts or sample forms that may be required by your organization.
- Problems and resolutions—Document the process of identifying, discussing, resolving and filing problems arising and resolved during the study.

- Change control processes—Evaluate if other change control processes may be encountered during the course of the study and describe them in the DMP.
- Blind data review specifications—Describe the expectation from data management if a blind data review will be conducted.
- Archival and record retention process—Describe when and how the archival process occurs. The processes described revolve around current organizational and governmental regulations. There are certain requirements that must be met according to applicable regulatory and/or sponsor requirements.<sup>1</sup> Document the record retention timeframe and communicate this timeframe to site personnel.

### ***Recommended Standard Operating Procedures***

- CRF Design and Development
- Database Design and Testing
- Data Management and Systems Roles and Responsibilities
- Coding Dictionary Management
- System Security
- Change Control
- Data Entry
- Internal Data Handling
- External Data Handling
- Data Cleaning
- SAE Data Reconciliation
- Quality Control
- Database Lock and Unlock

- Study Data and Documentation Archival

## **References**

1. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E6: Good Clinical Practice; Consolidated Guideline*. Geneva, Switzerland: Author; 1996.
2. *Code of Federal Regulations, Title 21, Volume 1, Part 11*. Washington, DC: US Government Printing Office; 1998.

## **Further Reading**

*Guidance for Industry: Computerized Systems Used in Clinical Investigations*. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Office of the Commissioner (OC); May 2007.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
December 2008	Initial publication.

---

# Project Management for the Clinical Data Manager

June 2010

## *Abstract*

Clinical data managers often assume some degree of project management responsibilities. This chapter discusses the discipline of project management and how to effectively apply project management principles to clinical data management. The chapter describes specific project management activities within a clinical data management department, and discusses the desired competencies of a data manager assuming project management responsibilities.

## *Introduction*

Project management is crucial for the success of any project or endeavor. However, many people mistakenly believe that project management skills only equate with being organized and being able to communicate. Although project management does require organization and communication skills, it encompasses much more than these two skills. Project management is a unique discipline that can be described as "...the application of knowledge, skills, tools, and techniques to project activities to meet the project requirements."<sup>1</sup>

The degree of project management activities performed by data managers varies widely between organizations. Many organizations will have separate departments for project management; however, data managers should know basic principles of project management, regardless of the extent of project management activities that are assigned to clinical data management (CDM). Effective application of project management principles results in improved quality and timeliness of CDM deliverables, as well as increased efficiency of CDM functions.

Although a clinical study can be thought of as a single project, each clinical study is made up of many components. One component is CDM, the ultimate

goal of which is to complete a study with a quality dataset that is appropriate for analysis. As the individuals responsible for overseeing CDM, data managers need to have varying degrees of project management skills. In addition to managing the internal resources and timelines of a study, CDM project managers may also manage external vendor relationships, necessitating awareness of contractual resources and scope restraints. For detailed information about managing external relationships, see the GCDMP chapter entitled “Vendor Selection and Management.”

## **Scope**

This chapter discusses project management principles and activities as applied to CDM within the context of a single study. Although project management within clinical research has a scope that encompasses much more than CDM activities, this chapter will not address project management activities and responsibilities that are beyond the CDM activities of an individual study. The activities described in this chapter are not applicable for all data managers, but do usually apply to those who are project leads or who assume project management responsibilities within CDM.

## **Minimum Standards**

- Identify all data management study team members, stakeholders, and respective alternates wherever possible and as early in study setup as possible. Ensure information is documented and updated regularly, with documentation centrally located or otherwise easily accessible to the study team regardless of their physical location. Clearly identify the individual(s) responsible for information updates. For an example of what should be included in a project plan, see Appendix A: Sample Project Plan Template.
- Identify, define, and document all study-specific processes. Any planned study-specific deviations from organizational SOPs and the rationale for the deviations should be brought to the attention of quality assurance personnel and logged for discussion during future SOP review cycles.
- Ensure clear, comprehensive, and technically feasible timelines with dependencies that are created and documented such that all personnel are

in agreement and can access timelines relative to their scheduled tasks. This may take the form of Gantt charts (derived from a project plan).

- Monitor, track and document projected costs and timelines against actual expenditures and deliveries (e.g., comparison of percentage of work completed to the percentage of budget spent).
- Identify potential risks to the project or study. Develop early warning signals and response strategies for each identified risk (e.g., risk mitigation plan). Review and adjust study-specific contingencies in accordance with study life cycle.
- Create and propose to the project team a communication plan, which, upon approval, shall be adhered to by all study personnel and stakeholders. The plan should be specific and easy to follow based on individual end user needs. The plan should identify a schedule for routine communications, the means by which these communications will be conducted, and how communications will be documented and archived. Common elements may include issue categories and associated severity codes, severity-based time/resource/cost impact, escalation rules, and resolution plans. For an example of what should be included in a communication plan, see Appendix B: Sample Communication Plan Template.
- Assure a thorough assessment has been made of CDM team members' familiarity with clinical study processes, disciplines, or functional lines.
- Ensure appropriate project- or study-specific training is delivered, maintained and documented for all study personnel performing CDM tasks.
- Ensure adequate and compliant electronic, virtual, and physical resources will be available for intake and archival of final accepted CDM deliverables. This may involve working with personnel from different departments, including information technology (IT), legal, and regulatory operations, as well as external vendors.

## **Best Practices**

- Create a responsibility matrix that describes activities to be conducted during the course of the study.
- Conduct regular meetings with the study team (may be conducted via Web or telephone conferences). During these meetings, track progress and upcoming milestones, and discuss corrective actions if needed.
- Continually assess project processes and modify processes as needed to function more efficiently. Ensure all process changes are communicated, documented, and version controlled. File this documentation within the study master file in effort to establish a clear audit trail.

## **Overview of Project Management**

A project can be defined as “A temporary endeavor undertaken to create a unique product, service, or result.”<sup>1</sup> Data managers should know basic principles of the formal discipline of project management to achieve the results desired from CDM. As with any other scientific or business-related discipline, project management employs basic theoretical constructs that underpin effective implementation. As a formal discipline, project management seeks to successfully complete specific projects in an effective and efficient manner by applying standard principles to project planning, organization and management.

### **Five Stages of Project Management**

Every project can be divided into the following five primary stages, although each of these stages can be subdivided into numerous smaller stages and steps.

- *Initiating* defines the scope and nature of a project, identifying the project’s primary goals and stakeholders.
- *Planning* lays the groundwork for a project by developing project timelines, establishing project milestones, identifying needed resources and personnel, and establishing processes to be followed and tasks to be completed during the project.



- *Executing* follows up on the planning phase by implementing the processes and tasks that were previously defined.
- *Monitoring and controlling* refers to processes and tasks that are intended to ensure project execution is progressing as intended. This phase encompasses assessing project metrics and implementing corrective actions, if needed.
- *Closing* encompasses activities undertaken as a project comes to an end, including file archival and documentation of lessons learned, which can subsequently be applied to future projects.

### **Nine Knowledge Areas of Project Management**

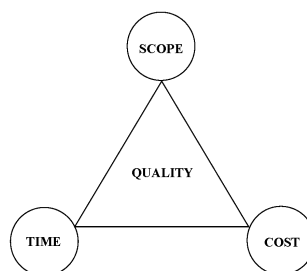
In addition to the five stages of project management, the discipline is divided into nine key knowledge areas. Effective project management should examine each of these areas to ensure all aspects of project needs are adequately addressed.

- Integration management “...includes the processes and activities needed to identify, define, combine, unify, and coordinate the various processes and project management activities.”<sup>1</sup>
- Scope management “...includes the processes required to ensure that the project includes all the work required, and only the work required, to complete the project successfully.”<sup>1</sup>
- Time management “...includes the processes required to manage timely completion of the project.”<sup>1</sup>
- Cost management “...includes the processes involved in estimating, budgeting, and controlling costs so that the project can be completed within the approved budget.”<sup>1</sup>
- Quality management “...includes the processes and activities of the performing organization that determine quality policies, objectives, and responsibilities so that the project will satisfy the needs for which it was undertaken.”<sup>1</sup>

- Human resource management “...includes the processes that organize, manage, and lead the project team. The project team is comprised of the people with assigned roles and responsibilities for completing the project.”<sup>1</sup>
- Communication management “...includes the processes required to ensure timely and appropriate generation, collection, distribution, storage, retrieval, and ultimate disposition of project information.”<sup>1</sup>
- Risk management “...includes the processes of conducting risk management planning, identification, analysis, response planning, and monitoring and control of a project. The objectives of project risk management are to increase the probability and impact of positive events, and decrease the probability and impact of negative events in the project.”<sup>1</sup>
- Procurement management “...includes the processes necessary to purchase or acquire products, services, or results needed from outside the project team. The organization can be either the buyer or seller of the products, services, or results of a project.”<sup>1</sup>

Although all nine of these knowledge areas are important components of project management, a special relationship exists between scope, time, and cost management. Sometimes known as the triple constraint, these three areas are often presented as a triangle, as depicted in Figure 1. If any one of these three components changes, the other two are also impacted. If the scope of a project increases, time and costs will typically increase as well. If the allotted time of a project is reduced, the scope must also be reduced in most cases. Although quality is an area in which most project managers do not want to compromise, changes to any of the three components of the triple constraint can negatively impact quality if changes are not properly balanced.

Figure 1. The Triple Constraint



## Meetings

Because meetings are an integral part of successful project management, regular meetings are specified for all five stages of project management. Each meeting should have a predetermined agenda and be documented via meeting minutes recorded by someone other than the individual leading the meeting. Ideally, meeting agendas and minutes should be formatted according to a standard predetermined template. Meeting attendance should be documented, and all meeting documentation should be appropriately archived.

Progress and upcoming milestones should be discussed during meetings, as well as corrective actions, if needed. As milestones are achieved, collect and compile lessons learned up to that milestone. This can be accomplished by adding the lesson learned as a note attached to the milestone in the project timeline. Collecting and processing lessons learned on an interim basis will facilitate earlier process improvements toward achievement of the next milestone, and make the final lessons learned meeting at the end of the study more robust.

## ***Project Management Activities Within CDM***

To effectively apply principles of the project management discipline to CDM, data managers should determine which tasks belong in each of the five stages of project management. Although this chapter classifies CDM activities according to each of the five stages, stage assignment of various tasks may vary between organizations and studies. Some activities may also relate to more than one stage and some of these activities may not be the direct responsibility of CDM personnel in all organizations.

### Initiating

During the initiation stage of a study, CDM tasks include, but are not limited to, the following activities.

- Share and discuss with individual(s) responsible for compiling requests for proposal the forecasted task/resource/time requirements as assessed for the study. This is the data manager's opportunity to clarify assumptions to be included in the contract. The result of this discussion will lead to an

accurate proposal and ultimately reduce out-of-scope hours or activities during the study life cycle.

- Define CDM contributions to the overall study team mission statement, scope, and goals in accordance with a funded portfolio or contract.
- Form or confirm core CDM team for the study: identify data manager(s) and CDM support personnel. Depending on the company's structure, the core CDM team may include IT support personnel, database programmers, and other team members who will contribute to CDM activities throughout the life cycle of the study. This core CDM team should meet regularly throughout the course of the study.

## **Planning**

The planning stage is crucial to the ultimate success of any project. During the planning stage, data managers assess study needs and determine how to best meet those needs. Planning includes, but is not limited to, the following CDM activities.

- If applicable, review the finalized study protocol or clinical development plan (CDP), which would typically include definitions of research questions, hypotheses, estimated study duration, and estimated number of subjects required to achieve statistical power. Note, data managers working for CROs may not have access to finalized CDPs.
- Assess resources and training needed for study execution (including personnel, hardware, software, and budget, as applicable).
- Document roles and responsibilities within the study. Responsibilities are often documented with a RACI chart (Responsible, Accountable, Consulted, Informed), which is a matrix specifying which individual(s) or group(s) will be responsible for each activity or group of activities. For an example of what should be included in a RACI Chart, see Appendix C: Sample RACI Chart Template where the RACI activities and designations can be changed according to your organization's needs. In some companies, the roles and responsibilities are described in SOPs and Work Instructions, and they should not be duplicated for the specific project.

- Confirm technical qualification of sites has occurred. Although IT personnel will handle many of the details of technical qualification, data managers should be involved with making certain specific functions (e.g., data uploads) operate as intended. This confirmation includes ensuring necessary hardware and software provisioning are in place at all sites, including validation documentation.
- Identify relevant stakeholders and confirm their individual roles and expectations. Facilitate introductions to stakeholders outside CDM who may impact or be impacted by CDM, and schedule regular meetings with these stakeholders.
- Identify vendors and service providers to be involved with the study, including confirmation of vendor qualification and contracts (e.g., requests for proposal development).
- Confirm identification to or from vendors regarding version of licensed tools used in the study, such as medical coding dictionaries. For example, if MedDRA is used, both the sponsor and the company coding data must have a current MedDRA license, although sites, monitors, biostatisticians, and CROs handling raw data do not need to be licensed. For any licensed tool, make certain the conventions of the license-issuing entity are closely followed.
- Develop high-level CDM project milestones and disseminate these milestones to the lead project manager for incorporation into the overall operations project timelines. More detailed CDM project milestones are addressed during the execution stage.
- Review the study protocol for consistency throughout and document and communicate inconsistencies to the trial manager or study team. Note that some organizations may not involve CDM until the protocol is approved.
- Review preprogrammed metrics reports and any other standard reporting tools for content, usability, and format. Share these templates with the team to solicit feedback and ensure end user reporting requirements are met. Identify and document report customizations or new metrics and reports requirements so necessary programming can be completed and outputs are ready immediately during study execution.

## **Executing**

As project management activities transition from the planning stage to the executing stage, many high-level planning tasks are defined in more detail. Executing includes, but is not limited to, the following CDM activities.

- Establish date for internal kickoff or initiation meeting (terminology may vary between organizations), and ensure all core CDM team members and support staff attend.
- Ensure technical and procedural training has been delivered to all internal or external staff, and that training is documented and archived. Provide similar training and documentation as staffing changes dictate.
- Ensure access (including passwords) to systems is enabled as appropriate, and confirm installation of any hardware or equipment, if applicable.
- Develop detailed CDM timelines that include, but are not limited to, a list of core deliverables such as CRFs, database build, edit check specifications, production database deployment, medical coding reviews, database lock, and interim or final quality reviews. After development, accepted timelines should be stored in a central location accessible to the study team. Within the description of each deliverable, identify the detailed CDM subtasks required. These detailed subtasks that drive intradepartmental staff activities may not need to be shared with the entire study team.
- Develop the data management plan (DMP), CRFs, CRF completion guidelines, database structure building plans, and other necessary documents and reports. For more information about DMPs and CRF completion guidelines, see the GCDMP chapters entitled “Data Management Plan” and “CRF Completion Guidelines.”
- Identify medical coding practices, dictionaries to be used, frequency of listing reviews (e.g., ambiguous term reports and unique term reports), and the frequency of dictionary upgrades or updates both during study conduct and at the point of final study reporting.
- Establish a detailed communication plan to ensure methods of communications between vendors, sponsor, and sites are clear, and storage locations are documented.

- Participate in investigator meetings or other appropriate training venues. It is critical to ensure site personnel and investigators receive adequate training and understand CDM expectations for the study, including proper completion of study-related documents. For more information about CDM's participation in investigator meetings see the GCDMP chapter entitled "CDM Presentation at Investigator Meetings."
- Perform an internal assessment to confirm the quality of the first group of data received.

### **Monitoring and Controlling**

Once the study is underway, CDM project managers should begin performing monitoring and controlling procedures, which include, but are not limited to, the following activities.

- Verify with major stakeholders that initiation plans continue to be aligned with the project plan.
- Conduct midstudy vendor/CRO assessment(s) as necessary, including confirmation that all vendor contracts are being adhered to in a satisfactory manner. Assessments are typically made using internal study timelines, predefined metrics reports, and the vendor contract as the basis for comparison.
- Conduct core CDM team meetings according to a predetermined schedule, although additional meetings may be held as needed. In addition to these meetings of the core CDM team, data managers should attend the project team meetings with other functional groups.
- Verify all planned production reports continue to meet user expectations. There should be minimal changes to these previously validated outputs in terms of content and format. At this point, the study team may need to provide justification because any further modifications or customizations could be considered out of scope. The first set of finalized outputs has probably been distributed based on previously agreed frequency and recipient lists.
- Carefully monitor study reports and metrics.

- Evaluate CDM team performance.
- Initiate corrective actions when deemed necessary using the risk management plan as a guide.
- Identify, plan, and carry out training/retraining as necessary, and provide additional training sessions for new staff that may come on board during study conduct. Ensure that staff who leave the project have returned all study materials, have study access removed and have completed appropriate exit consultations.

## **Closing**

Proper execution of the closing stage of CDM project management is crucial to ensuring the study's final deliverables meet expectations set at the initiation and planning phases of the study. Closing activities also help facilitate process improvements for future projects. Some of the CDM project management activities executed during closing include, but are not limited to, the following activities.

- Confirm all final deliverables are received or transferred and meet acceptable quality standards as defined by the organization's quality system. For information about quantification of quality standards, see the GCDMP chapter entitled "Measuring Data Quality," and for more information about quality systems, see the GCDMP chapter entitled "Assuring Data Quality."
- Ensure access (including passwords) to systems is restricted as appropriate, and confirm retrieval of any hardware or equipment, if applicable.
- Achieve, deliver and communicate database release to all relevant internal or external stakeholders. For more information about database release, see the GCDMP chapter entitled "Database Closure."
- Close all relevant contracts/procurements.
- Confirm all regulatory submission needs from CDM are met, such as annotated CRFs, sample blank CRFs, etc.



- Archive database(s), CRFs and data clarification forms (DCFs). For more information about data archiving, see the GCDMP chapter entitled “Clinical Data Archiving.”
- Archive CDM components of study master files.
- Convene closing meetings, which should encompass lessons learned.
  - An internal closing meeting should encompass corrective actions to improve processes in the future.
  - An external closing meeting should be held with sponsors and vendors to improve the working relationships for future collaborations.
- Confirm sites receive copies of electronic CRFs and DCFs.

### ***Competencies of Project Management***

A successful project manager should possess the skills needed to facilitate the success of each contributor to the project. Within the context of CDM, project management competencies should facilitate efficient production within CDM and departments affecting and affected by CDM. Although some of the competencies of a project manager are similar to those of a data manager, the interrelatedness of project management may require a higher level of proficiency.

### **Technical Knowledge**

Although project management is a unique discipline unto itself, a project manager should also be knowledgeable in the discipline(s) encompassed by the project. To be a good CDM project manager, one must first have the technical knowledge needed to be a successful data manager. In addition, a CDM project manager should have a good understanding of the operations of departments and stakeholders that impact or are impacted by CDM. A CDM project manager should also be well versed in the principles and practices of the discipline of project management.

## **Problem-solving Strategies**

Problem-solving abilities are crucial to the success of a data manager, but become even more crucial for a CDM project manager. Many of the problems that arise in CDM are similar to problems that may have been faced in the past, meaning SOPs and past experiences can often address these problems adequately. However, project management often faces unique challenges that require sound problem-solving strategies to devise unique solutions. The ability to accurately assess a potential problem and formulate a successful solution is imperative for an effective CDM project manager.

## **Facilitation/Communication/Mediation/Negotiation Skills**

Because project managers must coordinate with a wide range of roles and departments, CDM project managers need to have effective communication, facilitation, mediation and negotiation skills, and be able to summarize discussions and make appropriate decisions. The key to all of these skills is effective communication. A CDM project manager must be able to listen to study team members, understand their needs, and effectively communicate proposed solutions for meeting those needs. Without good communication skills, a CDM project manager cannot successfully manage the most important component of any project, which is the personnel involved with achieving the project's goals.

A CDM project manager typically interacts with personnel from multiple departments and stakeholders, necessitating the ability to facilitate and mediate communications and deliverables between those departments and stakeholders. When CDM project managers assume responsibility for managing external vendors, they may also be involved with negotiating contracts that fit the needs of the project while staying within time, scope and budgetary constraints.

## **Leadership**

Leadership is a quality that is needed in any individual managing and leading others. While this is a needed quality for data managers, leadership is even more important for those data managers assuming project management roles and responsibilities. Although data managers must provide leadership and direction for personnel in the CDM department, a CDM project manager must also provide leadership to individuals from other functional areas of the

project. Some of the specific leadership functions that a CDM project manager should perform include the following:

- A CDM project manager should establish standards of collaborative conduct that facilitate effective communication and teamwork between various personnel and departments. A CDM project manager should not only possess effective communication skills, but they should also be able to use their communication skills to facilitate effective collaboration, negotiation, mediation and teamwork between other members of the CDM and study teams.
- Standards of professional conduct are typically established by an organization's upper management. A CDM project manager should clearly adhere to these standards and lead others to do so by example.
- Team and individual coaching/mentoring are crucial to improving the skills of each individual and the team as a whole. A CDM project manager may mentor other data managers on the team, who will in turn coach or mentor other CDM personnel.
- A CDM project manager should establish what is expected of the team's performance, and continually assess team performance to gauge whether or not those expectations are being met. Performance assessments can identify areas where elements of the team are not meeting expectations, after which the CDM project manager can propose corrective actions to improve those areas. Any team is only as strong as its weakest link.

### ***Recommended Standard Operating Procedures***

- Vendor Management
- Contract Management
- Document Management and Version Control

## References

3. Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)—Fourth Edition*. Newtown Square, PA: Project Management Institute, Inc.; 2008.

## Further Reading

Terms used in this chapter may be found in the *Good Clinical Data Management Practices* Glossary.

International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*. Geneva, Switzerland: 1996.

Verzuh, E. *The Fast Forward MBA in Project Management*. New York, NY: John Wiley & Sons, Inc.; 1999.

## Chapter Revision History

Publication Date	Comments
June 2010	Initial publication.

## Appendix A: Sample Project Plan Template

Project Title

Project Plan Approval

Name	Role	Signature	Date
	Project Manager		
	Project Sponsor		
	Other Stakeholders		

Project Plan Table of Contents

### 1. Introduction

1.1. Project plan—Overview and the schedule and processes that will be involved in updating the project plan

1.2. Summary of the project

### 2. Business Requirements

2.1 Project history, business needs, and business drivers descriptions

2.2. Expected project benefits

2.3. Identification of responsibilities for any business-related changes

### 3. Project Definition

3.1. Project description, including project scope, costs, and schedule

3.2. Prioritized project objectives

### 4. Project Organization

4.1. Roles and responsibilities

4.2. Interfaces between different functional groups

4.3. Project authority hierarchy

### 5. Project Delivery Strategy

- 5.1. Lessons learned from previous projects
- 5.2. Project roadmap
- 5.3. Risk management strategy
- 5.4. Critical issues for project success
- 5.5. Project funding strategy
- 5.6. Contracting and procurement strategy
- 5.7. Monitoring and controls strategy
- 5.8. Project communication strategy
- 5.9. Project management tools to be used
- 6. Project Plan Details
  - 6.1. Revision history
  - 6.2. Revision strategy
  - 6.3. Review and approval
  - 6.4. Project plan distribution and archival
- 7. References
- 8. Appendices
  - 8.1. Project roles and responsibilities
  - 8.2. List of referenced documents
  - 8.3. Other appendices as needed for the specific project

## Appendix B: Sample Communication Plan Template

### Communication Plan Approval

Name	Role	Signature	Date
	Project Manager		
	Project Sponsor		
	Other Stakeholders		

### Communication Plan Table of Contents

1. Purpose (of the communication plan)
2. Scope (of the communication plan)
3. Overview (of communications planning methodology)
4. Roles and responsibilities
5. Project communications
  - 5.1 Communication goals
  - 5.2 Communication documentation
    - 5.2.1. Communications log
    - 5.2.2. Managing communication materials
6. Stakeholder management
  - 6.1. Stakeholder directory
  - 6.2. Stakeholder classification
7. Training
  - 7.1 Therapeutic area and protocol review
  - 7.2 Tools and systems
  - 7.3 SOPs and working processes
8. References
9. Communication plan revision history

10. Appendices

10.1. Role definitions

10.2 Stakeholder classification



Appendix C: Sample RACI Chart Template (3 pages)

<b>R</b>	<b>Responsible:</b> The individual(s) who performs a task. The doer, responsible for recommendation, action and implementation.																																																																																																																																																																																																								
<b>A</b>	<b>Accountable:</b> The individual (s) who is accountable for ensuring alignment with the overall plans and ensures the quality/performance/outcome of the activity.																																																																																																																																																																																																								
<b>C</b>	<b>Consulted:</b> The individual(s) or team(s) who are asked to provide input and/or insight prior to an action being taken or a recommendation being made.																																																																																																																																																																																																								
<b>I</b>	<b>Informed:</b> The individual(s) or team(s) who need to be informed when an action is taken or a decision is made either verbally or through documentation.																																																																																																																																																																																																								
	<table border="1"> <thead> <tr> <th></th> <th>Activity</th> <th>Lead CDM</th> <th>CDM staff</th> <th>Mgmt contact</th> <th>Quality contact</th> <th>Clinical contact</th> <th>S&amp;P contact</th> </tr> </thead> <tbody> <tr> <td rowspan="13">Study set-up</td> <td>Protocol review for CDM</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CRF review for CDM</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Randomization review</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Site Clarification Agreement Document</td> <td>A/R</td> <td>I</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Create study specification</td> <td>C</td> <td>A/R</td> <td></td> <td></td> <td>C</td> <td>C</td> </tr> <tr> <td>Complete test transfer for electronic data</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td>I</td> <td>I</td> </tr> <tr> <td>Set up Study File</td> <td>A</td> <td>R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Develop Data Validation Specifications (DVS)</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Data Capture Testing</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Creation of Data Entry Guidelines</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Run checks to test set-up</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td>I</td> <td>I</td> </tr> <tr> <td>Liasing with Central and/or Local Labs</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Develop and maintain Data Handling agreement</td> <td>I</td> <td>A/R</td> <td>C</td> <td></td> <td></td> <td>C</td> </tr> <tr> <td rowspan="6">Tracking</td> <td>Track receipt of CRF paper data from Site(s)</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Track receipt of CRF paper data from Data Entry Group</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Track and monitor miscellaneous data not captured in CRF i.e. electronic data</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Track outstanding data queries (DQ)</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Run missing page report</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Check received data against expected data (end of study/interim activity)</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="7">Data Processing</td> <td>Resolution of flags, automated and manual discrepancies including DCF creation and resolution</td> <td></td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Liaise with dictionary group for coding and query management of AEs and Con meds</td> <td></td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Action required as the result of Unique Terms report (UTR) and/or Ambiguous Terms report (ATR)</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Loading of electronic data</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Post-entry review of CRFs (ie: missing data/pages and reconcile DE flags)</td> <td></td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Cleaning of Lab data and liaison with lab personnel</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Complete SAE reconciliation activities (run reports, liaise with Pharmacovigilance Group (PVG) and dictionaries)</td> <td>I</td> <td>A/R</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Perform QC activities defined in the quality audit framework document</td> <td>I</td> <td>A/R</td> <td></td> <td>I</td> <td></td> <td></td> </tr> </tbody> </table>		Activity	Lead CDM	CDM staff	Mgmt contact	Quality contact	Clinical contact	S&P contact	Study set-up	Protocol review for CDM	A/R						CRF review for CDM	A/R						Randomization review	A/R						Site Clarification Agreement Document	A/R	I					Create study specification	C	A/R			C	C	Complete test transfer for electronic data	I	A/R			I	I	Set up Study File	A	R					Develop Data Validation Specifications (DVS)	I	A/R					Data Capture Testing	I	A/R					Creation of Data Entry Guidelines	I	A/R					Run checks to test set-up	I	A/R			I	I	Liasing with Central and/or Local Labs	I	A/R					Develop and maintain Data Handling agreement	I	A/R	C			C	Tracking	Track receipt of CRF paper data from Site(s)	I	A/R					Track receipt of CRF paper data from Data Entry Group	I	A/R					Track and monitor miscellaneous data not captured in CRF i.e. electronic data	I	A/R					Track outstanding data queries (DQ)	I	A/R					Run missing page report	I	A/R					Check received data against expected data (end of study/interim activity)	I	A/R					Data Processing	Resolution of flags, automated and manual discrepancies including DCF creation and resolution		A/R					Liaise with dictionary group for coding and query management of AEs and Con meds		A/R					Action required as the result of Unique Terms report (UTR) and/or Ambiguous Terms report (ATR)	I	A/R					Loading of electronic data	I	A/R					Post-entry review of CRFs (ie: missing data/pages and reconcile DE flags)		A/R					Cleaning of Lab data and liaison with lab personnel	I	A/R					Complete SAE reconciliation activities (run reports, liaise with Pharmacovigilance Group (PVG) and dictionaries)	I	A/R					Perform QC activities defined in the quality audit framework document	I	A/R		I		
	Activity	Lead CDM	CDM staff	Mgmt contact	Quality contact	Clinical contact	S&P contact																																																																																																																																																																																																		
Study set-up	Protocol review for CDM	A/R																																																																																																																																																																																																							
	CRF review for CDM	A/R																																																																																																																																																																																																							
	Randomization review	A/R																																																																																																																																																																																																							
	Site Clarification Agreement Document	A/R	I																																																																																																																																																																																																						
	Create study specification	C	A/R			C	C																																																																																																																																																																																																		
	Complete test transfer for electronic data	I	A/R			I	I																																																																																																																																																																																																		
	Set up Study File	A	R																																																																																																																																																																																																						
	Develop Data Validation Specifications (DVS)	I	A/R																																																																																																																																																																																																						
	Data Capture Testing	I	A/R																																																																																																																																																																																																						
	Creation of Data Entry Guidelines	I	A/R																																																																																																																																																																																																						
	Run checks to test set-up	I	A/R			I	I																																																																																																																																																																																																		
	Liasing with Central and/or Local Labs	I	A/R																																																																																																																																																																																																						
	Develop and maintain Data Handling agreement	I	A/R	C			C																																																																																																																																																																																																		
Tracking	Track receipt of CRF paper data from Site(s)	I	A/R																																																																																																																																																																																																						
	Track receipt of CRF paper data from Data Entry Group	I	A/R																																																																																																																																																																																																						
	Track and monitor miscellaneous data not captured in CRF i.e. electronic data	I	A/R																																																																																																																																																																																																						
	Track outstanding data queries (DQ)	I	A/R																																																																																																																																																																																																						
	Run missing page report	I	A/R																																																																																																																																																																																																						
	Check received data against expected data (end of study/interim activity)	I	A/R																																																																																																																																																																																																						
Data Processing	Resolution of flags, automated and manual discrepancies including DCF creation and resolution		A/R																																																																																																																																																																																																						
	Liaise with dictionary group for coding and query management of AEs and Con meds		A/R																																																																																																																																																																																																						
	Action required as the result of Unique Terms report (UTR) and/or Ambiguous Terms report (ATR)	I	A/R																																																																																																																																																																																																						
	Loading of electronic data	I	A/R																																																																																																																																																																																																						
	Post-entry review of CRFs (ie: missing data/pages and reconcile DE flags)		A/R																																																																																																																																																																																																						
	Cleaning of Lab data and liaison with lab personnel	I	A/R																																																																																																																																																																																																						
	Complete SAE reconciliation activities (run reports, liaise with Pharmacovigilance Group (PVG) and dictionaries)	I	A/R																																																																																																																																																																																																						
Perform QC activities defined in the quality audit framework document	I	A/R		I																																																																																																																																																																																																					

<b>R</b>	<b>Responsible:</b> The individual(s) who performs a task. The doer, responsible for recommendation, action and implementation.						
<b>A</b>	<b>Accountable:</b> The individual (s) who is accountable for ensuring alignment with the overall plans and ensures the quality/performance/outcome of the activity.						
<b>C</b>	<b>Consulted:</b> The individual(s) or team(s) who are asked to provide input and/or insight prior to an action being taken or a recommendation being made.						
<b>I</b>	<b>Informed:</b> The individual(s) or team(s) who need to be informed when an action is taken or a decision is made either verbally or through documentation.						
	<b>Activity</b>	<b>Lead CDM</b>	<b>CDM staff</b>	<b>Mgmt contact</b>	<b>Quality contact</b>	<b>Clinical contact</b>	<b>S&amp;P contact</b>
<b>Data Processing</b>	Run edit Checks	I	A/R			I	I
	Complete Database Release (DBR) and Database Freeze (DBF) authorization checklist	I	A/R				
	Complete data transfer to DM area	I	A/R			I	I
	Declare/send DBR and DBF communications	A/R	I			I	I
<b>Study Conduct</b>	Arrange study team kickoff meeting	A/R	C				
	Generate and communicate minutes	I	A/R	-			
	Protocol Amendments	A/R	I	-			
	Maintain DVS	I	A/R				C
	Develop Data Quality Assurance document for Clinical Study report	A/R	I		C		
	Maintain and complete Data Quality Assurance document for Clinical Study report	A/R	I		C		
	Maintain Study File	A	R				
	Data Management review of Reporting and Analysis Plan (RAP)	A/R	I				
	Generate and communicate status/metric reports	I	A/R				
	Review status/metric reports and action where needed to resolution	C	A/R				
	Provide outstanding DQ reports to Clinical if applicable	I	A/R			I	
	Provide missing page reports to Clinical if applicable	I	A/R			I	
	Problem solving and escalation of issues (to CDM)	C	A/R	I			
	Problem solving and escalation of issues (to Clinical and Statistics & Programming (S&P))	A/R /C	I			I	
	Negotiate timelines with Clinical	A/R	C	I		I	I
	Schedule validation batch jobs		A/R				
	Timely feedback on validation checks	I	A/R				
	Timely feedback on CRF completion (data quality and monitor performance)	C	A/R				
	Attend matrix team meetings (attendance from Clinical, S&P, CDM, Safety etc. at study level)	A/R	I			I	
	Attend DM update meetings	A/R	R				
	Monitor requirements for data looks and interim analyses	A/R	I			I	C
	Communicate to study team DBR achieved	A/R	I	I		I	I
	Communicate post DBR queries	A/R	I				
	Resolve post DBR queries	A	R			C	C
Adhoc Submission requests	A	R	I				
Action post DBR data edits	A	R			I	I	

<b>R</b>	<b>Responsible:</b> The individual(s) who performs a task. The doer, responsible for recommendation, action and implementation.						
<b>A</b>	<b>Accountable:</b> The individual (s) who is accountable for ensuring alignment with the overall plans and ensures the quality/performance/outcome of the activity.						
<b>C</b>	<b>Consulted:</b> The individual(s) or team(s) who are asked to provide input and/or insight prior to an action being taken or a recommendation being made.						
<b>I</b>	<b>Informed:</b> The individual(s) or team(s) who need to be informed when an action is taken or a decision is made either verbally or through documentation.						
	<b>Activity</b>	<b>Lead CDM</b>	<b>CDM staff</b>	<b>Mgmt contact</b>	<b>Quality contact</b>	<b>Clinical contact</b>	<b>S&amp;P contact</b>
<b>Study Conduct</b>	Post DBF edits approval request (after consultation with Manager)	A/R	I	I		I	I
	Request database lock	A	R			I	
	Unlock database and action edits if approval given	A	R			I	
	Reconciliation of Study File before archiving	A	R				
	Send documents to CDM in archive-ready state	A	R				
<b>Study Oversight</b>	Resource planning of processing activities (i.e. recruitment, staff changes on a study)	I	A/R	I		I	I
	Oversight of tracking tool	A/R	I			I	I
	Maintenance of RACI	A/R	C				
	Quality Review if planned	C	I	I	A/R		

**This page is intentionally blank.**

---

# Vendor Selection and Management

March 2010

## *Abstract*

Vendors provide services that are critical to the successful outcome of a clinical study, yet sponsors retain the ultimate responsibility for activities that are outsourced. If a sponsor is willing to give control of some study activities to a vendor, the sponsor should take measures to ensure the vendor is delivering products or services of acceptable and repeatable quality. This chapter provides recommendations for evaluating, selecting, and providing oversight of vendors to determine whether their services adequately meet quality expectations and regulatory standards.

## *Introduction*

Vendors are used in all aspects of clinical studies and have particular relevance in clinical data management (CDM) processes. Some examples of vendors relevant to CDM include contract research organizations (CROs), case report form (CRF) design and printing companies, electronic patient reporting tool providers, clinical laboratories, central readers, imaging, interactive voice response system (IVRS) providers, electronic data capture (EDC) and other software suppliers, and off-site storage and data hosting facilities. Before a vendor is selected, the end product or result desired from the vendor should be clearly defined and described.

International Conference on Harmonization (ICH) E6 states “Ultimate responsibility for the quality and integrity of the trial data always resides with the sponsor.”<sup>1</sup> Therefore, the sponsor must manage vendors in a fashion that ensures quality, integrity and reliability. Not only is this ICH statement relevant to the sponsor, it should also be important to all vendors having an impact on final data quality. Documented processes should be followed to ensure that quality data are received from vendors, as well as to consistently evaluate vendor services.

## Scope

The scope of vendor services differs widely across the industry, ranging from CRF printing to assistance with a regulatory submission. This chapter examines the communication of clear expectations between the vendor and the sponsor and some strategies for clearly documenting various areas of vendor oversight. The chapter also includes considerations for vendor qualification and the appropriate level of oversight needed, depending on the vendor's scope of work. Details and discussions regarding relationship management are beyond the scope of this chapter.

Some of the tasks described in this chapter may be joint responsibilities between different groups, just as there may be different groups involved in the implementation of various tasks. However, clinical data managers need to be conscious of whether or not these tasks have been performed in a satisfactory manner.

## Minimum Standards

- Document the sponsor's process and support functions that are needed to evaluate the use of vendor services.
- Evaluate and qualify (e.g., capacity, qualifications, experience, regulatory compliance, company stability, etc.) vendors *prior* to contracting for their services or products.
- Obtain a confidentiality agreement with the vendor prior to exchange of proprietary information.
- Create a contacts list that is centrally accessible to study team members.
- Determine and document whether the sponsor's or vendor's standard operating procedures (SOPs) (or a combination of procedures) are to be followed.
- Clearly define expectations, deliverables and responsibilities. Both the sponsor and the vendor must participate in establishing definitions of their roles.
- Conduct ongoing management of vendor activities. Communicate and assess the vendor's performance throughout the study.

## **Best Practices**

- Where feasible, evaluate from a CDM perspective the risk of utilizing or not utilizing vendor services related to the conduct and outcome of the study.
- Maintain an internally approved vendor list with regular evaluations (e.g., preferred vendor list or prequalified vendor list).
- Establish a cross-functional vendor auditing program based on established services, which should include plans to re-audit the vendor within a stated amount of time, if applicable.
- Oversee vendors by utilizing subject matter experts within a centralized organizational team to provide input into the processes of vendor evaluation, vendor audits, and issue resolution and escalation.
- Define and document a detailed statement of work and project plans that detail who is responsible for each task; who is responsible for reviewing and approving various documents; details of project reporting; or a checklist of tools, processes, and services to be performed by the sponsor and vendor at each phase of the study.
- Define and document detailed sponsor/vendor communication plans that clearly address the expected communication tools and frequency, as well as establish who is responsible for communications and how to escalate issues when deemed necessary.
- Identify other possible vendors or options as part of a contingency plan in case the vendor relationship is deemed unsatisfactory at any point during the course of the study.
- Establish a collaborative relationship based on partnership, trust and co-ownership of the project.
- If the vendor is providing services that involve computerized systems, ensure system support documentation is in place, such as a service level agreement (SLA), that describes in detail how much time it will take the vendor to respond to support inquiries, how long it will take to get a

database back online in case of a system failure, and other details related to supporting the sponsor's business requirements.

## ***Types of Vendor Services***

Each clinical study may require a variety of vendor services, depending on the needs of the study and resources already available within the organization. The following list contains some of the types of vendors most often utilized during the course of a clinical study:

- **Data management CRO**—An enormous range of services may be provided by data management CROs. Some CROs conduct all aspects of data management, while others may only perform select activities. Some of the types of specific services that may be encompassed by data management CROs include: project management; CRF creation; CRF guidelines creation; data management plan creation; database design and programming; edit check specifications development and programming; CRF tracking; data entry; data review and cleaning; data coding; serious adverse event (SAE) reconciliation; external data transfers; quality control audits; database lock(s); and database transfers.
- **CRF/document printer**—Not all printing companies have the equipment and expertise needed to print CRFs or other documents needed for a clinical study. Paper CRFs are typically printed in triplicate on carbonless copy paper. Ensure that the vendor used for printing these documents is capable of providing the end product needed for the study.
- **Translation services**—For studies requiring CRFs in multiple languages, accurate and reliable translation of the CRFs is crucial to collecting data that are consistently accurate and equivalent.
- **External data providers**—For vendors providing external data such as lab data or imaging and diagnostic data (e.g., ECG, MRI, CT), vendor evaluation should ensure vendors provide data that consistently meet quality standards defined for the study.
- **Software and hardware providers**—The technological needs of clinical studies have been increasing steadily over recent years. The advent of tools such as EDC and ePRO (electronic patient reported outcomes)



necessitate careful evaluation of vendors providing validated software, hardware, or database hosting services.

- Server or network providers—Whether servers and networks are hosted in-house or outsourced, an in-depth evaluation should be made to ensure servers and networks are stable, secure, and accessible only to authorized users. A disaster recovery plan should be available for inspection to help with evaluation of the provider.
- Coding services—Vendors providing coding services or licensing should be carefully evaluated to verify the appropriate training and experience of coding resources to ensure coding is performed accurately and consistently, and that all relevant licenses and documentation are maintained and up to date and the versioning frequency is assessed.

### ***Business Model Impact on CDM***

The business model followed by a vendor can significantly impact the relationship of the vendor with CDM personnel. The following are some of the more frequently encountered business models that may affect CDM personnel.

#### **Transactional Model**

The transactional model could be considered the traditional outsourcing model for clinical studies, in which a sponsor contracts vendors on a per project or per study basis. Transactional relationships may be more likely than other models to “...perform out-of-scope activities, resulting in cost overruns.”<sup>2</sup>

#### **Strategic Partnerships**

Strategic partnerships may be formed between companies with complementary resources and expertise, so as to increase efficiencies and lower overall costs. Strategic partnerships could be between a sponsor and a biotechnology company providing EDC or other electronic tools for clinical studies, or may be between a sponsor and a full-service CRO. Strategic partnerships may also be formed to gain location-specific resources needed for studies that span multiple countries or regions. Before forming a strategic partnership, carefully evaluate the potential partner to ensure there are no

significant differences between corporate cultures, philosophies or SOPs that could potentially lead to conflicts. Although strategic alliances may “...not result in lowest-bid providers, the long-term efficiencies, minimization of out-of-scope costs, and performance improvements theoretically surpass short-term cost savings.”<sup>2</sup>

### **Functional Service Provider (FSP) Models**

In contrast to outsourcing all data management aspects of a study to a single CRO, an FSP model may involve outsourcing only select activities. “Because project ownership remains in-house, companies that use functional outsourcing may experience higher levels of quality control yet have access to specific services at a lower overall cost. Sponsor companies benefit from being able to ramp up and draw down resources relative to their development activity levels without affecting their internal head count.”<sup>3</sup> Using an FSP model allows the sponsor to focus on their core competencies and outsource certain activities (such as CRF design or system validation) to niche vendors, rather than needing to hire additional personnel or provide additional training to existing personnel.

### **Application Service Provider (ASP) Models**

An ASP is a vendor that leases a software application to clients, and can involve contracts that are for the duration of a study, for a set amount of time, or on a per use (e.g., per user, per study, per CRF, etc.) basis. Using an ASP can shift much of the responsibility to the vendor for implementing, hosting, validating, maintaining, upgrading and supporting the software. However, because sponsors are ultimately responsible for data integrity and quality, a risk-based approach should be used to determine the scope and depth of any additional software testing and validation that may need to be performed.

## ***Vendor Qualification, Evaluation and Selection***

Before a vendor is selected, an evaluation of the vendor should take place. The sponsor and the vendor must understand specifically what services will be provided, and whether the data in the clinical study could be directly or indirectly affected by these services. Both parties should have a clear task ownership matrix defining who is responsible, accountable, consulted or informed with each specific task. However, before evaluating vendors,

qualification should be performed to ensure resources are not spent evaluating vendors that do not meet the needs of the sponsor. Vendor qualification should be determined by internally evaluating what services will be required, and defining desired attributes for the vendor that will eventually be contracted. For example, a sponsor may determine that they only wish to consider vendors of a certain company size or location, or they may have a preference of a full service CRO or a niche vendor.

The vendor evaluation should provide assurance that the vendor has sufficient staff to perform the contracted services under SOP requirements, and that staff are adequately qualified and trained to perform the regulated activities. A request for information (RFI) should be developed and sent to the vendor for precertification (see Appendix A for a sample RFI form). A full vendor evaluation should examine information provided in an RFI, and may include an on-site vendor visit to interview vendor personnel and review vendor processes and systems. A formal presentation at the sponsor site could also be conducted by the vendor as a response to the RFI, which may include the following:

- Company information, such as a historical overview of the organization, length of time in the industry, financial stability of the organization, and an explanation of the organizational structure
- Products and services
- Experience and areas of expertise (e.g., oncology, adaptive design, Phase 1, etc.)
- Product demonstration
- Computerized systems
- Results of previous regulatory inspections, as permitted

Sponsor staff or delegates who are subject matter experts in the activities being outsourced (e.g., clinical, biostatistics, medical monitors, or data management) should participate in vendor evaluation. The primary goals of vendor evaluation are to determine that adequate quality and sufficient resources are available to provide the defined services in compliance with regulatory expectations and defined quality standards. If a vendor is found to

have deficiencies, they may not necessarily need to be eliminated as a viable option. Deficiencies may be addressed by determining if actions can be taken to correct the deficiencies, or by determining what extent of controls need to be exercised over the vendor. An example of corrective action may be to implement process improvement or remediation. However, the remediation that will be required of the vendor should be defined and evaluated prior to entering into a contract. This remediation may take the form of SOP provisions, data management support, quality assurance (QA) or quality control (QC) advice, or documentation and system validation guidance.

The results of a vendor evaluation should be tailored to the services being provided. For example, CROs providing full service would require extensive evaluation, whereas a vendor that only provides printing for query binders may require a less comprehensive evaluation. From a business perspective, it is most advantageous to thoroughly qualify vendors *before* contracting with them for their services. Qualification prior to contracting can help avoid the need to have work redone due to the original work being of insufficient quality or in a system that is not compliant with relevant regulations. Redoing work or even an entire study can significantly lengthen project timelines and overall costs. See Appendix B for an example of topics that may be examined in an EDC vendor review.

Considerations when evaluating a vendor may include the following (not in order of priority):

- Financial stability of the vendor
- Mergers or acquisitions in the recent past and the impact on SOPs
- The vendor's experience with different business models
- The vendor's geographic capabilities
- The number of sponsors or studies currently supported by the available vendor staff
- References from previous customers
- Outcomes of previous regulatory inspections, as permitted

- Review of required accreditation in the vendor's field of work (e.g., lab certifications)
- Availability of documentation to regulatory authorities
- Review of the vendor's SOPs and work instructions to ensure soundness of processes and proof of regulatory and industry standards compliance
- Vendor's ability to adapt to sponsor's SOPs, if required
- Documentation of vendor's change control processes
- The vendor's quality system (e.g. computer systems, CDMS, databases, etc.) and proof of compliance to their quality requirements
- Evaluation of the vendor's QC/QA processes
- Sufficient staffing, including documented adherence to training and retraining plans
- Personnel qualification (through a review of curriculum vitae (CVs) of company personnel, job descriptions, organizational charts, training plans and documentation, etc.)
- Evidence of clearly defined project-specific training plans for new team members, and adequate transition processes to address staffing changes that occur during a study
- Documentation of system validation for regulated processes
- Data transfer processes
- Security of physical locations where services are provided (controlled facility access, server rooms, file rooms, independent backup procedures, etc.)
- Physical conditions of server and file rooms (limited access, fireproof, temperature and humidity controlled, etc.)
- Disaster/contingency plan(s) to protect business operations

- Evaluation of subcontractors and the vendor's management processes for those subcontractors, if applicable

After the vendor evaluation process is completed, a vendor is selected and is typically presented with an award letter providing notification of the vendor's selection. Larger organizations usually have law and procurement departments that handle award letters and final vendor selection activities, but CDM personnel may be involved in these processes at some organizations.

### ***Development of Contract and Scope of Work (SOW)***

Once potential vendors have been evaluated and vendor selections have been made, a contract and statement of the scope of work must be prepared and agreed upon by the sponsor and the vendor. Many large companies have separate departments that handle these details, but CDM personnel may be involved with these processes in some organizations.

### **Considerations for Sponsors, Vendors, and Data Managers**

The type of outsourcing business model used is the most important consideration in preparing the contract and the scope of work. Because numerous variations can exist between outsourcing relationships even when following the same outsourcing business model, the contract and the scope of work for each vendor relationship may also have unique variations.

When using models that involve more organizational integration, such as strategic partnerships or an FSP relationship, both organizations should commit to several levels of oversight (executive committees, operational committees, etc.) that focus on strategy and implementation to ensure that the partnership is successful. Each level of oversight should also be associated with a clear escalation path in case issues are unable to be resolved at a particular level. Governance models should ensure long-term senior management commitment from both sides.

For organizations using a transactional outsourcing business model, costs and scope of work are typically based on certain assumptions. Because some of these assumptions may be incorrect or based on changing information, the contract and scope of work should include provisions detailing how changes will be handled. These provisions should include a description of how changes

to underlying assumptions may result in change orders, as well as mitigation plans to resolve situations where the scope of work slowly evolves over time (i.e., scope creep).

Although typically the responsibility of a legal department, CDM personnel should be aware that contracts may include special clauses such as penalty clauses or bonus clauses. These clauses are intended to give vendors incentives for exceeding expectations, or disincentives for not meeting expectations.

### **Task Ownership Matrix**

A task ownership matrix identifies all tasks that may arise during execution of a clinical study. The matrix is intended to ensure all tasks are accounted for and to reduce the potential for duplication of effort. Not developing a task ownership matrix or developing one poorly can defeat the anticipated benefits that drove the parties to enter into an agreement in the first place. For example, if both parties duplicate efforts with a task because responsibility for the task is not clearly defined, duplicate costs are incurred and the desired monetary savings of the relationship may never be realized. The matrix should clearly identify four ownership responsibilities that occur with any task or document:

- Who is *responsible* for this task or document (e.g., creation, revision, approval)
- Who is *accountable* for this task or document
- Who is *consulted* for this task or document
- Who is *informed* for this task or document

The end result of a well-documented task ownership matrix, also known as a RACI (responsible, accountable, consulted, informed) table, will be a better relationship between the sponsor and vendor, as well as provide clear one-party accountability for success or remediation of various tasks. The task ownership matrix should be mutually agreed upon by both parties prior to study startup.

## **Bid Grid**

A bid grid (sometimes referred to as a roles and responsibilities or R&R matrix) is typically maintained by the sponsor procurement or vendor management office, although assignment of this responsibility may vary between organizations. A bid grid serves two primary purposes:

- A bid grid captures the sponsor's predefined study-specific cost drivers.
- A bid grid allows the outsourced partner to assign prices to specific tasks associated with cost drivers.

For all CDM cost drivers, a bid grid should include definitions of units, cost per unit, the estimated number of units expected, and total anticipated costs for each row. Columns may also capture which party is responsible for each activity and which party is accountable for each activity (the bid grid and task ownership matrix may be combined in some cases). The structure of a bid grid should be clearly aligned with the text portion of the SOW, and all tasks and units should be clearly defined, meaningful and measurable. In addition to specific CDM cost drivers, a bid grid may also include more high-level categories, such as CRF design or data cleaning.

Some high-level categories for pricing consideration include:

- CRF design
- Database development (including edit check specifications)
- Data management plan development
- Data cleaning
- Management of local lab reference ranges
- Data encoding (including panels to be encoded, dictionaries and versioning)
- Management of external data
- SAE reconciliation
- Quality control audit(s)



- Data transfers
- Database finalization/lock

Some examples of CDM cost drivers include:

- Number of unique CRFs (paper or electronic)
- Number of total CRF pages (paper or electronic)
- Number of subjects to be enrolled
- Number of cleanup listings
- Number of external data sources (e.g., central labs, electronic diaries, etc.)
- Number of local labs
- Number of queries expected
- Number of terms to be encoded
- Number of SAEs to be reconciled
- Number of data review rounds
- Number/types of data transfers
- Number of unique status reports
- Frequency of status reports
- Frequency of teleconferences
- Number of interim database locks

After relevant cost drivers have been shared with the vendor, the sponsor and vendor should discuss variables that could impact pricing prior to the vendor completing an initial bid. This discussion should include determination of which organization's SOPs will be followed. If the sponsor's SOPs are to be followed, training requirements will need to be determined for the vendor. Both parties should also consider which systems will be used and if any standards or efficiencies can be applied to the project(s). During this phase of

the relationship, clear expectations should be agreed upon and documented. Expectations to be discussed and documented should include:

- Communications (project status updates, escalation path, etc.)
- Quality (documents and data)
- Timelines and turnaround times
- Final deliverables

When working with a CRO, the final bid grid should be shared with the CRO parties in charge of managing the study and study deliverables. Both parties (sponsor and CRO) should review each task on the bid grid, line by line, to confirm understanding of the task and confirm the responsibility and accountability (responsible party, approving party, etc.). Each task should be explained to the CRO in sufficient detail prior to completion of the bid so that both parties fully understand what is to be included and priced. See Appendix C for an abbreviated sample bid grid.

## ***Vendor Oversight***

The vendor's responsibilities should be specified in detail and all deliverables in the contract should be clearly defined. The vendor should have frequent quality reviews to ensure compliance with contracted processes and deliverables. Continuous oversight of the vendor must occur throughout the study and until all final deliverables from the vendor have been accepted. A sponsor's quality management plan should include information regarding vendor oversight, and should be comprehensive enough to include processes and expectations.

Milestones should be based on defined deliverables that are mutually agreed upon throughout the study. These milestones are monitored through regular communication and status updates that are documented and provided to the sponsor. Out-of-scope or unexpected contractual issues should be discussed and handled as they occur, which will help ensure that misunderstandings and problems are minimized. All deficiencies should be analyzed to determine the appropriate preventive or corrective actions.

## **Governance Documents**

Every clinical study should include governance documents that explicitly describe how various processes will be carried out. These governance documents should relate directly to the data management plan (DMP) and be referenced within the DMP. The following are some of the governance documents that may be used within a clinical study, although some of these documents may be consolidated within other study documents (such as within the DMP itself).

- Training and on boarding plans
- Transition plans to cover staffing changes
- Key milestones on study timeline
- Performance metrics definitions
- Communication plan (to include roles and responsibilities)
- Escalation plan

Monitoring adherence to study governance documents may involve formal governance teams, depending on the size, value, or risks of the project or study.

## **Resource Management**

Resource management is a key component of providing sufficient vendor oversight. The sponsor must ensure that the resources needed to provide effective oversight are available. Similarly, both sponsor and vendor must ensure that they continue to possess the resources needed to fulfill their respective contractual obligations. In addition to physical resources (hardware, software, physical space, etc.), resource management should include personnel qualifications and availability, as well as the vendor's ability to ramp resources up or down as needed to fulfill contractual obligations.

## **Study Startup Oversight**

Vendor oversight may vary during different phases of a clinical study. The following are areas requiring vendor oversight that occur during the startup phase of a study.

- Depending on the business model selected, training on sponsor's systems and SOPs, if appropriate—The sponsor must ensure that the vendor's team is thoroughly trained on the sponsor's systems and SOPs where applicable.
- Project-specific training—Although this training should occur for both sponsor and vendor personnel prior to the start of the study, it is not intended to train personnel on the fundamental aspects of a therapeutic area or clinical research in general. The vendor evaluation process should have already demonstrated that vendor personnel have the level of knowledge and experience needed to perform their respective functions. Because both sponsor and vendor personnel should already have expertise in clinical research and the study's therapeutic area, project-specific training should focus on those aspects of the study that make it unique. Project-specific training should follow the Pareto principle, commonly known as the 80/20 rule, meaning that 80% of the training should focus on the 20% of study parameters and processes that are unique to the individual study.
- Retraining—Retraining study-specific processes and parameters should be carefully considered. Retraining necessitated by personnel changes (or other reasons) may occur for both the sponsor and the vendor during the course of the study, and it is crucial that responsibilities for project-specific retraining are determined at study startup. Although not always the case, each party (sponsor and vendor) typically assumes responsibility for retraining their own personnel when staffing changes occur.
- Timeline planning and review—Prior to the official start of the study, study timelines should be reviewed and mutually agreed upon by both sponsor and vendor. Any timeline changes should be agreed upon by both parties and communicated as soon as possible.

- Project milestone tracking tools/metrics—A set of predefined tracking tools and metrics should be established prior to the start of the study. The definition and expected use of these tools and metrics should be mutually agreed upon by the sponsor and vendor. The study’s communication plan should establish how often and through which medium the tools or metrics will be shared. Milestone tracking tools and metrics give better direction to sponsors in determining the correct level of oversight, while also helping vendors measure their own success, which may ultimately relate to bonus payments if stipulated in the contract.

### **Study Conduct Oversight**

After a study has begun, vendor oversight is needed to ensure the vendor continues to provide contracted services in accordance with agreements and timelines established during study startup.

- Timeline management—Both sponsors and vendors benefit from effective oversight and management of study timelines. Ultimately, timeline management should ensure the vendor is meeting sponsor needs and prescribed milestones, but effective timeline management will benefit the vendor as well by providing an assessment of vendor performance throughout the study. Timelines should be clearly documented from study startup until after study closeout, and should specify both major and minor study milestones. Any midstudy adjustments to study timelines should be mutually agreed upon, documented, and version controlled.
- Scope management—Throughout the course of a study, periodic evaluations should reexamine the scope of work being performed by the vendor. Although not always readily apparent, a slowly changing scope of work being performed by the vendor can result in unexpected cost overruns. Periodic scope of work reevaluations can benefit the vendor by ensuring all parties are aware of the amount of work that is being performed, as well as associated costs.
- Key performance indicators—Although established during the study startup phase, key performance indicators should be monitored throughout the study. Meeting or beating performance indicators may have an impact on the level of data quality oversight needed. Performance indicators may also be used to reward or penalize a vendor if such clauses exist within the

contract. For example, if a CRO achieved final database lock ahead of schedule, a monetary bonus may result. Conversely, if a vendor did not complete a data-entry system build by the date specified in the timeline, the sponsor may be entitled to a predetermined discount.

- Determination of meeting or beating performance indicators can be very subjective if not clearly defined. In the event of a discrepancy between achieving and not achieving a key performance indicator, an arbitration may occur between predetermined representatives from both sponsor and vendor organizations. In these situations, best practice is to use arbitration representatives who are separate from the day-to-day operations of the study.
- Data quality oversight—Because of the variations that may exist between quality requirements of different types of data, data quality oversight may be highly individualized for each vendor, and may be individualized within different parts of a single vendor’s scope of work. Data quality oversight may (and probably will) change during the course of a study. Different levels of oversight may be needed for different data elements in a study. For example, more data quality oversight may be needed for primary efficacy and safety variables than for a subject’s routine vital signs.
- Compliance—Sponsors should ensure regulatory compliance is consistently documented, complete, and timely, as compliance records may be audited at any time. Sponsors should also monitor and document strict adherence to all SOPs and ICH guidelines.
- Relationship Governance - It is advisable to have a face-to-face meeting between the sponsor and the vendor at least annually to evaluate the progress of the relationship and to determine the future shared vision.

## **Study Closeout Oversight**

The last stage of the vendor management relationship is the project closeout phase. During study closeout, the sponsor should conduct and document rigorous oversight of all closeout activities and processes. Some of the primary goals of the project closeout phase are to ensure all contractual obligations have been met, finalize the study’s objectives, conduct meetings to discuss lessons learned, complete and archive study records, and celebrate

successes. In terms of CDM, the specific activities associated with project closeout may include the following:

- Conduct database finalization activities (e.g., database lock, decommissioning sites, unblinding, preparing final analysis datasets).
- Conduct lessons learned meeting with all stakeholders to identify successes and challenges to be mindful of for future projects.
- Finalize and archive all relevant communications and data management documents according to applicable SOPs.
- Recognize the team for their contribution.

Continuous vendor oversight throughout the course of a study helps to ensure that study objectives are met effectively and efficiently, and that the vendor is fully aware of expectations. A successful sponsor–vendor relationship is one in which both parties’ business needs are met, and one where both parties would be willing to enter into a similar agreement with each other in the future.

### ***Recommended Standard Operating Procedures***

- Vendor Qualification and Selection
- Specifications for Outsourced Deliverables
- Ongoing Vendor Management
- Vendor Auditing Specific to Types of Vendors
- Vendor Relationship Termination

### ***References***

4. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Section 5.2.2 and 5.2.3; 1996.
5. Getz K, Zuckerman R. Clinical Research Outsourcing: Moving from transactional to strategic partnership-based outsourcing. *Contract Pharma*.

June 2008. Available at <http://www.contractpharma.com/articles/2008/06/clinical-research-outsourcing>. Accessed January 23, 2010.

6. Lucas K. Steps to Functional Service Provider Success. *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2008.

### **Further Reading**

Terms used in this chapter may be found in the *Good Clinical Data Management Practices* Glossary.

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
<b>September 2000</b>	<b>Initial publication.</b>
<b>May 2007</b>	<b>Revised for style, grammar, and clarity. Substance of chapter content unchanged.</b>
<b>July 2008</b>	<b>Revised for content, style, grammar, and clarity.</b>
<b>March 2010</b>	<b>Revised for content, style, grammar, and clarity.</b>



**Appendix A: Sample Request for Information (RFI) Form**

<b>Company Information</b>
1. Provide a brief description of the company's history, including length of time in the industry, origins of the company, mission statement and vision.
2. Provide an organizational chart. Include position and number of employees in each department (senior management, technical support, user support, technical and client service managers, sales and marketing, development, recruiting, quality assurance, training, etc.).
3. Describe quality assurance processes and roles. Is the quality assurance organization independent of the operational organization?
4. Describe the current level of company funding.
5. Describe the company's pricing model.
6. Describe the quality management system adopted by the company.
7. Describe the validation/change control processes of the computerized systems.
8. Describe results of prior audits.
9. Describe quality oversight on contractors (if applicable).
10. If applicable, provide the results of previous regulatory inspections.

<b>Products/Services</b>
1. Describe the evolution of your product or service.
2. How many clients are currently using your product or service?
3. Describe your user support services (IT, helpdesk, IVRS, etc.).
4. Describe the company's interpretation of 21 CFR 11 and how your product is in compliance with this regulation.
5. Can your company produce Clinical Data Interchange Standards Consortium (CDISC) compliant data? If so, which model or models?
6. Describe the company's involvement and specific recommendations for user training. Differentiate between clinical studies with a few sites and those with a large number of sites, if appropriate.
7. What other products or services do you offer?

<b>Experience</b>
1. How many studies has your company supported in the past _____ years?
2. What is the largest clinical study completed to date with respect to number of sites, number of subjects? What lessons did you learn?
3. What are some of the qualities of your company from a human resource perspective? (e.g., What is your rate of turnover? What percentage of your employees are contract versus permanent? What are your training procedures?)
3. What user feedback have you solicited or received from study site personnel or clients about your product or services? How was the feedback addressed?
4. Provide references.
5. Provide CVs and training plans for the proposed personnel.

**Appendix B: Sample Internal Review—Topics for EDC Vendor**

(for more discussion of EDC vendors, please see chapter entitled “EDC—Concepts and Study Start-up”)

Investigator Site	Electronic CRF interface Infrastructure Data entry Programmed validation checks – display and resolution Query resolution and workflow CRF status and workflow CRF design
CRA	Source document verification Query management CRF status and workflow Appropriate access to data
Data Management	CDM review of data Automatic query approval Manual query generation Query management Re-run of validation checks Appropriate access to data Security Change control Coding
External Data Integration	Laboratory Data Interactive Voice Response System (IVRS) Pharmacovigilance database (also known as a safety database—this will only be external in certain situations)
IT	Server requirements Client requirements Architecture Database Security measures
Company	EDC experience Funding Size
Helpdesk	24/7 service to support users Multiple languages Users activation and deactivation

**Appendix C: Sample Bid Grid**

<b>CRO SERVICES</b>	<b>Unit</b>	<b>Cost / Unit</b>	<b>Estimated Number of Units</b>	<b>Item Cost</b>
<b>DATA MANAGEMENT</b>				
Project Management	Month			\$0.00
CRF Creation	Per Unique Page			\$0.00
CRF Guidelines	Per Unique Page			\$0.00
Create Data Management Plan	Plan			\$0.00
Design Database	Per Unique Page			\$0.00
Program Derived Fields	Per Unique Page			\$0.00
Program Data Edit Specifications (to include number of edit checks to be developed)	Per Edit check/Per Unique Page			\$0.00
CRF Tracking	Page			\$0.00
Double Key Data Entry	Page			\$0.00
Query Rate (queries X pages X subjects)	Page			\$0.00
Line Listing Review (for Safety, Sponsor, etc)	Listing			\$0.00
Data Management Review	Page			\$0.00
Data Coding	Code			\$0.00
Provide Coding Dictionaries	Dictionary			\$0.00
SAE Reconciliation	SAE			\$0.00
Lab Normal Maintenance	Lab Site			\$0.00
External Data Loads	Load			\$0.00
QC Audit	Page			\$0.00
Database Lock	Lock			\$0.00
Database Transfer(s)	Transfer			\$0.00
<b>TOTAL</b>				<b>\$0.00</b>

**This page is intentionally blank.**

---

# Data Management Standards in Clinical Research

July 2009

## *Abstract*

Use of standards has become increasingly widespread within clinical data management. Standards can reduce setup costs for a study, reduce conversion errors, and most importantly speed a medical treatment's path to market. This chapter discusses the importance of standards within clinical research, the history of standards used in health care delivery, some of the standards already commonly used, and future directions for standards within clinical research. The chapter also provides readers with an overview of standards relevant to clinical data management. Links are provided for more information about each standard, including downloads for most of the standards.

## *Introduction*

Merriam-Webster defines the word "standard" as "something established by authority, custom, or general consent as a model or example." Within the context of clinical data management (CDM), standards are used to optimize the collection, transport and storage of data, and simplify the submission of data to regulatory bodies.

The advent of modern information technology has enabled widespread use of comprehensive standards. Today, standards encompass almost every part of data collection and handling. Although there are few regulatory mandates for using any particular standard, using standards in all areas of data collection and handling can greatly increase an organization's efficiency by shortening study setup time and incorporating effective and validated standards, thereby reducing overall time and expenses while maintaining consistency for data managers and those charged with collecting data at clinical sites. Most of the established standards currently in use are readily available and designed to be independent of any vendor or platform.

## **Scope**

This chapter provides an overview of established standards commonly used within clinical studies. In addition to giving an overview of each standard and its purpose and scope, the chapter directs readers to where more information can be found about these standards. In most cases, links to the download of the standards discussed are included. Additionally, information is provided about emerging standards within clinical data management. For specific information about implementation of standards listed in this chapter, please follow the provided links to the standards development organizations.

## **Minimum Standards**

- Use the most current version of any standard, if appropriate.
- Use standards required by regulatory agencies in the country where the study is conducted.
- Do not modify published standards.

## **Best Practices**

- Use accepted standards whenever possible, and strive for interoperability.
- Use all standards recommended by regulatory agencies in the locale of the study.
- Review implementation guidelines for any standard having associated guidelines documents.

## **Purpose and Benefits of Standardization**

The use of standards within clinical research involves using standardized names, codes, structures, and formats for data across different locations, studies, and organizations. Using the same formats, names, and codes for different studies can greatly decrease the time and money needed to set up a study, particularly in cases where similar studies have been conducted in the past. Standards provide benefits beyond study setup and can also help streamline processes for study conduct, data transfers, analyses, and

regulatory submissions. Ultimately, standards facilitate bringing safe and effective treatments to patients in a more timely and cost-effective fashion.

Although multiple standards exist for similar concepts, the ultimate goal is for researchers everywhere to use the same standards and naming conventions for their studies. This goal has not yet been realized, but the clinical research industry is trending in that direction. The US Food and Drug Administration (FDA) has strongly encouraged the use of the Study Data Tabulation Model (SDTM) for data submissions, and although this standard's use has not been mandated yet, it may become mandatory in the future. Data submissions in a standardized format allow the FDA and other regulatory bodies to expend fewer resources on their review of study data.

Another enormous benefit to standardization is that data can be more easily and accurately compared and combined across different studies. Although the Internet was originally created to promote sharing of scientific research data, the actual sharing of data has been somewhat limited, in large part due to researchers storing data in different file formats. Standards could potentially increase data sharing, as well as the compatibility of shared data. This increased data sharing could provide valuable benefits to science and humanity.

## ***History of the Development of Standards Organizations***

Before the advent of global communication tools such as the telephone and Internet, standards were typically limited to their locale of origin. As technological advances have sped globalization, organizations have emerged to promote standards for many industries. The following organizations have played integral roles in promoting the standardization of health care data used in clinical research.

### **International Organization for Standardization (ISO)**

ISO was created in 1947 after delegates from 25 countries met to discuss the creation of an international organization to create and maintain international standards for industry. From starting with 67 proposed ISO technical committees in 1947, ISO has developed over 17,000 standards encompassing the full spectrum of industries across the globe. In addition to standards

formulated for specific industries, ISO has created generic standards for product quality and management systems that are applicable to any endeavor.

In addition to general standards applicable to quality and management systems, there are multiple ISO standards specific to various processes involved with clinical research. More information about all ISO standards can be found at <http://www.iso.org/iso/home.htm>.

## **International Conference on Harmonisation (ICH)**

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) began in 1990 as an effort to standardize pharmaceutical regulatory requirements in Europe, Japan, and the US. The ultimate objectives of ICH are to 1) maintain safety and quality while increasing efficiencies in the use of human, animal, and material resources, and 2) help eliminate unnecessary delays in bringing new medical treatments to market. To achieve these goals, numerous guidelines have been released by ICH since its inception. Many of these have had a strong impact on standards development, particularly in regard to regulatory submissions. More information about ICH can be found at <http://www.ich.org>.

## **Health Level 7 (HL7)**

Founded in 1987, HL7 is a nonprofit Standards Development Organization (SDO) initially created to produce standards for hospital information systems. The organization's mission is to provide "...standards for interoperability that can improve care delivery, optimize workflow, reduce ambiguity and enhance knowledge transfer among all of our stakeholders, including healthcare providers, government agencies, the vendor community, fellow SDOs and patients."<sup>1</sup>

The following HL7 standards relate to clinical data management, and are discussed later in this chapter.

- Reference Information Model (RIM)
- Clinical Context Object Workgroup (CCOW)
- Clinical Document Architecture (CDA)

More information about HL7 standards can be found at <http://www.hl7.org>.



## **Clinical Data Interchange Standards Consortium (CDISC)**

Unlike ISO and HL7, CDISC was formed solely to create standards for clinical research data. Their mission statement reads, “CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC website.”<sup>2</sup>

CDISC began in 1997 with a meeting of 25 people interested in standards creation for use within clinical research. Since that time, CDISC has grown exponentially and today has the support of over 200 member organizations from around the world. In addition to consulting with recognized leaders in the clinical research industry, CDISC works closely with other SDOs such as ISO and HL7 to improve interoperability between the various standards. As CDISC has grown in membership and acceptance, their scope has expanded as well. The following standards have been developed by CDISC and are currently published and available for use.

- Clinical Data Acquisition Standards Harmonization (CDASH)
- Laboratory Model (LAB)
- Operational Data Model (ODM)
- Study Data Tabulation Model (SDTM)
- Analysis Dataset Model (ADaM)

For more information about CDISC, visit <http://www.cdisc.org/>.

## ***Standards for Clinical Research***

The standards discussed in the remainder of this chapter are primarily those relating directly to CDM functions within clinical studies. The majority of the standards discussed come from CDISC, but CDM personnel should be aware of any new standards gaining traction within the industry, regardless of the

origin of the standard. Where possible, all standards employed by an organization should be sufficiently interoperable to allow for a comprehensive standard practice to effectively manage clinical data.

## ***Clinical Data Acquisition Standards Harmonization (CDASH)***

The Clinical Data Acquisition Standards Harmonization (CDASH) standard released October 2008 by CDISC, is intended to streamline and standardize data collection at clinical investigative sites. The development of CDASH was a global effort, with feedback provided from all three of the ICH regions (US, Europe, and Japan). The published CDASH standard consists of a basic set of data collection fields (variable name, definition, metadata) that apply to the majority of case report forms (CRFs), regardless of therapeutic area or phase of development. Sponsors are expected to make additions for therapeutic area-specific data collection fields, as well as other data collection fields needed for regulatory requirements. The CDASH standard also includes best practice guidelines, regulatory references, and information about the development of the CDASH standard.

In order to ensure harmonization between standards, recommendations are provided for mapping CDASH data collection fields (or variables) into the Study Data Tabulation Model (SDTM) submission structure.

### **CDASH Domains**

The data collection fields, specified in CDASH, like SDTM, are divided into the following sixteen domains along with their associated codes.

- Adverse Events (AE)
- Comments (CO)
- Concomitant Medications (CM)
- Demography (DM)
- Disposition (DS)
- Drug Accountability (DA)
- ECG Test Results (EG)

- Exposure (EX)
- Inclusion/Exclusion (IE)
- Laboratory Test Results (LB)
- Medical History (MH)
- Physical Examination (PE)
- Protocol Deviations (DV)
- Subject Characteristics (SC)
- Substance Use (SU)
- Vital Signs (VS)

An implementation guide is under development to accompany the standard, and is targeted for completion in the third quarter of 2009. Please see <http://www.cdisc.org/standards/index.html> for more information about the CDASH standard, including a link to download the most recent version of the standard.

### ***Laboratory Model (LAB)***

The CDISC LAB standard was initially released in 2002, and was designed to be a standard for the transfer of laboratory data. Other standards already existed for laboratory data, but those standards had limited applicability to clinical research. Use of the LAB standard is estimated to save 30% to 50% of laboratory costs, which has an enormous impact on overall costs considering that 60% to 80% of clinical data is estimated to come from laboratories.<sup>3</sup>

### **Data Field Levels**

Data for this standard are categorized into the following 12 levels and associated data fields.

- Good transmission practice—version of LAB model used, local (and universal) date and time data file was created, identification code and name of organization that is the source of the data transmission

- Study—identification code and name of the study and whether the data transmission is incremental or cumulative
- Site—identification code of the site
- Investigator—identification code and name of the investigator
- Subject—identification code of the subject before and after randomization (and possibly an extra subject identifier code), subject initials, subject gender (and possibly gender code), subject date of birth, subject race (and possibly race code). Note: When collecting subject identification data, follow local regulations relating to subject privacy
- Visit—name of the visit, identification code or number of the visit, whether the visit was scheduled or unscheduled, and whether the visit was physician ordered, a retest, or early termination of the subject's involvement with the study
- Accession—name and identification code of the laboratory delivering the data, identification code of the kit used for the subject visit, local (and universal) date and time of the last modification made to the record
- Record extension type—specifies if any extension to the base LAB model was used, as described below
- Base specimen—identification code of an individual kit item used at the visit, actual and planned local date and time of specimen collection from the site, time discrepancy between planned and actual specimen collection, local (and universal) date and time of specimen receipt at laboratory, specimen condition, laboratory and investigator comments, specimen identification code, specimen name (e.g., blood, urine), subject age at collection, units of subject age at collection, fasting status of subject at collection
- Base battery—name and identification code of the battery, panel or group to which the test belongs
- Base test—name and identification code of laboratory, name and identification code of the test as defined by site, name and identification code of the test as defined by laboratory, LOINC (Logical Observation Identifiers Names and Codes) code and code list identifier, test status

(done, not performed or cancelled), test comments, local (and universal) date and time of testing, test type (study test, non-study test, unscheduled study test)

- Base results—this level contains 32 fields providing all test result names, codes, reference ranges, units, results, statuses, toxicity grades, flags, reporting time, and record type

## Extensions

In addition to the LAB base model, the standard has several extensions designed for specialized laboratory data. The extensions currently published or in development include:

- Microbiology
- Pharmacogenomics
- Electrocardiogram (ECG) interpretation
- Specimen handling
- Edit/data query capabilities

Please see <http://www.cdisc.org/standards/index.html> for more information about the LAB standard, including a link to download the most recent version of the standard.

## ***Operational Data Model (ODM)***

The first ODM standard was released by CDISC in 2002 to address the structure of data rather than naming conventions. The ODM standard is designed to “...support the end-to-end data flow within clinical trials, from the operational database through analysis to regulatory submission. The role of the ODM is to facilitate the movement of clinical data collected from multiple acquisition sources to an operational database, but it also has application in the subsequent exchange and archiving of such data.”<sup>4</sup> In addition to providing a standard format for transporting data, the flexibility of the ODM creates the possibility of automating creation of electronic CRFs used in an electronic data capture (EDC) system.<sup>5</sup>

The ODM uses the extensible markup language (XML) to create a file with the four following primary elements.

- Study information such as study name and metadata
- Administrative information such as users, sites, and authorizations for the study
- Reference data (e.g., normal ranges)
- Clinical data from the study <sup>4</sup>

### **Supported Data Formats**

The ODM was designed to be vendor-neutral and platform-independent, and supports numerous data formats including integers, decimals, text strings, Boolean terms, hex binary, base 64 binary, dates and times, partial dates and times, intervals, durations, and more.

Please see <http://www.cdisc.org/standards/index.html> for more information about the ODM standard, including a link to download the most recent version of the standard.

### ***Study Data Tabulation Model (SDTM)***

The first implementation-ready version of the SDTM was released by CDISC in 2004, and was developed to provide a standard for the organization, structure, and format of tabulation data to be submitted to regulatory agencies. Tabulation datasets contain collected data from a clinical study, and should not be handled in the same manner as the other three types of data submitted to regulatory agencies (e.g., analysis datasets, patient profiles, and listings). The FDA has strongly recommended using SDTM for data tabulation submissions, but this has not been mandated.

### **Variable Classification Scheme**

According to the SDTM, each variable, which normally corresponds to a column in a dataset, can be classified according to its *Role*. A Role determines the type of information conveyed by the variable in describing an observation. Variables can be classified into five major roles:

- *Identifier* variables—identify the study, the subject (individual human or animal) involved in the study, the domain, and the sequence number of the record.
- *Topic* variables—specify the focus of the observation (such as the name of a lab test), and vary according to the type of observation.
- *Timing* variables—describe the timing of an observation (such as start date and end date).
- *Qualifier* variables\*—include additional illustrative text, or numeric values that describe the results or additional traits of the observation (such as units or descriptive adjectives). The list of Qualifier variables included with a domain will vary considerably depending on the type of observation and the specific domain.
- *Rule* variables—express an algorithm or executable method to define start, end, or looping conditions in the Trial Design model.<sup>6</sup>

\*The SDTM further divides qualifier variables into five subclasses of grouping qualifiers, result qualifiers, synonym qualifiers, record qualifiers, and variable qualifiers. See the SDTM implementation guide for detailed descriptions of these qualifier variables.

## Standard Domains

The SDTM contains the following domains and respective codes, which fall into six general categories.

- Special Purpose Domains
  - Demographics (DM)
  - Comments (CO)
  - Subject Elements (SE)
  - Subject Visits (SV)
- Interventions

- Concomitant Medications (CM)
- Exposure (EX)
- Substance Use (SU)
- Events
  - Adverse Events (AE)
  - Disposition (DS)
  - Medical History (MH)
  - Protocol Deviations (DV)
  - Clinical Events (CE)
- Findings
  - ECG Test Results (EG)
  - Inclusion/Exclusion Criterion Not Met (IE)
  - Laboratory Test Results (LB)
  - Physical Examinations (PE)
  - Questionnaires (QS)
  - Subject Characteristics (SC)
  - Vital Signs (VS)
  - Drug Accountability (DA)
  - Microbiology Specimen (MB)
  - Microbiology Susceptibility Test (MS)
  - Pharmacokinetic Concentrations (PC)
  - Pharmacokinetic Parameters (PP)
  - Findings About (FA)



- Trial Design Domains
  - Trial Arms (TA)
  - Trial Elements (TE)
  - Trial Visits (TV)
  - Trial Inclusion/Exclusion Criteria (TI)
  - Trial Summary (TS)
- Special Purpose Relationship Datasets
  - Supplemental Qualifiers (SUPPQUAL)
  - Related Records (RELREC)

### **SDTM Implementation Guide (SDTMIG)**

CDISC has also released an implementation guide to augment the SDTM standard. This implementation guide is intended to guide the format, organization, and structure of tabulation datasets. Any organization using SDTM should also utilize this implementation guide.

See <http://www.cdisc.org/standards/index.html> for more information about the SDTM standard and implementation guide, as well as a link to download the most recent version of the standard and implementation guide.

### ***Analysis Dataset Model (ADaM)***

ADaM was initially released by CDISC in 2004 as a standard model to create analysis datasets for submission to regulatory bodies, and can be thought of as an extension to the SDTM standard. The ADaM describes the proposed content, structure, and metadata of analysis datasets, including analysis dataset metadata, analysis variable metadata, and analysis results metadata. The standard includes examples of datasets created using the ADaM.

### **Four Key Principles for Analysis Datasets**

The ADaM standard is based on the following four general principles.

- Analysis datasets should facilitate clear and unambiguous communication
- Analysis datasets should be useable by currently available software applications
- Analysis datasets should be linked to machine-readable metadata
- Analysis datasets should be analysis-ready

### **ADaM Implementation Guide (ADaMIG)**

As with the SDTM standard, CDISC has released a draft implementation guide to augment the ADaM standard. This implementation guide is intended to guide the format, organization, and structure of analysis datasets. Any organization using ADaM should also utilize this implantation guide.

See <http://www.cdisc.org/standards/index.html> for more information about the ADaM standard and implementation guide, as well as a link to download the most recent version of the standard and implementation guide.

### ***Electronic Common Technical Document (eCTD)***

The eCTD standard was developed by the ICH to provide a standardized format for submitting files from pharmaceutical studies to regulatory bodies. Unlike some standards used in clinical research, eCTD focuses more on data and file structures than naming conventions. The eCTD relies heavily on the Document Type Definition (DTD) specification of the XML markup language. These DTDs are used to create a detailed hierarchical folder structure for each eCTD.

In addition to the structure of an eCTD, the standard is designed to support high-level functional requirements. Some of these functional requirements include the ability to copy and paste, view and print documents, have annotated documentation, and export to databases. An eCTD should also allow users to search both within and across applications and allow navigation throughout the eCTD and any subsequent amendments or variations.<sup>7</sup>

## **eCTD Modules**

Every eCTD consists of five modules, four of which are common to all countries and regions. The first of the following five modules may vary between different ICH regions.

1. **Regional Administrative Information and Prescribing Information**—Module One contains administrative information and forms that may vary between countries and regions.
2. **Common Technical Document Summaries**—Module Two contains summaries of the information contained in Modules Three, Four, and Five.
3. **Quality**—Module Three provides detailed information about the treatment being studied and details of the product's development and manufacturing processes.
4. **Nonclinical Study Reports**—Module Four provides detailed pharmacological, pharmacokinetic and toxicological information.
5. **Clinical Study Reports**—Module Five contains the results of the study, including data related to background and development rationale, efficacy, safety, benefits and risks.

The eCTD has become the recommended format for regulatory submissions in the European Union, US, Canada and Japan, and may become mandatory in time. Many companies sell eCTD submission solutions, but more free information about eCTD can be found at <http://www.fda.gov/cder/Regulatory/ersr/ectd.htm>.

## **HL7 Standards**

Although HL7 does not design standards specifically for use within clinical research, the increased use of electronic health records within hospitals gives CDM personnel an incentive to become familiar with the following HL7 standards.

- Reference Information Model (RIM)—This standard provides structure, naming and coding conventions to be used among disparate organizations and platforms.
- Clinical Context Object Workgroup (CCOW)—This is a vendor-independent standard designed to enable different computer applications to communicate with each other effectively.
- Clinical Document Architecture (CDA)—This standard is based on the RIM, and uses the XML markup language to specify the coding, structure, and semantics of clinical documents to be exchanged.

See <http://www.hl7.org/> for more information about HL7 standards, including links to downloads of the most recent versions of the standards.

## ***Future Directions***

With the numerous standards that currently exist, the ultimate goal is to make these standards interoperable to the degree that any health-related data can be easily shared between different researchers and institutions. The US National Cancer Institute (NCI), FDA, HL7, and CDISC are all collaborating to create the Biomedical Research Integrated Domain Group (BRIDG) model. This standard is being designed to integrate HL7 and CDISC standards, which will reduce potential errors and streamline the flow of data from health care providers to clinical researchers.

CDISC is also creating a standard called the Protocol Representation Model (PRM), which identifies, defines and describes over 300 common protocol elements and maps those elements to elements within the BRIDG model. The PRM model is intended as a standard to be used in designing a study, selecting investigative sites, developing data collection tools, and describing an analysis plan and study procedures.

The FDA is piloting a program for a standard known as the Summary Technical Document (STED), which is a harmonized format for medical device regulatory submissions that is already accepted by multiple regulatory bodies worldwide. For information about the pilot program, see <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/SummaryTechnicalDocumentSTEDPilotProgram/default.htm>. The FDA has already released eSubmitter, a

standardized tool that is part of an electronic submissions program originated in the Center for Devices and Radiological Health (CDRH). The eSubmitter program evolved from two very successful pilot programs (eLaser and Turbo 510(k)) at CDRH. FDA eSubmitter is an improved and expanded package for a variety of submission types and is now available for voluntary use by sponsors and manufacturers in certain device and radiological health and blood regulated industries. Like other attempts to standardize, the goal is to improve efficiencies in the regulatory submission and review process. See <http://www.fda.gov/ForIndustry/FDAeSubmitter/default.htm> for more information.

Another interesting and evolving initiative is the National Cancer Institute's cancer Biomedical Informatics Grid (caBIG<sup>®</sup>), which is intended to simplify collaboration by leveraging shared expertise and large multidisciplinary data collections to speed many of the processes of cancer research. The four key principles of caBIG<sup>®</sup>—open access, open development, open source, and federation—have guided the development of interoperable software tools, data standards, and a computing infrastructure conceived to advance basic and clinical research. Originally designed solely for cancer research, the caBIG<sup>®</sup> initiative may expand outside cancer research to serve as a model for improving collaboration, data sharing, and patient outcomes in other therapeutic areas in the future. For more information about caBIG<sup>®</sup>, see <https://cabig.nci.nih.gov>.

Standards across medical research are contributing to more efficient research activities. This success has spawned a mounting interest in standards development and resulted in an increasing number of new and revised standards. Staying abreast of standards that affect CDM is a challenge. One effective strategy is to visit the Web sites of organizations that have been involved in standards development to keep informed of their progress.

### ***Recommended Standard Operating Procedures***

- CRF Design
- Database Design
- Medical Coding

- Data Transfers
- Regulatory Submissions

## References

1. HL7 (Health Level 7) Web site. Available at: <http://www.hl7.org/>. Accessed November 24, 2008.
2. CDISC (Clinical Data Interchange Standards Consortium) Web site. Available at: <http://www.cdisc.org/about/index.html>. Accessed November 24, 2008.
3. Bassion S. The CDISC Lab Model: Saving Users Time and Money. *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2005.
4. Ibersen-Hurst D. The CDISC Operational Data Model: Ready to Roll? *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2004.
5. Aerts J. CDISC Standards Energize EDC. *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2006.
6. CDISC Submissions Data Standards Team. *Study Data Tabulation Model: Version 1.1*. Austin, TX: Clinical Data Interchange Standards Consortium; 2005. Available at <http://www.cdisc.org/models/sdtm/v1.1/index.html>. Accessed November 24, 2008.
7. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. *ICH M2 EWG Electronic Common Technical Document Specification*, V. 3.2, Feb 2004.

## Further Reading

CDISC Analysis Dataset Modeling Team. *Analysis Data Model: Version 2.0*. Austin, TX: Clinical Data Interchange Standards Consortium; 2006. Available at <http://www.cdisc.org/models/adam/V2.0/index.html>. Accessed November 24, 2008.

CDISC CDASH Core and Domain Teams. *Clinical Data Acquisition Standards Harmonization (CDASH)*. Austin, TX. Clinical Data Interchange Standards Consortium; 2008. Available at <http://www.cdisc.org/standards/cdash/index.html>. Accessed November 24, 2008.

CDISC Operational Data Modeling Team. *Specification for the Operational Data Model (ODM) Version 1.3*. Austin, TX. Clinical Data Interchange Standards Consortium; 2006. Available at <http://www.cdisc.org/models/odm/v1.3/index.html>. Accessed November 24, 2008.

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
July 2009	Initial publication.

**This page is intentionally blank.**



---

# Design and Development of Data Collection Instruments

October 2010

## *Abstract*

Clinical data can be collected with a variety of tools, but case report forms are the most frequently used data collection tool. Case report forms may be paper based or electronic and include data entry forms used by patients as well as health care providers. This chapter provides guidelines for the design of case report forms, emphasizing accurate, consistent and logical data collection in accordance with a study's protocol. The design and development processes discussed highlight the importance of a case report form's clarity and ease of use. The chapter also discusses referential questions, redundancies, edit checks, standards, case report form completion guidelines, and distinctions for studies using paper CRFs, electronic data capture and/or patient-reported outcomes.

## *Introduction*

Although the study protocol is arguably the most important document used during a clinical study, case report forms (CRFs) are of vital importance as well. Because CRFs are the most frequently used tools for data collection, great care must be given to ensuring each CRF accurately and consistently captures data specified in the study protocol. An informative and well-structured CRF simplifies database design and data validation processes as well as manipulation of data during statistical analysis. The quality of study data relies first and foremost on the quality of the tool used to collect the data. If the data points specified in the protocol are not accurately collected, a meaningful analysis of the study's outcome will not be possible. Therefore, the design, development, and quality assurance processes of a CRF must receive the utmost attention.

The International Conference on Harmonisation's *Guidance for Industry: E6 Good Clinical Practice* defines the term "case report form" as, "A printed, optical, or electronic document designed to record all of the protocol-required

information to be reported to the sponsor on each trial subject.”<sup>1</sup> This chapter discusses considerations for CRF design, development, and quality assurance, including distinctions for studies using paper CRF, electronic data capture (EDC) and/or patient-reported outcomes (PRO). Because CRFs are related to numerous aspects of clinical data management (CDM), references are provided to other chapters of *Good Clinical Data Management Practices* (GCDMP) that provide more in-depth information in certain areas.

## **Scope**

This chapter focuses on the design and development of CRFs used to acquire clinical data. Consideration is given to topics including questions with dependent relationships (referential questions), redundancies, edit checks, standards, CRF completion guidelines, and distinctions for studies using paper CRF, EDC and/or PRO. For information about laboratory data and data acquisition through external data transfers, see the GCDMP chapters entitled “External Data Transfers” and “Laboratory Data Handling.” For more detailed information about EDC, see the GCDMP chapters entitled “Electronic Data Capture—Concepts and Study Start-up”, “Electronic Data Capture—Study Conduct,” and “Electronic Data Capture—Study Closeout.” For more detailed information about different collection methods for PRO data, see the GCDMP chapter entitled “Patient-Reported Outcomes.”

Although some of the specific topics addressed by this chapter may not be the direct responsibility of CDM personnel, data managers must have an ongoing awareness of requirements and ensure these tasks have been completed in accordance with the principles and standards of their organization, regulatory bodies, and good clinical practice.

## **Minimum Standards**

- Design CRFs to collect the data specified by the protocol.
- Document the process for CRF design, development, approval, and version control.
- Document training of clinical site personnel on the protocol, CRF completion instructions and data submittal procedures prior to subject enrollment.

- Verify CRFs based on rating instruments created by an independent source (e.g. Health Status Questionnaire, Beck Depression Inventory, etc.), have been properly licensed for use and follow prescribed formatting or copyright requirements.
- Ensure CRFs are available at the clinical site prior to enrollment of subjects.

### ***Best Practices***

- Establish and maintain a library of standard forms and associated edit checks (CRFs, CRF completion guidelines, subject diaries, etc.).
- Use a multidisciplinary team to provide input into the CRF design and review processes. Data entry personnel, biostatisticians, the internal study team, and clinical operations personnel may be able to provide valuable perspectives to help optimize CRFs.
- Design CRFs with safety and efficacy endpoints in mind. Consult the protocol, study biostatistician(s) or review the statistical analysis plan (SAP) (if available) to ensure all key endpoints are collected.
- Keep the CRF's questions, prompts, and instructions clear, concise and conformant to CDISC CDASH standards, where possible.
- Design the CRF to follow the data flow from the perspective of the person completing it, taking into account the flow of study procedures.
- Whenever possible, avoid referential and redundant data points within the CRF. If redundant data collection is used to assess data validity, the measurements should be obtained through independent means.
- Use carbonless copy paper (NCR) paper or other means to ensure exact replicas of paper collection tools.

### ***Design and Development Processes***

As with most aspects of clinical research, best results can be achieved through a multidisciplinary approach to designing and developing CRFs. Input from

CDM, statistical, clinical, safety monitoring and regulatory personnel will help ensure the data collected with CRFs meet the needs of the study from all pertinent perspectives. This collaborative approach also allows more thorough consideration of what data should be collected and how the data will be used to meet study objectives.

To ensure the protocol specifies data collection strategies that are reasonable and achievable, CRF design should be taken into consideration before the protocol is finalized.<sup>2</sup> However, this may not be possible for a contract research organization (CRO) that has been contracted to develop CRFs. The process of CRF development may make apparent that certain data points are not as easy to quantify as originally anticipated. If CRFs are developed after the protocol has been finalized, any data points found to be undesirable or unattainable may require a protocol amendment to correct. When the protocol and CRFs are designed concurrently, the quality of both the protocol and the CRFs can be improved through continuous collaboration and feedback.

Although collection of data specified by the protocol is the main impetus of CRF development, care should also be taken to ensure CRFs do not collect data that ultimately will not be used for analysis or will not support analyzed data. Extraneous data can adversely affect overall data quality by drawing the attention of site personnel away from key variables.<sup>3</sup> Key variables are typically those that measure safety parameters or study efficacy endpoints. These key variables should be defined before or during CRF development to ensure they are captured on study CRFs.

All CRFs should contain certain specific elements. All data must be attributable to a subject; therefore each CRF should accurately link the data to the correct subject. Each section that can be separated or viewed separately must contain sufficient identifiers to uniquely identify the data contained in the section. CRFs based on rating instruments created by an independent source (e.g., Health Status Questionnaire, Beck Depression Inventory, etc.), may require special licensing agreements be in place prior to use and that prescribed formats be used or specific copyright information appear on the CRF. All CRFs should also contain a provision for investigator signature to allow timely documentation of the investigator's review of the data as represented and in the event data are subsequently changed.

Data collected on CRFs will ultimately be consolidated for statistical analysis, therefore using standard data structures will help facilitate this integration.

Although the clinical database(s) will impart the structure of the dataset(s), collecting data on forms that promote a common structure will avoid the need for mapping or conversion at a later time. To facilitate this continuity, some organizations have standardized protocol templates, CRFs, database structures, validation procedures, and reporting tables.

## ***Clarity and Ease of Use***

CRF completion is subject to human error. Improving a CRF's ease of use and clarity will result in improving the quality of data collected in the CRF. A number of factors contribute to ensuring a CRF is easily understood and used.

These factors include, but are not limited to:

- CRF layout,
- wording,
- coding,
- use of minimal referential questions,
- minimized redundancies, and
- consideration of distinctions between different collection strategies (such as paper-based CRFs versus EDC-based CRFs versus PRO).

In addition to the need for a CRF to be easily understood by those completing the CRF, the data collected on a CRF should be easily understood as well. Therefore, all questions on a CRF should be carefully examined to determine if the resultant data could potentially be ambiguous.

For example, if possible symptoms are listed with instructions to check all that apply, all check boxes that remain unchecked could be interpreted in two ways: either no symptoms were present or the individual completing the CRF skipped this section. If each symptom is accompanied by two check boxes for the responses "Present" and "Not Present," the potential for ambiguity is removed. Similarly, many questions can have potential ambiguity removed by adding response options for "Not Applicable" or "Unknown."<sup>4</sup>

## **Layout**

A CRF's data fields should be arranged in a manner that is clear and easy to follow. Data that are logically related should be grouped together whenever possible, taking into account any limitations or constraints of the clinical data management system (CDMS) that will be used. Multiple choice answers are a better alternative to free text fields, but if free text fields are used, make certain that fields provide sufficient space to record the information intended for the field.

Throughout all CRFs used in a study, maintain consistency in the order of similar answer choices. For example, the placement of "None," "Not Applicable," or "Other" within a series of choices should not change throughout the CRFs. Similarly, all questions with answer choices of "Yes" and "No" should present these two answer options in the same order. All questions should indicate whether multiple choices can be selected (i.e. check all that apply) or if a question can only have a single answer choice (i.e. check only one).

When designing a CRF layout, format it consistently, including font size and the use of color (if used), and take into account the intended use of the form. The flow of a CRF should closely follow the flow of data from the perspective of the person completing the form. For example, CRFs completed by site personnel might look quite different from those completed by subjects. If a CRF is completed based on information from source documentation (e.g., a medical record) the CRF should be organized in a similar sequence as would appear in the source documentation to facilitate easy transcription of information. If a CRF is to be completed by each subject every three months, a separate CRF should be provided and labeled for each interval to minimize the potential for redundant or ambiguous data.

## **Wording**

All questions and prompts should be concise, specific, and clear enough to ensure that complete and comparable data are obtained from the various people (subject, site personnel, etc.) using a set of CRFs. Always avoid leading questions, and where possible, phrase questions in the positive to avoid the potential confusion that negatively stated questions can cause. For example, use "Did the subject follow the instructions?" rather than "Did the subject fail to follow the instructions?"

Where possible, questions should solicit data that are directly measurable, rather than soliciting interpretations of measurable data. For example, the question “Did the subject have hypertension?” is better posed by asking for the blood pressure range, length of time sustained, or specific interventions performed for the condition.

Once again take into account the intended use of the form from the perspective of the person completing it (i.e. site personnel versus subject).

### **Coded Responses**

Because a large percentage of data must be coded prior to analysis or reporting, data should be collected in a coded format whenever possible. Examples of coded formats include multiple-choice pick lists and yes/no check boxes, where each of the possible responses may be associated with a specific code. Careful use of coded formats can provide for multiple responses where needed, track the total number of responses, and simultaneously encourage the individual completing the form to select at least one response. In cases where possible responses are known, responses can be conveniently structured as a pick list and can be coded without biasing the distribution of responses.

Ideally, CRFs should be designed such that site personnel complete the CRF by selecting, checking or ticking responses. Site personnel will typically be in the best situation to pick the correct assignment because of the availability of source documents and the familiarity of these personnel with each subject. This approach minimizes errors and reduces data processing time. With the possible exception of providing details about safety issues such as adverse events, free text is rarely useful.

### **Referential Questions**

Referential questions are those where the answer (or lack of an answer) to one or more questions is contingent upon the answer to another question. An example of this would be: “Does the subject have child bearing potential? If yes, did the subject agree to use acceptable contraception throughout study duration?”

These types of questions set up a dependent relationship that requires both levels to be completed correctly. Because of this relationship between levels, referential questions can lead to problems during CRF design and maintenance. For example, during CRF revision, one level of a question may be deleted while the other level remains.

Referential questions can also be associated with challenges to proper CRF completion. If instructions are not explicitly clear, subjects or site personnel may not answer all levels of a set of referential questions, leading to unnecessary queries. To minimize potential confusion, referential questions should only be used after careful consideration. Instructions should note where to skip to, not what to skip. They should also be clearly grouped together, apart from other questions or prompts. Referential questions should not refer to another question contained in a remote section of the CRF packet.

### **Minimizing Redundancy**

Data based on the same measurement should not be collected more than once or in more than one place. Doing so creates unnecessary work for site personnel and creates a need to check for consistency between redundant data points, resulting in increased work for clinical and data management teams. Because of the potential for inconsistencies and errors resulting from scores calculated by different parties at different times, collecting raw data is typically preferable to collecting calculated values.<sup>2</sup> Raw data are also easier to verify from source documents. For example, a CRF should not have site personnel calculate the BMI (body mass index) since this can be computed more efficiently by the statistician at the time of analysis based on the recorded height and weight responses. The CRF should also allow the site to record data in their customary units of measure (e.g. inches, centimeters, pounds, kilograms) per their normal practice, which can then be converted, if necessary, by the data management team in the edit check specifications or the statistician at the time of data review/analysis.

Situations do exist where redundant data collection is used to assess data validity, particularly in cases where other means are not practical.<sup>5</sup> If redundant data collection is used to assess data validity, the measurements should be obtained through independent means. For example, two pregnancy tests may be administered during the same visit but on different types of



samples (i.e., serum and urine). If both tests produce the same results, the data can be considered valid.

Some data, such as adverse events or concomitant medications, may be collected via logs rather than individual CRF forms, in which case the elimination of redundant data collection should be carefully considered.

## **Paper-Based Distinctions**

If a paper CRF is poorly designed, organized, or printed, there is a greater potential for missing data due to questions being overlooked. Avoiding certain pitfalls can greatly reduce the odds of questions being overlooked. For example, all printed CRF pages should be single sided and should use a clearly legible font size. Trying to squeeze too many questions onto a single page can lead to questions being overlooked, because the page may become too crowded for the eye to easily discern different items. In part because copies and faxes can be less legible and can cut off part of a page, data should only be recorded on original CRFs.

Paper CRFs should also contain certain design elements. For example, each CRF page should contain both the page number and the total number of pages in the CRF module or packet, which will reduce the likelihood of a page being overlooked. Each CRF page should also be clearly linked to the correct site, subject, visit and follow-up interval (if applicable).

Where dates are requested on a paper CRF, the proper date format (e.g., mm/dd/yyyy, dd/mm/yy) should be clearly stated, especially in studies that span multiple countries or geographic regions. However, dates ideally should be formatted according to the CDASH standard of using a 3-letter abbreviation for the month, which avoids the potential confusion of inconsistent date formats (dd/mmm/yyyy). It is also important to consider how partial dates should be entered if the exact date is not known. If times are requested they should be recorded using the 24-hour clock (HH:MM). Unit of measure (e.g., kilograms or pounds, centimeters or inches) should also be clearly identified.

## **EDC Distinctions**

EDC systems use electronic CRFs (eCRFs), which may offer functionality that helps to avoid potential problems that can occur with paper CRFs. For example, an electronic CRF can enable dates to be chosen from a pop-up calendar, avoiding the potential for entering inconsistent date formats. Care should be taken; however, that if a pop-up calendar is used to enter dates, there remains a method to enter a partial date if the exact date is not known. Electronic CRFs can also group multiple pages into a set for a single subject in such a way that a subject and/or site identifier need only be entered once for the module, therefore avoiding potential errors associated with inconsistent subject/site ID records. System edit checks programmed within the EDC application validate the data at the point of entry and sometimes provide instant feedback to the person entering the data, giving an opportunity to correct the error(s) right away. Paper CRFs, on the other hand, silently accept the error until it is caught by the clinical monitor or the data manager.

However, electronic CRFs must take certain factors into account that do not apply to paper CRFs. For example, electronic CRFs should be thoroughly validated to ensure they function as intended and meet regulatory guidelines.<sup>6,7</sup>

Referential questions that create difficulties when designing paper CRFs can sometimes be addressed with the use of dynamic forms in electronic CRFs. Some EDC applications allow the form(s) to be added dynamically through a script or an edit check. For example, a pregnancy form will not appear unless gender is reported as female on a demographics form.

Electronic CRFs offer the capability to tab through fields in a prescribed sequence, which can help minimize the chances of a question being overlooked. For more information about electronic CRF design, see the GCDMP chapter entitled “Electronic Data Capture—Concepts and Study Start-up.”

## **Patient-Reported Outcomes Distinctions**

Information that is directly reported by subjects is known as patient-reported outcomes (PRO). This type of data is crucial to studies that attempt to quantify subjects’ subjective experiences such as pain intensity or quality of life using rating scales and questionnaires. Because these data are recorded by subjects

themselves rather than trained site personnel, the tools used to collect these data may differ from CRFs intended for completion by study personnel.

Because study subjects will not undergo the same rigorous training as site personnel, the wording of questions and instructions on a CRF collecting PRO data should be clear and understandable to the subject population. These CRFs should avoid the use of any terminology that might be considered jargon common to the clinical research industry.

Some PRO data may be collected on a CRF that is based on a rating instrument created by an independent source (e.g., Health Status Questionnaire, Beck Depression Inventory, etc.), in which case the validity of that instrument must be maintained. If any changes in content or format are necessary, the independent source should be consulted to ensure that the validity of the tool has not been compromised by the changes. Maintain documentation of all changes and the continued validity of the tool. Also, confirm that all necessary licensing and copyright requirements have been satisfied.

Paper CRFs can be used to collect PRO data, but PRO data can also be collected with a variety of electronic tools, commonly referred to as ePRO. For more information about PRO data collection, including considerations specific to use of paper-based PRO or ePRO, see the GCDMP chapter entitled “Patient-Reported Outcomes.”

## ***Edit Checks***

Regardless of how well CRFs are designed, edit checks should be programmed into the database or clinical data management system (CDMS). Edit checks are intended to ensure data integrity and improve data quality by bringing attention to data that are out of the expected range, inconsistent, illogical or discrepant. When data meet the predefined criteria of an edit check, a flag or warning notifies CDM personnel that the data point should be carefully examined to ensure the accuracy of the data point.

Although the majority of edit checks do not differ between paper-based and EDC studies, there are some distinctions in edit checks between the two data collection modalities. For example, edit checks for paper-based studies tend to

focus more on potential transcription errors. For more information about edit checks, see the GCDMP chapter entitled “Edit Check Design Principles.”

## ***Review and Quality Control Processes***

Before being used to collect study data, all CRFs should undergo a quality control review. As with the design process of a CRF, the review process should include input from a variety of sources. First and foremost, CRFs should be examined in conjunction with the protocol to ensure all protocol-specified data are captured. In addition to the various personnel groups that may be involved in CRF design (e.g., statistical, clinical, safety monitoring, regulatory), certain types of CRFs (e.g., translations) may require specialized input into the quality control review.

- CRFs translated into multiple languages (including Braille for the visually impaired) should be carefully reviewed to ensure the translations are truly equivalent. One method to ensure equivalency would be for one party to translate the CRF to the target language and then a second party translate back to the source language and compare the results to the original document.
- CRFs collecting PRO data based on an independent rating instrument may need to be reviewed by the source of the rating instrument, especially if any modifications are made or the instrument is translated into a different language.
- Paper CRFs should be carefully reviewed prior to printing by preparing a prototype using the paper size that will be used for printing (standard paper sizes vary by region, so notebooks, file folders, or other means for housing, filing, faxing or copying the forms should be considered). Upon completion of the printing process, paper CRFs should be examined to ensure acceptable quality of the printed forms prior to releasing the forms to the sites.
- Electronic CRFs should undergo user acceptance testing (UAT) to ensure the CRFs meet the needs of the users who will be entering data. The team performing the UAT should consist of the database developer, clinical research associate, data manager and/or data entry personnel.

- Electronic CRF review may require input from data managers, programmers, or other information technology personnel to ensure the CRFs are properly validated. For more information about validation of electronic CRFs, see the GCDMP chapter entitled “Database Validation, Programming and Standards.”

## ***Standards in CRF Design***

Use of standards can greatly decrease both the cost and time of CRF development. Some organizations create and maintain a library of standard CRF templates and associated edit checks, allowing CRFs to be easily modified to meet the needs of each individual study. Apart from organization standardized CRFs, standards that might impact CRF design come from various sources.

- Regulatory standards may have an impact on CRF design, particularly in regard to data privacy or CRFs that are translated into multiple languages.
- Software platform-specific standards frequently impact CRF design for studies using EDC.

## **CDASH**

In October 2008, the Clinical Data Interchange Standards Consortium (CDISC) first released the Clinical Data Acquisition Standards Harmonization (CDASH), which was intended to standardize data collection fields used on CRFs. The CDASH standard provides a set of data collection fields that are divided into sixteen domains, and was designed to be applicable to clinical studies regardless of therapeutic area or phase of development. For more information about CDASH and other standards that impact CDM, see the GCDMP chapter entitled “Data Management Standards in Clinical Research.”

## ***CRF Completion Guidelines***

To help ensure CRFs are completed correctly, all CRFs should include clearly stated instructions and have associated CRF completion guidelines. These guidelines are used not only to train site personnel, but also to help clinical monitors when reviewing data on completed forms. In many cases, CRF

completion guidelines may also encompass instructions regarding acceptable methods of correcting or changing the data.

Instructions and completion guidelines should take into account the data collection method used (paper versus EDC) and should be tailored to the individuals who will be completing the CRF. Instructions and completion guidelines may look very different for CRFs completed by subjects rather than those completed by study personnel. Also, paper-based CRFs typically use printed CRF completion guidelines, while EDC systems may use on-line help screens in lieu of printed guidelines. For more information, see the GCDMP chapter entitled “CRF Completion Guidelines.”

### ***CRF Change Control and Versioning***

Any time CRFs undergo changes, appropriate authorization should be obtained, relevant personnel should be consulted (including biostatistics, clinical, regulatory, etc.), and all the changes should be clearly documented. Each revision of the CRF should contain a clearly identified version number or code. Versioning strategies vary widely between organizations, but any successful versioning strategy should clearly identify the correct sequence of CRF versions. When CRFs are revised, the changes made and reasons for those changes should be documented. If CRFs are revised during an ongoing study, ensure all sites use the latest version for subsequent data collection.

### ***Data Privacy***

Although each CRF must correctly represent the subject from whom data are being collected, CRFs must also avoid collecting data that could lead to direct or indirect identification of the subject. Some examples of data that could identify a subject include, but are not limited to, subject names, initials, addresses, or genetic information. Each subject should be assigned a unique code to be used for identification of that subject within the study without jeopardizing his or her privacy. For more information about privacy issues in clinical research, see the GCDMP chapter entitled “Data Privacy.”

## **Future Directions**

The GCDMP chapter entitled “External Data Transfers” provides information on data that are presently routinely directly transferred to a clinical database, such as data from an interactive voice response system (IVRS), a diagnostic imaging device, or an ePRO device. As more physicians and hospitals transition to using electronic health records (EHR), more opportunities arise to streamline collection of clinical data. Several companies are already developing applications that will integrate EHR data with clinical databases used in clinical research. Also known as Retrieve Form for Data-capture (RFD), this approach will streamline data acquisition by eliminating steps (such as source data verification by the monitor during a site visit) currently needed to transport clinical data from a physician’s subject medical charts to a study’s clinical database. Because every data processing step introduces the potential for error, RFD may soon be a huge contributor to improving data quality while also reducing study costs and timelines.

## **Recommended Standard Operating Procedures**

- CRF Design
- CRF Development
- CRF Quality Assurance
- CRF Approval Process
- CRF Version Control Process
- CRF-Related Training

## **References**

1. International Conference on Harmonisation. *Harmonized Tripartite Guideline for Good Clinical Practice*. 2<sup>nd</sup> ed. London: Brookwood Medical Publications; 1996.
2. Spilker B. *Guide to Clinical Trials and Developing Protocols*. New York, NY: Raven Press; 1984.

3. Califf RM, Karnash SL, Woodlief LH. Developing systems for cost-effective auditing of clinical trials. *Controlled Clinical Trials*. 1997;18:651–660.
4. Proschka S. *Practical Guide to Clinical Data Management, Second Edition*. Boca Raton, FL: CRC Press; 2007.
5. Calvert WS, Ma MJ. *Concepts and Case studies in Data Management*. Cary, NC: SAS Institute Inc; 1996.
6. Code of Federal Regulations, Title 21, Part 11, Electronic Records; Electronic Signatures. Washington, DC: US Government Printing Office; 1997.
7. US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Washington, DC: US Department of Health and Human Services; 2007.

## **Further Reading**

Terms used in this chapter may be found in the *Good Clinical Data Management Practices Glossary*.

CDISC CDASH Core and Domain Teams. *Clinical Data Acquisition Standards Harmonization (CDASH)*. Austin, TX. Clinical Data Interchange Standards Consortium; 2008. Available at [http://www.cdisc.org/stuff/contentmgr/files/0/9b32bc345908ac4c31ce72b529a3d995/misc/cdash\\_std\\_1\\_0\\_2008\\_10\\_01.pdf](http://www.cdisc.org/stuff/contentmgr/files/0/9b32bc345908ac4c31ce72b529a3d995/misc/cdash_std_1_0_2008_10_01.pdf). Accessed April 23, 2010.

Kyung-hee Kelly Moon. *Techniques for Designing Case Report Forms in Clinical Trials. Considerations for Efficient Data Management and Statistical Analysis*. ScianNews Vol. 9, No. 1; Fall 2006.

McFadden E. *Management of Data in Clinical Trials*. New York: John Wiley & Sons; 1997.

Pocock SJ, ed. *Clinical Trials: A Practical Approach*. 1st ed. New York: John Wiley & Sons; 1984.

Spilker B, Schoenfelder J. *Data Collection Forms for Clinical Trials*. New York: Raven Press; 1991.



## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
October 2010	Revised for content, style, grammar, and clarity. Chapter title changed from “Data Acquisition” to “Design and Development of Data Collection Instruments.”

**This page is intentionally blank.**

---

# **Edit Check Design Principles**

December 2009

## ***Abstract***

Edit checks are invaluable tools for increasing data quality and providing greater efficiency during data review and cleaning activities. This chapter discusses the process of edit check creation, including balance and efficiency considerations. The chapter also describes different types of edit checks, edit check validation, strategies for edit check specification creation, training related to edit checks, and considerations for using edit checks in studies that are paper based or use electronic data capture.

## ***Introduction***

The ultimate goal of clinical data management (CDM) is to complete every study with a dataset that accurately represents data captured in the study. No matter how much care is taken in collecting and entering data, discrepancies and data errors will invariably find their way into a clinical database. The vast majority of these data inconsistencies and errors can be alleviated with careful review and data-cleaning activities.

Review and cleaning of various data types may be performed by different personnel according to their knowledge and training. For example, data managers may not have the relevant medical knowledge to determine if an out-of-range lab value is indicative of a possible adverse event (AE) unless explicitly defined in the protocol or data management plan (DMP). Similarly, data entry personnel may not have the level of knowledge needed to recognize data indicative of protocol violations. Although responsibilities vary between organizations, CDM typically reviews triggered edit checks in addition to reviewing data that may be outside the scope of data entry personnel's experience.

Carefully designed edit checks can greatly increase efficiency and data quality by automating many data review processes within the clinical database or clinical data management system (CDMS). CDM personnel and members of the study team should collaborate to determine what edit checks should be in place to fulfill study requirements and reduce potential data errors and inconsistencies. Although assignment of responsibilities varies between organizations, CDM may be involved with all phases of edit check specification and testing, with the possible exception of edit check programming.

## **Scope**

This chapter discusses the use of edit checks in clinical studies, including the purpose of edit checks, types of edit checks, creation processes of edit check specifications and development, and edit check testing. The chapter is intended as an overview of edit checks from a CDM perspective, and does not discuss details of programming and conditional statements used in edit checks.

Roles and responsibilities vary between organizations, and some of the topics discussed in this chapter may be the responsibility of different departments in different organizations. Regardless of role assignment, CDM personnel should be aware of the processes discussed in this chapter and how they impact their roles as data managers.

## **Minimum Standards**

- Finalize protocol and complete initial database specifications prior to designing edit checks.
- Specify edit checks based on parameters of case report form (CRF) pages and safety and efficacy parameters from the protocol.
- Specify edit checks for all primary study endpoints and safety data.
- If applicable, specify edit checks with external data (e.g., laboratory data) for reconciliation purposes.
- Ensure all edit checks are programmed, validated, and documented in accordance with established standard operating procedures.

- Ensure all edit checks specification documents are appropriately version controlled.
- Provide training to relevant personnel on the impact of edit checks on their individual roles in entering and managing clinical data.

### ***Best Practices***

- Where appropriate, specify edit checks to compare study inclusion and exclusion criteria and any data (that are collected in CRF pages) that could be indicative of protocol violations.
- Design edit check specifications so redundant output does not occur when edit checks are run.
- Review edit checks with appropriate clinical and statistical personnel to ensure edit checks meet study needs and help identify inconsistencies in study endpoints.
- Specify edit checks for all study endpoints and all data supporting safety data and study endpoints.
- Develop a library of standard CRFs and edit checks based on standards used, such as CDASH or company-specific standards.
- Perform a quality control review of edit check design and specifications prior to performing user acceptance testing (UAT) of edit checks.
- Evaluate the effectiveness of edit checks once in active use, and modify, delete or create new edit checks accordingly.

### ***Purpose and Process of Edit Checks***

The purpose of edit checks is to draw attention to data that are inconsistent or potentially erroneous. Edit checks may be described as automatic warnings or notices that are generated by a database, CDMS, or other data entry application, and are triggered by data that are missing, out of range, unexpected, redundant, incompatible or otherwise discrepant with other data or study parameters. Most edit checks are triggered during the data entry

process, and may prompt the data entry operator to double check a value before saving the data. Other edit checks may be triggered by characteristics of related or aggregate data, and are more likely to notify CDM personnel of potential data errors after data entry has occurred. The potential data errors identified by triggered edit checks may prompt CDM personnel to perform data-cleaning activities such as performing self-evident corrections or generating queries to a site.

## **Balance and Efficiency Considerations**

When creating edit check specifications, balance and efficiency considerations should be taken into account. Although edit checks can save considerable time and money in regard to data accuracy and cleaning, an edit check should not be created simply because it is possible to do so. Edit check specifications should be carefully designed to ensure checks are in place for critical data fields such as efficacy and safety variables. However, for variables not related to study endpoints or safety parameters, an evaluation should be made to determine whether the benefit provided by an edit check justifies the resources needed to create and test the edit check. This process should also evaluate the benefit of the edit check against the resources needed to review and close discrepancies generated by the edit check, as well as the resources needed to conduct the query process once the study is in progress.

When evaluating balance and efficiency factors for edit check specifications, consider that some edit checks may be less feasible or efficient than a manual review. For example, although edit checks can be created for free text data fields, manual review of listings by CDM may be more efficient, reliable, and cost-effective for this type of data. Even if an edit check could be programmed to account for every possible variant in a free text field (which is doubtful), a manual listing review would typically be more efficient and better suited to identifying unanticipated entries.

Some data irregularities may be more appropriately identified by biostatisticians than through edit checks or manual reviews. Some unexpected data trends may be indicative of systemic problems with data collection or processing and may not be easily identified by an edit check or manual review. In many cases, these types of data trends are most accurately and efficiently identified during preparation for statistical analysis. Because biostatisticians may also be able to suggest edit checks that can make their

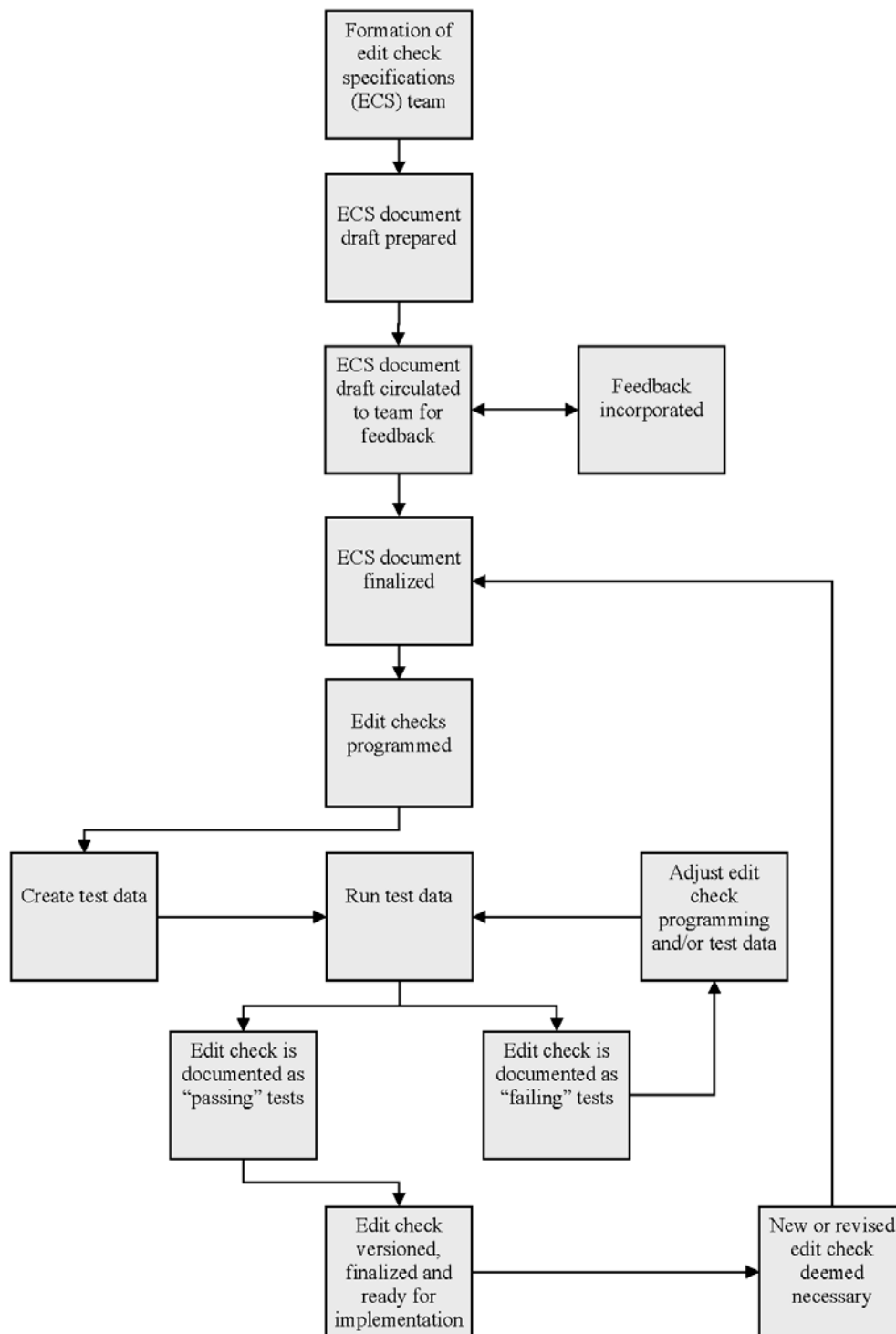
work more efficient when performing statistical analyses, consult the biostatistician(s) when designing edit check specifications.

Other potential data errors may be most efficiently identified by clinical research associates (CRAs), medical monitors, or medical coders. In many cases, a CRA or medical monitor may identify potential data errors by noting a trend and requesting a listing. Subsequent review of the requested listing may allow the CRA or medical monitor to confirm or deny the presence of the suspected data error(s). Additionally, medical coders may identify inconsistencies while coding data and subsequently bring these inconsistencies to the attention of appropriate CDM personnel.

### **Process of Edit Check Development**

Edit check development is a process that requires information from a variety of sources, and should ideally incorporate a multidisciplinary approach to ensure appropriate and effective edit checks are implemented.<sup>1</sup> Although some details of edit check development processes may vary between organizations, the general steps should be similar between organizations for both edit check creation and testing. Figure 1 presents an overview of edit check development and testing processes.

Figure 1. Flowchart of Edit Check Development and Testing Processes





## **Creating Edit Check Specifications**

Edit check specifications are crucial to identify invalid data, missing data, inconsistent data, and out-of-range values. Edit check specification planning requires information from a number of sources and should be performed with a comprehensive strategy for specification development in place prior to creating the initial draft.

Sources of information for edit check specifications may include:

- Study protocol—The study protocol describes the intent of a study, identifying inclusion/exclusion criteria, safety parameters, and primary and other study endpoints.
- Data management plan—Although the study protocol provides a broad overview of study parameters, a DMP typically describes in more detail data conventions for the study and identifies variables for which edit checks may need to be designed.
- Annotated CRFs and database design documentation—After identifying variables for which edit checks will be created, annotated CRFs and the database design should be examined to ensure edit checks are properly aligned with answer choices and the database structure.
- Standard edit check macros—Developing and maintaining a repository of commonly used edit check macros can save considerable time and money by avoiding duplication of work across studies or datasets.<sup>2</sup>
- Biostatisticians—Biostatistician(s) can provide direction regarding areas where edit checks may be desired to facilitate delivery of data that are suitable for statistical analyses without needing further cleaning or manipulation.
- Study personnel—Site personnel or other study personnel may be able to identify data fields that have been particularly prone to errors, inconsistencies, or out-of-range values in previous or similar studies.

Edit check specifications are typically documented in a table or spreadsheet format using various software applications. Although format, structure, and

level of detail may vary greatly between organizations, Table 1 presents an example of how an edit check specification table might be organized.

**Table 1. Sample Edit Check Specification Table**

CRF	Field Name (Number)	Check Name	Edit Check	Edit Check Message
ENROLL	Subject ID (2)	DUP_REC	Duplicate subject ID number	This subject ID number has already been assigned for this site. Please confirm correct ID number.
DEMOG	Subject ID (2)	NO_SUBJ_ID	Missing subject ID number	A subject ID number has not been entered for this record.
DEMOG	Subject DoB (6)	INVLD_AGE	Subject age is out of range	The date of birth value entered may be invalid. Please confirm correct date of birth.

### Hierarchical View of Edit Checks

Some edit checks may be more important than others. Although a risk management approach can help identify edit checks that are crucial to the success of a study, a hierarchical approach to designing edit checks may be more efficient and provide similar results. The following items are an example of how a hierarchical sequence of edit check specification creation might be designed.

- General clinical data checks—These are checks designed to ensure key clinical data are accurate, reliable, and consistent. Although most edit checks fall under this category, some are more crucial than others.
  - Endpoint checks—Primary and other study endpoints should have checks in place to identify missing, erroneous, or out-of-range values. These are the variables for which statistical analyses will in part determine whether a study’s primary and secondary hypotheses are accepted or rejected. As such, the integrity of these data is crucial to the success of a study.
  - Safety checks—Edit checks should be created to help ensure any deviations from key safety parameters are noted and handled

accordingly. For example, if an AE is noted but no AE form is present, an edit check should flag this discrepancy so appropriate action can be taken.

- Protocol compliance checks—Data indicating adherence to study inclusion and exclusion criteria should be subjected to edit checks, as well as other protocol-specified parameters such as acceptable follow-up visit intervals.
- Programmed checks—For greatest efficiency, the majority of edit checks should be programmed into the clinical database or data capture system. These checks automatically trigger when certain predetermined conditions are met, such as missing data from a particular field or inconsistencies between data fields.
- Manual checks—Manual checks should be used for those data that cannot be easily checked through programmed edit checks, such as free text fields. Manual checks may also be used to verify key information such as site and subject identifiers on paper CRFs.
- Listings checks—Edit checks may also be designed for listings, which are used for checking multiple data points (where both correct and discrepant values may reside) across a single subject or module. Reviewing listings for discrepancies is typically a manual process.
- External checks—In some cases, most commonly in large complex studies, some checks may be programmed to run against data transferred from an outside source (e.g., labs). These checks are often run on multiple subjects with data from multiple datasets but only output data for subjects who fail the check.

## **Use of Standards for CRFs and Edit Checks**

Use of standard edit checks based on standard CRF templates can save time and money while increasing quality, as well as potentially make the programming of edit checks easier. Use of standard edit checks can also decrease the amount of time needed for programming, therefore decreasing overall study timelines. Standard CRF templates may be prepared using CDASH or corporate standards, as appropriate, and version or change controls

should be applied. Although the types and scope of clinical studies may vary within a single organization, maintaining a central repository of CRFs and corresponding edit checks can reduce time and expenses for subsequent studies.

Standard edit checks should clearly identify the version of the corresponding standard CRF template. If a standard edit check template needs to be customized in some areas (e.g., a page number must be specified), the customized area should be flagged to draw attention to it. For more information about the CDASH standard, see the GCDMP chapter entitled “Data Management Standards in Clinical Research.”

### **Consistency in the Edit Check Specifications Document**

The edit check specifications document should be consistent in its wording and conventions. The specifications document should also be consistent with the CRFs for which the edit checks are specified. The following are some examples of areas that should be reviewed for consistency within an edit checks specifications document.

- Use generic terms, such as "Subject" rather than "Patient," although a global change to “Patient” may need to be made for some studies.
- Note field names exactly as they are provided on the corresponding CRF (e.g., "Date of Birth" rather than "Birth Date," if “Date of Birth” is how the field is identified on the corresponding CRF).
- All descriptions in the edit check specifications document should be stated in complete sentences, using consistent terms such as "Visit Date must be present," or "If Not Done is marked, Result must be blank."
- Use consistent formatting conventions such as capitalizing all field names, or adding brackets only when a sentence is not clear without them (e.g., “A response must be marked for [Were any Adverse Events experienced?]”).
- Note any exceptions or special instructions for the reviewer (e.g., “NOTE: Do not query if page is lined through.”).

## **Message Wording**

In addition to the care that must be taken to ensure edit checks are in place for key variables, the wording output by edit checks should be clear, unambiguous, and not leading. Any manually added queries to a clinical site should follow the same conventions as edit check output wording. The wording of both queries and edit check output messages should be carefully chosen to clearly and unambiguously relay the following information:

- Study, site, and subject or subject record—While adhering to data privacy conventions and regulations, queries and edit check outputs should clearly identify the study, site, and subject record for which an edit check or query is triggered.
- Variable name and value—Queries and edit check outputs should clearly identify what field, variable and value triggered the edit check or query.
- Supporting values—If an edit check or query is triggered from a derived value or is associated with other fields, supporting values should also be identified. For example, if an edit check is triggered by an out-of-range value for computed body mass index, the output message should indicate the value's relationship to the supporting fields containing subject height and weight.
- Message composition—Queries and edit check output messages should clearly identify the discrepant data and acceptable options for discrepancy resolution, but should not introduce bias or pose leading questions in any way. For example, an edit check for blood pressure should not output a message that specifies the expected range. Rather, the message should simply state that the value is out of the expected range and request confirmation or correction of the blood pressure.

## ***Types of Checks***

Edit checks are created to identify a number of different types of data inconsistencies or potential data errors. Although most edit checks are programmed into the database or CDMS and are triggered automatically when predefined conditions are met, data inconsistencies and potential data errors may also be found through manual data review.

Some of the most commonly used types of programmed edit checks include the following:

- Missing values—Edit checks for missing values are not usually applied to all data fields, but should be used for critical variables such as site and subject identification numbers or primary safety and efficacy variables.
- Missing CRF pages—In contrast to edit checks for missing values, edit checks for missing CRF pages may be applied to all CRFs. The intent of these checks is to highlight that an entire page or multiple pages have not been entered, which may be an oversight by the data entry operator or may result in a query to the site.
- Range checks—These are some of the more commonly used edit checks, and are intended to identify values that may be the result of an entry error or that may be indicative of a value outside of those expected for the subject population. Some examples may include height, weight, blood pressure, and other physiological parameters for which a particular range of values might be expected.
- Checks for duplicates—These checks are intended to negate the potential for the same data to be entered into the database more than once. Duplication may take the form of a duplicate subject identification number being used, a follow-up form being entered twice for a particular subject and interval, a single AE being entered twice, or any other situation where duplicate pages or data are entered.
- Logical inconsistencies across single CRF—The nature of potential logical inconsistencies may vary greatly between studies, but one example would be a CRF indicating that the subject is pregnant, but also indicating the subject is male. An edit check for this type of logical inconsistency can flag a data error that may not have been noticed otherwise.
- Inconsistencies across CRF pages or modules—Edit checks for logical inconsistencies are not limited to inconsistencies on a single CRF. Edit checks can also be programmed to identify discrepant data across CRF pages or modules. An example could be an edit check flagging an AE form that indicates that a medication was prescribed without the medication being recorded on a corresponding concomitant medications form.

- Checks of external data—Programmed edit checks are not limited to CRF data, but may also be applied to external data (lab data, ECG data, etc.). Many of these types of checks are primarily designed to help ensure that external data are consistent with the subject data within the database.
- Protocol violations—These checks are designed to identify specific data that may be indicative of protocol violations, and may take the form of range checks. One example would be calculating date ranges for follow-up visits to ensure all follow-up visits were within protocol-specified time windows. Another example would be checking subject eligibility forms to ensure all inclusion criteria were met and no exclusion criteria were met.

### ***Front-End vs. Back-End Edit Checks***

Edit checks that are triggered upon data entry are often referred to as front-end edit checks, whereas edit checks across multiple forms are often known as back-end edit checks. Front-end edit checks are typically limited to a single field or CRF page. An example of a front-end edit check would be a flag or warning that appears when an entry operator attempts to enter an impossible visit date, such as February 30 or a date in the future. Although front-end edit checks are usually more numerous, back-end edit checks are typically more complicated and therefore more difficult to program. An example of a back-end edit check would be one that notifies CDM personnel that a BMI (body mass index) entry is not consistent with the subject's reported height and weight.

Although details vary between studies and organizations, Table 2 presents which types of edit checks are more likely to be implemented as front-end checks, back-end checks, or both.

**Table 2. Comparison of Edit Check Types**

<b>Type of check</b>	<b>Front-end check</b>	<b>Back-end check</b>
Missing values	<b>X</b>	
Missing CRF pages	<b>X</b>	<b>X</b>
Range checks	<b>X</b>	
Checks for duplicates	<b>X</b>	<b>X</b>
Logical inconsistencies across single CRF	<b>X</b>	
Inconsistencies across CRF pages or modules		<b>X</b>
Checks of external data		<b>X</b>
Protocol violations	<b>X</b>	<b>X</b>

### ***Electronic Data Capture (EDC) vs. Paper-based Edit Checks***

Edit checks used in paper-based studies may differ somewhat from those used in EDC studies. For paper-based studies, some organizations may choose to limit the number of front-end checks. This ensures that potentially critical errors or discrepancies will be addressed directly by qualified CDM personnel. For studies using EDC, checks for transcription errors are not as necessary. However, more care must be taken in EDC studies to ensure the data entry design and front-end edit checks catch potential errors as they are entered. Because the electronic record may be considered the source document in some situations, there may be no other documentation to check against if possible errors are discovered later. The potential lack of additional source documentation in EDC studies also increases the importance of ensuring all edit checks are in place prior to the start of data collection. For more details about edit checks in studies using EDC, see the GCDMP chapter entitled “Electronic Data Capture—Concepts and Study Start-up.”

### ***Validating Edit Checks***

As with other aspects of a clinical database or EDC system, edit checks should be thoroughly tested and validated. Details may vary between different organizations and electronic systems, but the following process gives an overview of how edit checks should be validated.

- Creating test data—After edit checks are programmed, a set of test data should be created to mimic the type of data that are expected during the



study. This test data should not only include expected values, but also missing values and values that are out of range or that may not be expected from actual study data. These test data are typically created by CDM, although in some organizations database programmers or a quality assurance department may also be involved.

- Testing edit checks with test data—The test data used should include out-of-range or discrepant values that should trigger edit checks, as well as within-range or consistent values that should not trigger edit checks. The test data should contain all different scenarios that can occur for that check. For example, if an edit check is testing for a blood pressure range that is not between 80 mmHg and 200 mmHg, the out-of-range test data should ensure the edit check is triggered for anything below 80 and anything above 200 while the in-range data ensures the edit check is not triggered for anything between 80 and 200, including values that are exactly 80 or 200.
- Testing feedback loop process—This process may vary between organizations and is dependent upon who is doing the testing, which is usually the responsibility of CDM and programming personnel. CDM may give programming personnel the test plan and have the programming personnel test edit checks against the plan. If something in the plan does not occur as expected (e.g., an edit check was not triggered when it was supposed to or was triggered when it was not supposed to), the programmer notifies CDM, who may then modify the test data or add additional test data. Regardless of who performs the testing, data management should attempt to ensure all possible scenarios are tested, and should clearly document if any possible scenarios are not tested.
- Documentation—Every step of the edit check testing and validation process should be thoroughly documented. Both test data and edit checks may be documented on electronic or paper CRFs. If an organization does not have a formal test plan, these annotated test CRFs may suffice as a test plan provided the edit checks are described in sufficient detail. Documentation should also exist from the database showing where checks were triggered or not triggered. Any changes made to edit checks or test data during testing should also be documented. How documentation is achieved varies between organizations. One approach is to consolidate the test plan with the edit check specifications document, including a

“pass/fail” column that must be initialed and dated by the individual who is testing edit checks. Another approach may be to compile a binder with edit check specifications, the programming code behind edit checks, and test output from the database showing where each check was triggered or not triggered, with the initials and date of the individual who tested the check.

- Quality control (QC)—Although QC responsibilities may vary between organizations, some form of QC should be performed for the entire edit check validation process, final edit check programming, and all associated documentation. In different organizations, some or all of these QC processes may fall under the responsibilities of CDM personnel, project managers, database programmers, quality assurance personnel, or a manager of database development.
- Validation of new or revised edit checks—If any edit checks are added or revised during the course of a study, the same steps should be followed as are used for edit checks created at the beginning of the study.

## ***Maintenance of Edit Checks***

After edit check testing and validation has been completed, all responsible parties should provide written approval of edit check documentation prior to using the edit checks with actual subject data. CDM typically maintains an edit check document, ensuring that the document is kept current and incorporates proper version or change control. If substantial changes are made to the edit check document or the study is ongoing for more than a year, prior to study closeout CDM may request an additional review and approval of the final edit check document or changes made to the document. This re-review is intended to ensure that the needs of all parties continue to be met.

The edit check document should be considered a living document throughout the life cycle of the study. Edit checks may need to be changed as a result of CRF changes, or errors discovered in logic or terminology that need to be corrected. In addition, database programmers may suggest changes that result in more efficient data processing. As data are processed, new checks may be designed to identify discrepancies noted by monitors, biostatisticians, or other reviewers.

## **Change Control**

Ideally, all changes to edit checks should be tracked within a single edit check document. However, a separate document may be employed if needed. All changes should be accompanied by the responsible individual's initials, the date of the change, and the reason for the change. If a change was approved or directed via e-mail, the date and sender of the e-mail should also be identified within the change document. Any new or changed edit checks should be thoroughly tested in accordance with established edit check testing procedures.

## **Version Control**

Although different organizations may employ different strategies, a common strategy is for the first approved version of an edit check document to be considered Version 1 (V1.0). With this approach, minor administrative changes may be made at any time, and will change the version number by one-tenth (e.g., V1.1, V1.2, etc.). If CRF changes or other substantial changes occur, when the edit check document is subsequently updated, the version is updated by 1 (e.g., V2.0, V3.0). Regardless of the specific methodology used, all versions of an edit check document should be clearly documented.

Upon conclusion of a study, the final version of the edit check document should be archived with all other pertinent study documentation.

## ***Edit Check Training***

All data entry and CDM personnel who will be entering data, reviewing data, or reviewing the output of edit checks should be trained prior to data entry into the database. All personnel involved with these processes should have basic training in the formats, terminology, and use of edit checks, and the documentation of this training should reside in training folders. Training can be tailored to each individual role. For example, data entry personnel may only be trained on those edit checks that may be triggered upon data entry.

Data entry and CDM personnel may also need to undergo study-specific training for any edit checks that are unusual or unique to the study. If needed, a brief overview of the study and a review of the CRF may be included in the training. Study-specific training should also have clear documentation, and

may be maintained in training folders if confidentiality is not a concern. Otherwise, documentation of study-specific training may be maintained by data management and archived with all other pertinent study documentation at the close of the study.

### **Recommended Standard Operating Procedures**

- Database Design
- Edit Check Specifications
- Edit Check Validation

### **References**

1. Rondel RK, Varley SA, Webb CF, eds. *Clinical Data Management Second Edition*. West Sussex, England. John Wiley & Sons Ltd. 2000:113–119.
2. Gupta S. Standards for Clinical Data Quality and Compliance Checks. *Data Basics*. Winter 2008.

### **Further Reading**

N/A

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
December 2009	Initial publication.

---

# **Electronic Data Capture—Concepts and Study Start-up**

September 2008

## *Abstract*

Electronic data capture (EDC) has emerged as a proven tool for sponsors of clinical trials. Understanding the principles of EDC is more important than ever for clinical data management (CDM) professionals. This chapter reviews the regulations and guidance that currently apply to EDC during pre-production and study start-up, and emphasizes the important role that CDM professionals have in the adoption, development, and improvement of EDC systems.

## **Introduction**

Electronic data management for research emerged in the 1970s and has evolved into a suite of processes and tools to enhance the management, quality control, quality assurance, and archiving of clinical trial research data. In the 1990s the development of electronic data capture (EDC) tools for clinical trials research became more focused. Today, EDC is gaining in popularity, and regulatory agencies are readily accepting submissions in which validated EDC tools are used. EDC systems should be more than just a means to an end and quality EDC systems can be drivers of the entire clinical trial's information management process. Data managers provide significant value in designing processes to make the transition from paper systems to EDC systems efficient while ensuring data integrity is maintained.

The return on investment has been proven for the automation of clinical trials information management processes from data entry through the summarization and archival processes. Although remote data entry (RDE) processes emerged in the 1970s,<sup>1</sup> these processes languished for 20 years without significantly impacting clinical trials. By the mid-1980s, personal

computers (PCs) were introduced to clinical trials for clinical data capture, which led to a major transformation in the way clinical data was captured. Prior to that time, site professionals collected data on paper case report forms (CRFs) and sent the forms to a centralized sponsor facility where data computerization took place. This method of data capture was called “centralized” because data was entered into a computer system in a single facility by professional data entry staff. The investigators’ main responsibilities were the original completion of the paper CRFs and responding to queries that arose following review of computerized data.

Having PCs at the investigator site allowed for the introduction of “decentralized” clinical data capture, which became known as remote data entry (RDE). This development began a paradigm shift in clinical trial conduct, placing the responsibility for electronic data entry on site staff. Many sponsors developed proprietary hardware and software solutions to manage RDE at investigator sites. Computerized data was routinely transferred from each investigator site to the sponsor through some type of periodic file transfer, for example, using file transfer protocol (FTP). The FTP process was usually done via phone lines and took some time to complete, depending on the volume of data to transfer.

In the late 1990s, Web-based approaches to clinical data capture were introduced in an effort to gain efficiencies that other industries had realized by moving processes to the Internet. The acronym RDE was subsequently replaced by EDC as data transfer was expedited by Internet technologies rather than FTP, resulting in more frequent and rapid data transfers.<sup>5</sup> The introduction of Web-based EDC led to greatly expanded use of decentralized clinical data capture.

## **Scope**

This chapter provides information on the concepts and start-up activities related to EDC systems for companies who are considering transferring some or all processes from traditional paper data collection to EDC. It concentrates on establishing an environment conducive to incorporating EDC technologies into the clinical trial process from the viewpoint of data management. Practices, procedures, and recommendations are proposed for data managers to prepare for and start an EDC system that will properly align electronic data capture technology to support statistical and medical research needs.

Comparisons between paper data collection methods and EDC are also presented. The primary focus in this chapter is on start-up activities to support EDC for electronic CRFs (sometimes called eCRFs, although the term will not be used in this document) and data integration with non-CRF data.

Many of the tasks described in this chapter may be joint responsibilities between different groups, just as there may be many different groups involved in the implementation of various tasks. However, clinical data managers need to be conscious of whether or not these tasks have in fact been performed in a satisfactory manner.

Recommendations for proper study conduct and study closeout using EDC will be addressed in the chapters entitled “Electronic Data Capture—Study Conduct” and “Electronic Data Capture—Study Closeout.” Recommendations for patient diaries and interactive voice response systems (IVRS) will be addressed in future chapters of the GCDMP.

### ***Minimum Standards***

- Ensure compliance with 21 CFR 11 and consistency with the Food and Drug Administration’s (FDA) *Guidance for Industry: Computerized Systems Used in Clinical Trials*.<sup>2,3</sup>
- Stated quality standards should support the utilization of automated data capture, management and archiving.
- Ensure requirements are defined for data transfers and integration with other systems.
- Software systems validation should be scheduled and completed prior to EDC study implementation.
- Ensure user acceptance testing (UAT) is completed prior to implementation and deployment to sites.
- Verify training is provided for all users of the EDC systems and that all training is documented and minimum competencies are met.
- Verify access to data is limited to authorized individuals.

- Determine roles and responsibilities in data review and query management.
- Software technical support should be provided to users and a toll free phone number should be available for the help desk.
- Ensure sites have access and control of data up to database lock.

### **Best Practices**

- Use business process analysts (possibly external, for objectivity) to establish EDC-specific workflow processes and identify required transitions from current processes.
- Do not apply paper study processes to studies using EDC.
- Identify stakeholders in current processes, as well as additional stakeholders required for new EDC processes.
- Plan studies to avoid “last minute” system modifications that introduce errors and complexity to study-specific CRFs.
- Develop CRFs or data collection tools with teams of individuals from monitoring, data management, statistics, regulatory affairs, and medical, ensuring adequate attention to the collection of safety data.
- Ensure systems are user-friendly and flexible for data entry.
- Ensure EDC systems do not restrict answers site staff can provide in a way that introduces bias into the clinical study.
- Ensure adequate edit check procedures and query management tools are built into EDC software.
- Before the start of a study, conditions (e.g., SDV completed, all queries resolved) for locking forms and/or casebooks should be set according to a set of criteria, such as, all SDV complete, all data review complete, no outstanding queries or missing data exist.
- When coding in an EDC environment it is recommended not to display coded terms back to the site user.



- Ensure data can be traced from the time of original input through the reporting and analysis files via easily accessible audit trails.
- Ensure ease and quality of all data transfers by testing data transfers prior to deployment of EDC systems.
- Ensure your EDC system integrates as needed with other databases by testing integrations with your EDC system prior to initiating any trials using the system.
- Ensure processes are defined to integrate laboratory and other non-CRF data with data obtained from the CRF.
- Ensure all user acceptance tests are documented.
- Ensure change control procedures include complete documentation.
- Ensure all documentation for use by site staff is adequately reviewed before being provided to site staff.
- If 24 x 7 x 365 support is not available, the help desk should cover the work days/times of all regions included in the study.
- The help desk should support the minimum number of languages needed to communicate with all users and all languages, including local dialects.
- Develop and follow standard operating procedures (SOPs) for electronic data capture, data validation, and data archiving.
- Assess current SOPs for potential impact created by EDC workflow processes and update SOPs as necessary.
- Include SOP modification time in project plans for EDC implementation.
- Assume that both the new workflow and SOPs will be in transition for some period of time as the staff interact with the EDC system following any modification of SOPs.
- Identify issues that merit periodic reminders, such as user ID and password security, and schedule recurring reminders.
- Provide an instruction manual for study workflow processes.

- Verify all users have documented training prior to being granted access to the system.
- Create a training environment in which users can practice, and create training cases as examples that are pertinent to the study.
- Provide a “Train the Trainer” program for clinical research associates (CRAs), data managers or others to be able to provide training to sites.
- Provide training customized to each user’s role. A study coordinator may need in-depth training of most system functions, while users with read only access may need minimal instructions.
- Document all training for trial master files as well as site files.
- Integrate metrics on process and cost/benefit into the EDC process to enable better EDC versus non-EDC comparisons and comparisons across EDC technologies.
- CRF specifications should be finalized prior to finalization of edit check specifications, although development of both should be performed concurrently.

### ***Differences Between EDC and Paper-based Studies***

Four important areas that differ between EDC and paper-based studies are the manner in which data will be collected, the timeline necessary to prepare for the study, the manner in which collected data will be verified, and disaster recovery planning.

### **Offline vs. Online vs. Hybrid Studies**

The three primary modes of capturing data for a study are:

- Offline—the traditional paper-based method for collecting, sending, and collating or an EDC system that works without a constant Internet connection.
- Online—the EDC method, typically using networked resources to record clinical data in electronic forms, which are then stored at a central server location

- Hybrid—a combination of offline and online methods that are either a combination of paper-based systems using EDC to manage some aspect of the data-collection process, or that involve the use of both offline and online EDC methods

The mode chosen is generally dependent upon the capabilities and limitations of the sponsor and EDC software used, as well as of sites that will participate in the study. Therefore analysis and planning are essential to determine which mode should be used for a given study.

EDC solutions are inherently technical implementations that vary in their degree of complexity and level of competence required by users. The EDC process extends data collection (and in some situations, data cleaning) to the site and/or subject. It is critical to accurately assess the ability of sites to use and manage the technology on which the EDC application is based. If it is apparent that a one or more sites lack the requisite technical capabilities to use an EDC solution, the sponsor should consider a paper-based or hybrid study as an alternative.

The results of the following assessment examples, as well as any others that are pertinent to the sponsor, will guide the determination of which data collection mode is best suited to a study.

- Site readiness: including technical capability, staff training and competencies, systems infrastructure, and past EDC experience
- Edit checking complexity: the study's degree of dependency on robust edit checks and their impact on system performance
- Audit trails: the importance of capturing the *entire* audit trail electronically as stipulated in 21 CFR part 11
- Subject population: for studies that can utilize subject-oriented EDC solutions such as ePRO, an assessment of the overall subject population's ability to understand and operate the technology successfully
- Study timelines: the need for short turnaround times
- Study management strategy, for example, the level of monitoring required at each site

- An assessment of the ability of the sponsor's clinical trials management system (CTMS) to interface with an EDC solution

## **Study Development and Start-up Timelines**

Because the study database should become active upon enrollment of the first subject, study start-up is critical for EDC studies. For those working at the sponsor facility, many start-up activities may only need to be performed when EDC is initially adopted. Many of the typical CDM start-up activities for both paper-based and EDC studies include: protocol approval, CRF design, CRF annotation, edit-check specification, user acceptance testing (UAT), and documentation preparation. The differences in CDM start-up timelines for EDC studies are based largely on the increased number of tasks that must be completed before the study may begin. In addition to typical start-up activities, several additional activities may need to be considered for EDC that could impact study development and start-up timelines, including:

- Revision of SOPs to support the EDC process (documentation preparation). For sponsors this activity may be done once, while for CROs this may occur with each sponsor with whom they contract.
- Define roles and access to data by authorized sponsor and site staff
- User account management, which may include access control logs and account management documentation
- Definition and creation of new or modified standard data collection forms
- Trial specific programming and UAT (e.g., edit checks, screen designs)
- Preparation of coding dictionaries and processes as needed
- Design, programming, and testing of reports. Establish standard reports that can be reused across compounds.
- Communicating trial status impact on timelines
- Definition and requirements testing for data transfers and integration with other systems or third party data. Utilize industry standards where possible (ODM, LAB, etc.).

- Selection of task-specific applications (e.g., a grant payment system) that may need to be integrated with the EDC system
- Site assessments for the ability to use EDC
- EDC system- and trial-specific training
- Help desk support for users
- Disaster recovery planning

With these CDM start-up activities for EDC, it is important to remember that these activities are highly cross-functional.

### **Source Document Verification (SDV)**

The FDA has issued requirements for electronic records and signatures in 21 CFR Part 11, which provides criteria for considering electronic records as equivalent to paper records and electronic signatures as equivalent to handwritten signatures. Determining the level or amount of SDV is not within the scope of DM, however, it is important to determine if the SDV process impacts the database in any way. In principal, conducting source data verification (SDV) on electronic records is the same as for paper records. Electronic records, like paper records, must be accurate, original, legible, attributable, and contemporaneous. It is important to determine how SDV processes will function before the start of an EDC study so the database can be configured to support access, workflows and reporting requirements. Validation of computerized systems is a completely different, but very important, aspect of electronic records that must be fulfilled as well.<sup>2</sup> Some systems will allow different SDV strategies and some will not. This needs to be agreed upon up front in case any study specific configuration is required.

The *ICH Harmonised Tripartite Guidelines for Good Clinical Practice*, the *WHO Guidelines for Good Clinical Practice for Trials on Pharmaceutical Products*, and the *Code of Federal Regulations* require that source data verification must occur for all clinical trials in phases I–IV. An evaluation of the conformity of data presented in CRFs with source data, SDV is conducted to ensure data collected are reliable and allow reconstruction and evaluation of the study. The SDV responsibilities of the principal investigator, sub-investigator, study coordinator, monitor, quality assurance auditor, and the

clinical trial manager must be made clear at the outset of the clinical trial, and adequate training should be provided to all staff involved. So there are no misunderstandings or errors when SDV is undertaken, special emphasis should be placed on confidentiality and direct access to data. All staff involved must realize that SDV adds to the scientific and ethical integrity of a clinical trial.<sup>1,3</sup> Records of what was done and found, including an evaluation of findings, must be made in the same way as for any other aspect of the trial.<sup>4,5</sup>

In the SDV process, information reported by an investigator is compared with the original records to ensure that it is complete, accurate, and valid. Strictly speaking, every item of data that appears in a CRF should be documented somewhere else to allow verification, audit, and reconstruction. The main objective of SDV is to confirm that the information collected during a clinical study is complete, accurate, reliable, and verifiable so as to give confidence to the sponsor and the regulatory authorities in the data being used to support a marketing application. SDV is also required to provide confidence in any data reported, for example, in published manuscripts and at scientific conferences. Without SDV or stringently controlled electronic source data collection methods, no scientist can have confidence in the data presented and in the conclusions derived.<sup>4,5</sup>

All information in original records of clinical findings and in certified copies of original records are necessary for the reconstruction and evaluation of the trial. These records may include hospital records, clinical and office charts, laboratory notes, memoranda, subjects' diaries or evaluation checklists, pharmacy dispensing records, recorded data from automated instruments, microfiches, photographic negatives, microfilm or magnetic media, X-rays, subject files, records at the laboratories and at medico-technical departments involved in the clinical trial, observations, and documentation recording activities in the clinical trial. The following data are considered key data in SDV and any gross errors in these data might be detrimental to the scientific and ethical quality of the clinical trial:

- Primary efficacy data
- Inclusion/exclusion criteria
- Medical and medication history

- Physical examination and vital signs
- Visit dates
- Adverse events
- Concomitant medication
- A record that the patient has entered a clinical study and the date of informed consent

### **Disaster Recovery and Business Continuity Planning**

When determining a move to EDC, ensure that your facility and selected vendor has a plan in place for Disaster Recovery. Disaster Recovery Plans (DRP) are very similar between EDC and paper-based trials, but are always a key consideration. In the context of this section, a disaster is an event that significantly disrupts operations, either temporarily or permanently. This event could be due to fire, theft, or a weather-related incident that removes access to data on the servers; the sudden unavailability of key members of internal or external (e.g., vendor) staff; or the EDC vendor becomes insolvent. The goal of an organization's Disaster Recovery Plan should be to minimize the loss of operational control in the event of a disaster and to restore business activities quickly with minimal disruptions. As no single response pattern is appropriate to all organizations, a DRP should be flexible in its design.

Data management should cooperate with the information technology (IT) department to ensure that a plan is in place for all hardware and software being used to implement the EDC system. The location of all components of an EDC system should be known and documented to ensure that each of these possible points of failure are addressed by the DRP.

A tiered DRP establishes different levels of response for different levels of failure. Examples of tiers, listed by increasing levels of severity, might include the following:

- Localized failure—one system drive becomes nonfunctional
- Server failure—an entire server becomes nonfunctional

- Office building failure—all resources become unavailable at a building where business operations are conducted
- City failure—all resources become unavailable within a geographic region

Accompanying the DRP should be a Business Continuity Plan (BCP) that guides continuation of a study during the recovery of failed systems. Depending on the number of EDC vendors and contract research organizations (CROs) used by the sponsor, the BCP may be included in the EDC project's data management plan, or exist separately as a plan applicable to all EDC projects. The BCP should identify alternative processes in the event the EDC system becomes temporarily or permanently unavailable. For example, a project may revert to faxing paper CRFs and queries. The BCP should also establish the process by which sites will be informed of EDC system downtime, and the alternative method for collecting data while the system is unavailable.

### ***EDC Deployment Considerations***

Considerations for deployment of an EDC system should be taken into account at the organization level and used when researching, interviewing and assessing EDC vendors. Considerations for this step in the EDC process fall into three main categories:

- Understanding different types of software technology (pdf-based, XML-based, etc.)
- Understanding different EDC system capabilities
- Researching general information about vendors

### **Thin Client and Thick Client Technology Comparison**

There are several issues to consider when selecting an EDC vendor and client-server application for a clinical trial. A key decision is whether the bulk of the workload will be done on the client (investigational site) computer or on the server. This decision can determine the costs of clients and servers, the robustness and security of the application as a whole, and the flexibility of the design for later modification.



A thick client (also known as a “fat client” or “rich client”) is client software that performs the bulk of data processing operations and does not necessarily rely on the server. For example, a word processing program installed on a personal computer is an example of a thick client. All documents are created and stored on the PC without the need for processing by a server. For study coordinators to perform data entry, the thick-client approach requires software to be downloaded to computers at the investigational sites.

The use of the thick-client approach introduces a number of challenges. In today’s hospital environment, concerns for privacy and security must be considered. Users may not have administrative rights to install software and existing firewalls may block communication with servers. A thick client may require the use of dedicated internet connections and provision of IT hardware. Each installation of the client software requires validation in accordance with 21 CFR Part 11. Thick clients can encounter versioning issues, as users have to connect to the remote server to retrieve software updates, and accurate records must be kept to ensure all users are using the most current, approved version of the client software. Also, if a user must synchronize the client with a central server to submit data, contact may be lost before synchronization is complete, resulting in inconsistencies.

However, advantages of thick clients include the following:

- **Local processing:** Complex logical checks and coding can be carried out immediately.
- **Less burden on server:** Because thick clients handle much of the application processing, a thick client does not require as high a level of server performance as a thin client. The use of a thick client also reduces server loads by being able to use the client machine for processor intensive tasks like reporting and analysis.
- **Better multimedia performance:** Thick clients have advantages in multimedia-rich applications, which are bandwidth intensive when delivered over a thin client.
- **Flexibility:** On some operating systems, software products are designed for PCs that have their own local resources. Running such software as a thin client can be difficult.

- Thick clients allow sites with low bandwidth to still remain electronic and transmit their data on-demand.

In client-server architecture, a thin client depends primarily on the server for processing activities. For example, thin clients include browser-based EDC platforms that require the user to log on with the combination of a user name and password. Information is entered and stored centrally, and no data are retained on the investigational site's PC.

There are several advantages to using a thin-client model. The study coordinator does not have to use one specific computer to access and enter data. This capability is especially helpful when staff must share a limited number of PCs. Installation of study-specific software is not required, and centrally managed updates and patches ensure all users have identical client software. Dedicated network connections are no longer considerations, allowing for much greater user flexibility.

Additional advantages of using a thin client include the following:

- Lower IT administrative costs to the project: Clients are managed almost entirely by the server, and the hardware has fewer points of failure. The local environment is highly restricted, thereby improving protection from malicious software.
- Easier to secure: The client can be designed so that application data is only displayed in the browser but never resides on the client PC in any form.
- Lower hardware costs: Thin client hardware does not contain a disk, application memory, or a powerful processor and therefore can go long periods without needing an upgrade or becoming obsolete. The total hardware requirements for a thin client system are usually much lower than for a thick client system. With thin clients, memory can be shared.
- Less network bandwidth: Since terminal servers typically reside on the same high-speed network backbone as file servers, most network traffic is confined to the server room. When a thick client is used, a large file that is accessed may be transferred in its entirety from the server to the client PC. When the same file is saved or printed, the thick client sends the entire file over the network. If sites have limited access to bandwidth, this process can be highly inefficient. When a thin client is used, only mouse movements, keystrokes, and screen updates are transmitted between the

server and end user, thereby enabling large files to be accessed with far less bandwidth.

- **More efficient use of resources:** A typical thick-client is designed to handle the maximum processing needs of the user. However, such a design can be inefficient when allocated processing resources are not fully used. In contrast, thin clients are typically designed to use only the amount of resources required for the user's current task.
- **Simple hardware upgrade path:** In the thin-client model, if the peak resource usage is above a pre-defined limit, boosting resources is a relatively simple process (e.g., adding another rack to a blade server), and existing units can remain in service alongside new units. However, in the thick-client model, resolving this issue may require an entire PC to be replaced, resulting in downtime for the user and concerns regarding the secure disposal of the old unit.

When using thick clients, the following questions must be addressed prior to implementation:

- Will the site's IT department permit external software to be installed?
- Will the site's network firewall and security systems interfere with communication between the client and server?
- Who will be responsible for maintaining software and ensuring updates are provided? Will maintenance result in any downtime for users, and if so, how will downtime be managed?
- Will a dedicated PC or internet connection be used for the study? Does the study's budget include the cost of these resources? Does the site have space for the equipment required by a thick client?
- Will there be any restrictions regarding use of Internet access, such as periods when the investigational site staff are unable to connect to the Internet due to scheduled network maintenance?
- Will technical support be provided, and if so, by whom?

When using thin clients, the following questions should be addressed prior to implementation:

- Will the site's network firewall and security systems interfere with communication between the client and server?
- Who will be responsible for maintaining software and ensuring updates are provided? Who will be responsible for maintaining records regarding new updates?
- Who will be responsible for ensuring browser compatibility? If a site does not have a compatible browser, how will this issue be addressed?
- Will a dedicated PC or Internet connection be used for the study?
- Will there be any restrictions regarding use of Internet access?

### **Application Service Provider (ASP) vs. Technology Transfer**

The decision to use either an ASP or technology transfer model of EDC depends largely on the sponsor's long-term strategy. The determining factors are usually based on the frequency of use of the software (e.g., how many studies for which it will be used) versus the cost of purchasing and maintaining the software.

#### **Application Service Provider**

An ASP is essentially a company that offers its software for use by another company at a cost. The software itself is not purchased, only the opportunity to use that software. The vendor retains full ownership of software, and the client pays for it on a "per use" basis. When an organization uses EDC software in the ASP model, the software resides on the vendor's hardware and under the vendor's authority. It is accessed by the client through a browser or other client software provided by the vendor.

The incentive for the sponsor to adopt an EDC system that uses the ASP model is that implementing, hosting, and validating software are left to the vendor, as are the issues of upgrading, maintaining, and supporting the software. A risk-based approach should be used to determine the scope and depth of any additional sponsor software validation to be performed. The ASP pricing structure takes all of these issues into consideration, and therefore ASP pricing per study is typically higher than when using a technology or knowledge transfer system.

Some advantages to using an ASP model for an EDC system include:

- Little or no setup time is needed, and limited or no software integration is required to begin using the client software.
- Pay per use of the software or “pay as you go”
- Costs for software development and upgrades are shared by multiple clients rather than solely by one client.
- In-house experience is not required, and niche employees do not need to be hired.
- Vendor handles the challenges of system up-time, reliability, security, and scalability.
- IT costs are maintained at a predictable level, and fewer expensive or specialized IT staff are required.
- Installation of heavy infrastructure is not required.

Some disadvantages to using an ASP model for an EDC system include the following:

- Clients must usually accept the software “as is”. Customizations usually do not occur unless several clients have made the same request and the vendor is willing to change the software.
- Loss of control of data: because the software is owned and maintained by the vendor, clients must ensure that proper service level agreements for system up-time and application availability are in place.

### **Technology Transfer**

In a technology transfer scenario, software under one company’s authority is moved to the environment of the sponsor or CRO. Many different levels of technology transfer are possible, ranging from transfer of just the build of a study to transferring all services from a vendor. Alternately, a sponsor or CRO may bring only certain services in-house, such as the help desk and user training. Traditionally, most companies handle building the study but not its

hosting. A sponsor determining whether to bring software in-house should consider the following questions:

- Is the sponsor ready to build studies internally?
- Is the sponsor able to provide hosting services?
- Can the sponsor provide help desk services to end users?
- How many trials using EDC are planned for this year and subsequent years?
- Does the sponsor have sufficient IT staff to provide technical assistance?
- Does the sponsor have trainers to provide necessary skills to users of the software?
- Does the sponsor have a dedicated project team to handle implementation?
- What is the overall scope/timeline for implementation of the EDC system, and can these deadlines and goals be met?

## **EDC System Capabilities**

Adopting an EDC system offers an opportunity to implement features and functions that enhance operations. To ensure success, software must be qualified and validated, and key features thoughtfully addressed.

### **Software Qualification and Validation**

Because CRFs are used by many end users, the functionality of each CRF should be tested and validated to ensure that data are entered, assessed, cleaned, committed to the study database, extracted and delivered in a known, regular, and repeatable fashion. Issues to consider when planning and executing the validation of an EDC system include the following:

- Whether the EDC application is “off-the-shelf” or custom-developed
- The amount of application validation that has been performed previously. A risk-based approach should be used to determine the level of validation to be performed.

- Notably, the amount of validation and qualification required for an EDC system can affect the time to start-up, the cost of start-up, site initiation and qualification, site maintenance, and software patch maintenance and deployment.

### **Support of Library Functionality**

Library functionality is the ability to reuse forms, fields, edit checks and other functions within EDC software. The ability to reuse pieces of a study for newly developed studies will allow you to gain efficiencies in the design and build process. Choosing a system or developing a process that supports a library of CRF components will greatly enhance the speed at which you can develop studies.

### **Electronic Investigator Signatures**

Prior to creating a study's process workflow, electronic signature capabilities of the EDC system being considered must be clearly understood. Questions to be answered may include:

- Can an investigator signature be applied at the form level, visit level, or casebook level? Is there a mechanism to easily tell which CRFs, subjects, or casebooks are awaiting signature, which have already been signed, and which have been edited since signing?
- Does the system send notification that a page is ready to be signed or has been signed? How is that notification delivered? Must the investigator be logged into the system to receive the notification?
- Can the investigator reject a request for signature and provide a comment with the rejection?
- Does the EDC study workflow require multiple signatures? If so, at what level (CRF, visit, and casebook)? Does the system have the ability to apply multiple signatures at this level?
- Does the system workflow automatically create signature notification based on specific status flags? If so, can this workflow be modified?
- Can a CRF be signed with open queries?

- Can frozen or locked forms be signed?
- Can the signature capability be turned off for a study?
- Is there an ad hoc search feature for all system users to filter by signature status flags? For example, not awaiting signature, awaiting signature, forms that have been signed, no longer signed, and so on. Can this search be performed at site level, visit level, by subject, and/or by specific form?

The ability to quickly search in this system interface may be particularly useful for CDM and CRAs when ascertaining a site's completion status at the end of a defined visit, and also at the end of the study.

### **Electronic CRF Archiving**

The method of transferring an electronic CRF to a read-only format for sites to use in the future must be determined. Factor the process for obtaining the CRFs into the contract's study timelines and expectations. See the chapter entitled "Electronic Data Capture—Study Closeout" for more information.

### **Export Formats**

Export formats as well as restrictions to the availability of exports should be documented. Possible formats for exported data may include the following:

- Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM)
- Microsoft® Access
- SAS®
- American Standard Code for Information Interchange (ASCII)
- Character delimited files
- Oracle®
- Extensible Markup Language (XML)
- Microsoft® SQL



The timing and delivery of exports is important, therefore the process for exporting and delivery of data should be robust and flexible.

## ***Integration***

Utilizing EDC has added complexity to data integration needs. Data managers must now understand how data collected or maintained outside an EDC system will be used, who will use it, and for what purpose it will be used. Knowing the answers to these questions will determine the path integration efforts must follow. In the event data integration does not occur as expected, a clearly defined roll-back plan should be established. To ensure project goals are met, the data manager must articulate these needs to technical or IT staff in clear terms. This section discusses considerations for various types of data integration that data managers may encounter during an EDC study.

### **Clinical Data Management System (CDMS) Integration**

Unless a fully integrated EDC or data management solution is being purchased, data managers must consider how an EDC system will integrate with new or existing data management systems. The EDC vendor may be able to help with integration through an add-on component specifically designed to meet the system needs. Some organizations should consider a custom solution that will involve technical and/or IT staff. Integration should encompass data and queries, while avoiding manual transcription of queries into the CDMS when automated edit checks occur in the EDC system.

Integrations should also consider the reporting needs for EDC data. Data from EDC, ePRO, an external vendor or other sources oftentimes must be viewed together to assess data quality. A third party reporting tool may be needed to achieve this, or your organization may need to rely on clinical programming or other support groups to merge data via SAS.

### **SAS® Integration**

The data manager, in collaboration with other functions, should decide whether EDC data will be directly integrated into the SAS® environment, or first integrated with a back-end CDMS.

## **ePRO Integration**

If patient reported outcomes will be collected via the Web, an e-diary device, or other data device, data managers should consider where and how data will be integrated with CRF data captured through the EDC system. Many EDC systems can import bulk data from external sources. If data collected using ePRO is of interest to the investigator, it may be worthwhile to upload ePRO data feeds into the EDC system. Integration of external data into the EDC system may also facilitate the archival and submission process by enabling all data to reside in one CRF. Consideration must be given to integrating ePRO data that has the potential to unblind the study.

## **CTMS Integration**

Integration of the EDC system and the CTMS can be a powerful way to gain efficiency in the conduct of clinical trials. Specifically, the data manager may want to integrate user account management. If site staff information is already being captured in the CTMS, this information may be transferred to either a help desk or directly into the EDC system, thereby eliminating manual creation of EDC accounts. Additionally, integration of visit information from the EDC system to the CTMS can facilitate monitoring and tracking of patient enrollment and completed patient visits. In turn, this information can be used to trigger site payments and grants. Integration of EDC with the CTMS also creates an ideal way to consolidate metrics used to assess overall trial performance.

## **Paper Component Integration**

If data is collected using paper CRFs, a method must exist to integrate these data with data collected using the EDC system. In most instances, data collected on paper is integrated into the back-end data management system. In some cases, it may be more appropriate to merge the data using a SAS<sup>®</sup> environment. Several EDC systems now also have the capability of integrating paper data entry into the same EDC database with EDC data.

## **Laboratory Data Integration**

Even if central laboratories are used, it is sometimes helpful to have all or key laboratory parameters available to site staff within the EDC system. The data manager must consider this need with the clinical team. Having all data stored

in one database can facilitate more robust edit checks across other CRF data in the EDC system.

### **External, Non-laboratory Data Integration**

If data such as an electrocardiogram will be received from external vendors other than central laboratories, data management should analyze the importance of data integration. As with ePRO integration, if sites require access to this data, the data manager should plan on uploading data into the EDC system. More information on this topic may be provided in other chapters.

### **Other Important Integrations**

As new technological tools are developed constantly, it is important to be mindful of other systems that may need to be integrated with an EDC system. In addition to the integrations discussed above, data managers should be aware of the need to also integrate an EDC system with coding, IVRS, and reporting tools other than SAS<sup>®</sup>. In the future, electronic health records (eHR) may also become an important consideration.

## ***International Study Considerations***

EDC systems are routinely used in international studies. The role of data managers in international EDC trials is similar to the role played in paper studies. However, planning is critical if the deployment of CRFs and hardware is to be completed prior to the first site initiation. Data management must work with clinical research to understand language needs of the CRF or any components of the CRF. Issues to consider include the following:

- Ascertain whether the local language can be used in a multi-national study. Many coordinators speak more than one language. Data management may avoid unnecessary work by asking this simple question or challenging the status quo in this area.
- Plan early for CRFs that must be programmed in multiple languages. Significant lead time is required to translate CRFs and verify translations.

- If applicable, ensure hardware deployment timelines are increased. Country specific laws may delay shipments significantly.
- Establish a plan to manage time zone differences, especially in relation to time and date stamping.
- Ensure hardware and software can be used at study sites, and that sites are prepared to use the tools that will be deployed to them.
- Develop a plan to manage system upgrades, which is particularly important if the system is being used 24 hours a day.
- Consider the wording of manual queries to ensure they will be understood by speakers of other languages.
- Consider issues posed by language barriers to staff training. For example, investigator meetings could provide simultaneous translation for all languages spoken by participants, a train the trainer strategy could be employed, or training materials could be translated into the users' native languages.

## ***System and Vendor Assessments***

For most organizations, moving to EDC is a significant decision, with an effect that is not limited to only data management. When launching an EDC system assessment, stakeholders from different parts of the organization are needed to develop the requirements checklist. Included in the requirements checklist should be:

- The types of data formats to export from the system
- The formatting and process for final archived CRFs
- The required software functionality, such as types of edit checks and the process for obtaining investigator signatures
- Details of service level agreements (SLAs)

This list should include a minimum of topics and the priority of each requirement (e.g., necessary or “nice to have”). A suggested method for determining overall requirements for the software and vendor is a grid to

which perceived values can be added. Additionally, identifying the availability of alternatives for some requirements may be useful, as selected vendors and software may not meet the requirements precisely.

The suggested list of minimum topics for the vendor/system assessment grid is as follows:

- About the EDC system:
  - User friendliness of the EDC system's interface
  - Study start-up process, including time, expectations, and what is included
  - Configuration limitations and the amount of customization that will be required
  - Hardware provisions and associated costs
  - Variable costs
  - Upgrade options and restrictions
  - Process for change management
  - Capability for establishment of a standards-based library
  - Reporting capabilities
  - Export formats available
  - Integration of IVRS and laboratory data (if needed)
  - Types of edit checks possible (e.g., cross-form, cross-visit, dynamic)
  - Handling investigator signatures
  - Process for data archiving at the end of the trial
- About the EDC vendor:
  - Stability of the vendor company

- Software and system validation and validation approaches
- Prospect of the vendor's continued existence for conduct of the intended study
- The vendor's offered help desk support, if needed
- Languages offered by the vendor for training of EDC users
- Languages offered by the vendor for support to EDC users

If a vendor is sought for the purpose of securing an engagement for more than one project or product, the grid should also include:

- The vendor's approach to addressing your product portfolio needs
- A comparison of the vendor's EDC tool suite and roadmap to the study sponsor's EDC strategy

Once developed, the grid can be evaluated to identify company requirements. The following sections will further discuss the requirements listed above.

For more information, see the GCDMP chapter entitled "Vendor Selection and Management."

## **Other Considerations**

At the point in time that you are still investigating a move to EDC, the following topics should be discussed with your prospective vendors.

### **Change Control Process**

These plans do not need to be fully developed prior to selection of a vendor, but you will want to know and understand that your vendor has a structured, detailed, and documented plan for change control. This includes change control for software upgrades as well as for midstudy amendments. A well developed plan should be in place for both types of change control. For a more detailed discussion of change control processes, see chapter entitled "Electronic Data Capture—Study Conduct".

### **Escrow Agreements**

Assess if there is a need for your company to have an escrow agreement put in place with your vendor. Some companies offer source code escrow services that assure the Licensee continued availability and usefulness of the software in the event that the software vendor fails to maintain the license agreement, or becomes insolvent.

### **Related Services: Hosting and Help Desk**

A critical factor in the success or failure of an EDC study is the technical support received by users encountering problems with the system. Technical software support is often managed through a help desk. If using an outside vendor, standards and expectations for all trials using the vendor's help desk software should be documented in the vendor's contract for this service.

As the EDC market expands, EDC vendors continue to add functionality that makes their solutions unique. However, most vendors offer the services of hosting and help desk. How services are provided and fees for services can be very different among vendors.

The basic questions of how the help desk will be handled needs to be discussed with your team at this stage of deployment. You may consider bringing EDC help desk functions in-house with your organization, or you may prefer to use an outsourced help desk. The outsourced help desk could be facilitated by the EDC vendor you choose, or you could enlist the services of a technical call center or help desk. The sponsor's decision to use an internal or external help desk is primarily determined by the amount of internal resources available to perform this function and the level of response time or needs that the study or project dictates. For example, if an EDC study is of a very simple design, of short duration, and has a small number of sites, the use of an internal help desk may be the best choice, so long as qualified staff are available to take inbound calls. Internal help desk agents will have a better understanding of the company's policies, procedures, and protocol design elements. As good service will have a direct impact on their employer and future business with that particular user or site, internal staff will also be more interested in determining the quickest and best way to resolve an issue.

A global study with numerous sites, complicated and varied forms, and new users represents an opposite example. Use of an outside help desk could

provide advantages, such as covering additional time zones and languages without taxing internal resources. Moreover, external help desk agents are typically evaluated on performance of help desk ticket resolutions, and will have a vested interest in being courteous to users. However, they will not be able to address clinical-related questions, which will need to be forwarded to internal resources.

## ***Vendor SLAs and Performance Reports***

When using an external vendor, the sponsor must emphasize the writing and managing of service level agreements for performance. A contract should be established between the sponsor and vendor providing help desk services (such as the EDC vendor or other outsourced agency). This contract and/or SLA should include, but is not limited to, the following identified functions and associated costs:

- Core languages covered
- Translation fees for additional languages
- Vendor project management fees
- Portal or other web access to see open/resolved calls and problem resolutions
- Computer telephony costs
- Determination of fees based on studies vs. by site or by call
- Documented allowable number of calls per month
- Documented allowance for number of inappropriately handled calls, which should be no higher than 4%–6% of all calls
- Study setup fees, if applicable
- Documented percentage of expected system uptime

Several reports can assist with the management of a help desk, especially an external help desk. These reports include:

- Aging reports



- Escalation reports
- Summary of activity per week
- Pareto analysis of problem areas to address (Pareto analysis is a statistical technique used for selection of a limited number of tasks that can produce a significant overall effect. It is based on the Pareto principle, which says that by doing 20% of work one can generate 80% of the advantage of doing an entire job.)
- New tickets per week report
- Ticket closure patterns

The main point to remember at this time is that technical support for end users is crucial to the success of your move to EDC. Ensure that your first level of help desk coverage is available to all users, has enough language coverage to accommodate your sites and that the hours of support are sufficient for your user community.

### ***Detailed Help Desk Planning***

Once an EDC vendor is selected, roles and responsibilities of the help desk staff and the sponsor's staff must be established. Consideration should be given to the number of help desk staff available. The number of studies conducted by the sponsor will determine the number of help desk staff required and indicate whether help desk services should be provided by an external party. If only a small number of studies require support, it may be feasible for the sponsor to provide help desk support with internal staff. The timeframe during which users should be able to contact the help desk must be considered. Typically, 24/7 coverage is not required unless the EDC system is deployed globally. When the help desk is provided by an external organization, service level agreements should be established concerning the timeframe in which each call will be answered, as well as any other metric your organization feels is important.

Software support is commonly separated into different levels or tiers based on the technical expertise needed to correct the issue. Tier 1 software support is the lowest level of support needed and includes activities such as unlocking

user accounts and resetting user passwords. Because it is the most common support required by users, tier 1 software support is vital to a study. Users that require this level of support are often unable to access the system in any fashion. Therefore, to minimize the negative impact to both users and study conduct, it is critical to provide assistance to these users as soon as possible.

For various types of anticipated user issues, clear escalation paths must be identified for second and third level support. Data managers frequently serve as the second or third level of help desk support for EDC studies. The most common issues escalated to data managers are trial-specific data entry or query resolution issues. The data manager should be prepared to discuss the problem's solution with the level one help desk agent or with the user directly. This new role may require multilingual expertise from CDM. This new role also strengthens the relationships between CDM, clinical researchers, and sites. Cooperation among all three parties may be required to solve problems related to the EDC system. To ensure that users are satisfied with the EDC system, CDM should ensure that help desk escalation procedures are followed and working correctly.

Examples of issues and their required level of support include:

- Account activation: usually requires only level one support
- Technical error messages: may require level two or level three support
- CRF design issues: level two or level three support is typically provided by CDM
- Data entry issues: sometimes may be handled by level one support, but many are escalated to CDM as a level two or level three support issue
- Supporting systems issues: usually escalated to the IT department
- Query resolution issues: usually escalated to CDM

Deployment of computer equipment and Internet connections can also be handled by your help desk. However, these services can drain resources in day-to-day operations and involve the complexities of international shipping and tracking.

In addition to the issues listed above, data managers should ensure that specific information is provided for issues concerning account management, tier one software support, and requirements for multilingual capabilities.

### **Tier-One Software Support**

Steps must be taken to ensure that training materials for help desk staff are complete and clearly identify the correct issue-escalation procedures. Ideally, help desk staff should be trained on usage of the EDC system. For each study, data management should provide help desk staff with a document outlining study-specific areas of concern, such as common issues with data entry encountered by users. This document will enable help desk staff handle calls more efficiently and will minimize issues that are escalated to data management.

Tier 1 support may be needed whenever a user attempts to access the EDC system and encounters a problem. Therefore, the support center must be available whenever users will access the system. At a minimum, standard business hours should be represented (e.g., in the United States, 9:00 AM to 5:00 PM), but even determining what standard business hours are for specific users can pose a challenge since sites can be located in different time zones or countries. Another consideration is whether or not support will be available on weekends and holidays. While the gold standard for support availability is 24 hours per day, 7 days a week and 365 days per year (24 x 7 x 365), this ideal may provide significantly more coverage than is needed and unnecessarily increase help desk costs. As with language localization, help desk availability must to be determined prior to the start of the study.

### **Providing Toll-Free Support**

Tier 1 software support most commonly involves individual users contacting a support center or help desk for assistance. To ensure convenient access to technical assistance, users should be provided with a toll free phone number or calling cards to contact the help desk.

### **Multilingual Capabilities Required**

To handle calls in international studies, the help desk staff should be fluent in applicable world languages.

To be effective, the help desk managing tier 1 support must be able to successfully communicate with all system users. Clinical studies are frequently multilingual and it cannot be assumed that all users will be conversant in English. Several options are available when determining how multiple language needs will be addressed. One option is for the help desk to fully support any and all languages with on-site support staff. Another, more frequently used option, is for the help desk to support several of the more common languages with on-site staff and to make use of a translation service to provide access to translators fluent in languages less likely to be needed. If providing multilingual help desk support is not possible, CDM representatives should discuss this issue with the CRO local to the international site. In many cases, monitoring staff may be fluent in local languages and can handle certain types of support.

### ***Gap Analysis between Existing SOPs and EDC Requirements***

It is critical to determine how implementation of an EDC system will require changes in the sponsor's current set of SOPs and other controlled documentation. Identifying these gaps is an effort of technical and clinical operations that must be shared among all stakeholders.

At a minimum, requirements should be written for each new process that has the potential to impact study data. These requirements must later be tested and will form the basis of validation efforts. Functional requirements must be developed to test overall functionality of the solution, and business requirements must be developed to test how the solution meets needs of the sponsor. Examples of procedures and processes for which requirements, testing, and validation should be performed include data entry, data verification, discrepancy management, data lock, user roles, user security and privileges, data reporting, subject freezing or locking, database backup and recovery (if not covered elsewhere), financial reports, study design of data objects, edit check procedures and derivation procedures.

- In conjunction with the set of requirements you establish for your EDC solution, you can also use the metrics and performance targets you have already established. These can aid in analyzing the set of SOPs that must be modified to include new practices for EDC. The primary stakeholder should be responsible for driving each new or updated SOP, however all

impacted stakeholders should have review and approval status before the SOP is put into effect.

Requirements for updating SOPs and the study process for EDC include:

- Identifying metrics and performance targets
- Performing a gap analysis between current SOPs and requirements for EDC

Data management may also establish goals for EDC projects based on calculated return-on-investment. However, most organizations will find it necessary to modify their processes to accommodate EDC during the start-up phase. It should be expected that the start-up phase will be iterative and will be impacted by many variables, including:

- The complexity of projects implemented
- The variation between projects
- Requirements for user training
- The type of EDC system implemented
- The number of staff affected by the EDC transition
- The preparation required by each site

This transition phase of EDC initiation may be planned through careful selection of the first several projects that will implement the EDC solution. Limiting the number of projects to use EDC will enable the sponsor to transition smoothly into the solution and manage expectations of stakeholders as necessary. Using this model, the sponsor may identify discrete phases of EDC implementation and formulate each phase to increase the complexity of projects using EDC. Ideally, a review step can be established at the end of each phase to inform stakeholders of the structure and expectations of subsequent phases. This review step should include all stakeholders and should analyze how closely the phase met set targets for metrics and performance.

## **Staffing Evaluation and Staffing Change Plans**

Before initiating an EDC trial, a sponsor should carefully compare the resources needed to manage the people, technology, and processes to the resources that are currently available. Any deficiency in resources should not be underestimated. Initiating a trial using EDC without identifying and providing necessary components may result in failure to meet study objectives in terms of time and cost. Moreover, the study team (including the sponsor, contract staff, and site staff) may be negatively affected. To ensure an EDC study is achievable in terms of cost, time, and quality of deliverables, the sponsor must commit to meeting the study's resource and training needs.

Resources available to the sponsor include staff (and their skill sets), established processes, and vendors. The staffing evaluation plan should analyze the sponsor organization for any gaps between available resources and the resources needed to conduct and manage the EDC study. During this analysis, the following issues should be considered:

- Will a vendor or the sponsor provide the EDC system?
- Who, within CDM, will serve as the liaison with project management to approve the EDC system's design and oversee its production?
- Who will be responsible for testing and validating the system: CDM, a clinical project manager, a clinical research associate, or a targeted site participant?
- Who will provide help desk support to system users? To what level? What type of questions will be answered?
- Who will be responsible for training staff and maintaining training records? How often will training be required?

After assessing the staffing needs to conduct an EDC trial, the sponsor should evaluate necessary changes, if any, that are needed within the organization. The role of CDM may change from a task manager (managing the data itself) to that of a project manager (facilitating data and communication flow between all other study participants). CRAs may take on more responsibility understanding the technical aspects of data collection, training site staff, and answering questions from site staff about the EDC system. Information technology (IT) and programming resources may also have increased

visibility to other study team members, especially during the EDC system configuration, EDC suitability at prospective sites, and/or the help desk communication to resolve user questions.

To facilitate these changes in staffing structure, the sponsor should:

- Perform and document a process flow evaluation addressing all study team members in the workflow, communication plan, information flow, and reports.
- Perform an evaluation of required staff and their skill sets as compared to required resources.

### ***Metrics and Performance Targets***

Ideally, performance targets set for EDC projects will be based on the sponsor's foundational reasons for switching to EDC. These targets should represent the first level objectives for EDC projects. The next set of objectives can be developed during rollout of the EDC solution and should include feedback from all stakeholders.

Data management must also identify any additional metrics that may not be applicable for paper-based studies but that will be needed for EDC projects. Examples of EDC metrics may include average time for discrepancy resolution by site, average number and severity of help desk calls, and percent of EDC system downtime.

For EDC and paper-based studies, sponsors need to determine which metrics reports are required and establish processes for collecting, analyzing, and reporting the metrics data. Decisions need to be made to determine reporting frequency, at which level metrics need to be reported (e.g., by phase of study, by site, by trial, by therapeutic area, by region, etc), and who is responsible for reviewing and assessing data results. All metric reports must be clearly defined, with definitions clearly understood by all individuals reviewing and making decisions about data. These reports should be validated against the sponsor's computer system policies and procedures, and should be standardized so they can be used for multiple trials.

The following is a list of minimum recommended EDC metric reports:

- Study build timeline metrics
- Number of subjects screened and/or enrolled
- Subject visit date to data entered in system for that visit (measure in number of days)
- Current status and history of SDV
- Number of queries outstanding (measure in number of days)
- Percent of data/visits clean
- Number of queries per site (manual and automated)
- Query frequency counts per data element
- Time from last patient visit (LPV) to all data entered and/or data cleaned
- LPV date to data lock date

### ***Reports to Support Study and Process Management***

Data are more readily available in studies using an EDC system. This availability is particularly advantageous as it enables the sponsor to be more active in managing data management workflow processes and timelines, as well as site progress and performance. Early data reporting capabilities enable the sponsor to perform more timely assessments and take action to drive productivity, improve site performance, and reduce overall study timelines.

Just as in paper-based trials, creating report and listing specifications should be a collaborative effort between CDM and other members of the research team, especially those responsible for clinical, statistical, and safety-related functions. Report and listing specifications should be documented in a data review or data management plan that clearly indicates who is responsible for reviewing listings and reports. This plan should also clearly indicate the frequency of review and what action may be taken (e.g., creating manual queries, contacting the site, or retraining the site).

Performance targets and goals need to be established at the organizational level and individual study level, as discussed in the previous section devoted



to metrics and performance targets. Expectations based on these targets need to be communicated to both the study team (e.g., data managers, monitors) and site staff (e.g., study coordinators, principal investigators) to drive improvements in data management processes, trial timelines, and site performance. Due to the technology differences, metrics goals and expectations for EDC studies can be more aggressive than for paper-based studies. For example, in an EDC study the overall study goals for query turnaround time and subject-visit-date to date-entered-in-system can be shorter and more aggressive. This difference is also true for data-management process metrics. For example, goals for the time from LPV to data lock can be shorter on average in an EDC study than in a paper-based study, because the data can be received, reviewed, and validated in a much more timely fashion.

To ensure efficient data cleanup activities, it is recommended that reports that aid in data cleaning be identified during the pre-production period, as well as reports necessary to gather metrics about study conduct processes. Reports which capture protocol deviations should be programmed at the beginning of the study and run frequently to monitor compliance. Tasks that should be completed include the following:

- In addition to a project plan, create a flow chart that outlines each report deliverable and the person responsible for approving that report design.
- Schedule meetings to review and obtain feedback on reports to be used during the study.
- Determine the metrics important to the study and design reports to capture these metrics.
- If the study involves a CRO, consider what reports will be required from the CRO.
- Define CRO performance requirements and design a report to track performance of and provide feedback to the CRO.

### ***CDM Deliverables During Study Start-up***

The following section concerns CDM deliverables during the start-up phase of an EDC study, including CRFs, edit checks, and coding.

## **CRFs**

In an EDC system, electronic CRFs replace traditional paper CRFs. However, data captured through electronic instruments or computer software is not immediately considered electronic CRF data. CDISC defines an electronic CRF as a CRF in which related data items and their associated comments, notes, and signatures are linked electronically.<sup>6</sup>

The process for designing CRFs in an EDC system is integral to study start-up. Development of an electronic CRF is more complex than simply modeling a paper CRF in a word processor. Creating an electronic CRF entails designing a truly electronic form with user interface elements that reduce challenges posed by electronic data entry, as well as facilitating data collection to improve data quality. Interface elements such as check boxes, option buttons, and menus enable users to record data less likely to be queried, which is a goal of EDC.

Although organizations such as Clinical Data Interchange Standards Consortium (CDISC) have initiatives underway, such as CDASH, there are currently no widely adopted standards for electronic CRFs, and those standards that do exist are evolving.<sup>6</sup> However, as with paper-based CRFs, standardization of electronic CRFs by the sponsor can:

- Facilitate data exchange
- Remove the need for mapping during data exchange
- Enable merging of data between studies
- Allow consistent reporting across protocols and projects
- Enhance monitoring activity and investigator staff efficiency
- Provide increased efficiency in processing and analysis of clinical data
- Provide capabilities not traditionally available when using paper-based CRFs
- Promote reusability of CRFs across studies through development of CRF libraries

A well-designed EDC system should assist site staff with accurate entry of study data. An EDC system can be designed to guide site staff to appropriate forms and to correctly enter data into the fields of those forms. The design of CRFs should avoid the following shortcomings:

- Using multiple pages for an CRF which could be displayed on one page
- Requesting an excessive amount of information on one page
- Using unfamiliar jargon
- Using checkboxes that do not include every applicable choice
- Using codes that are only relevant to data processors
- Requiring overly complex edit checks

The following practices should be followed during the design of electronic CRFs:

- The protocol should determine what data should be collected on the CRF.
- All data must be collected on the CRF if specified in the protocol.
- Data that will not be analyzed should not appear on the CRF.
- Data required by regulatory agencies should be collected.
- Data questions should be clear and concise.
- Duplication of data should be avoided.
- Use of free-text responses should be minimized.
- Units should be provided to ensure comparable values.
- Instructions should be provided to reduce misinterpretations.
- For each question, choices should be provided to enable summary generation by computer.
- “None” or “Not done” should be available as answer options when applicable.

As mentioned previously, standards for electronic CRFs are still developing. Generally, an EDC system should be flexible enough to capture questions as they would appear on a paper CRF. Data management should be dedicated to identifying and maintaining standards for the design and functionality of CRFs to be used in the EDC system. Prior to a study's production release, a process should be established to ensure that development of CRFs adheres to these standards.

### **Multilingual CRFs**

Even when an EDC system has multilingual capabilities, translation of text appearing on CRFs can be a time-consuming process. Data management should determine whether English is the best language to be used for text appearing on CRFs. Because many studies are already conducted in English, most sites do not object to the use of English for CRFs. However, data management should work with each site to evaluate its multilingual requirements. Translation should be considered for printed and electronic user manuals, as well as training materials supporting the EDC system.

Although users of the EDC system may speak English or have years of English training, they may still misunderstand text that appears on an untranslated CRF. Idioms that may cause confusion or may not translate clearly should be avoided. A medical linguist may be required to translate certain terminology. Ad hoc forms should be presented in the local language of the applicable site. To check the quality of translations, back-translation by a third party should be used.

### **Dynamic Forms**

Dynamic forms appear only when a subject meets a certain criterion, or when a particular data point is entered. A common example of a dynamic form is a form for pregnancy information which only needs to be entered when the patient is female. However, because dynamic forms do not always appear to be available in the EDC system (e.g., the pregnancy form will not appear in the system if a patient is male), they have the potential to confuse users. For example, the sex of a patient might be entered incorrectly as female and subsequently changed to male. In this case, implementation of dynamic forms on the EDC system determines what happens to the pregnancy form and its data, which is no longer needed for the male patient.

Determining how to best implement dynamic forms depends on capabilities of the EDC system and complexity of the study. Clarity with the functionality of dynamic forms can be achieved through the following practices:

- Keep functionality of dynamic forms simple; sometimes the ability to have dynamic forms can create too many permutations, and may frustrate users.
- Ensure that the development team understands the challenges of designing and implementing dynamic forms.
- During validation and qualification of the EDC system, test the design and implementation of dynamic forms by entering “incorrect” data for a patient and subsequently changing it.

### **Dynamic Visit Structures**

Dynamic visits are similar to the dynamic forms discussed in the previous section. However, instead of forms, visits become available based on data entered for a patient. In an oncology trial for example, when a patient meets a certain criterion he or she may move to a different treatment group. Dynamic visits enable this type of capability, and the same best practices should be followed as for the design and implementation of dynamic forms. Dynamics may impact data entry efficiency and system speed, so data managers should be aware of the possibility of overloading sites with confusing or complicated dynamic functionality.

### **Derived Variables**

EDC systems can provide derived variables. This may be helpful for sites, as some data will not need to be entered by site data entry personnel. Some commonly used derived variables would be conversions from one measurement system to another (pounds to kilograms or inches to centimeters) as well as averages or computations from other entered fields.

It is necessary for CDM to communicate how these features work and help users understand the impact on monitoring. This holds true for dynamic forms as well as derived variables.

## **Edit Checks**

The use of edit checks in an EDC system offers data managers a unique opportunity to resolve data issues by interacting directly with clinical site coordinators. Due to the ability to gain access to data shortly after it is entered, data managers and clinical personnel can initiate issue resolutions with site staff in a more timely manner. For example, direct contact by phone with site staff promotes an active approach to completing a CRF and resolving edit check issues. Direct contact with site staff also promotes an active approach to completing a CRF and its edit checks within a short period of time from each other. Moreover, during study development, data managers can truly collaborate with database developers when programming edit checks. The technical nuances can be explained by the database developer, and the data manager can provide necessary data management principles to ensure implementation of a functional edit check.

The approach to programming edit checks should be chosen during development of the EDC database specification, and in consultation with all stakeholders involved in data validation. To define and review edit checks prior to production release of an EDC study, data managers coordinate activities of clinical, IT, quality control, quality assurance, and other groups. This is essential to the correct functioning of edit checks in an EDC system.

### **Edit Checks in an EDC System versus a Back-end CDMS**

The approach to programming edit checks depends on the architecture of the EDC system, which can typically be described as either of the following:

- An EDC front-end data capture system with a robust back-end CDMS. With this architecture, an analysis of whether to program the edit checks within the EDC data capture system and/or in the CDMS should be determined.
- Complete EDC data management system where all edit checks must be programmed in the EDC system

The following considerations concerning the EDC system architecture should also be made in determining the approach to programming edit checks:

- How complex is the edit check? Will performance of the EDC system be adversely affected if programmed in EDC? Suppose the data manager

needs to confirm that the last date of subject contact is the last chronological date in the database. In this case, the edit check program should pull all dates from each module in the database and compare those dates against the date of last contact. This type of edit check might access the underlying database thousands of times and noticeably degrade server response times.

- Are all the data available in the EDC system? For example, if coding of terms occurs on the back-end CDMS, edit checks requiring coded terms should be programmed on the CDMS.
- Will programming back-end edit checks require manual re-entry of data into the EDC system for query resolution? The resources needed to manage this activity should be considered.
- Is a reporting database structure better suited to handling complex edit checks? For example, if an edit check is too complex, it may be best handled through listings with manual re-entry of the query into the EDC system for resolution by the site users.

In both EDC system architectures (and for paper-based studies as well), consideration must be given to the edit check specification as a whole. However, in an EDC system some issues impact the site user rather than the sponsor. For example:

- Are all of the data elements needed to properly open a given edit check actually collected in the study? If all data are not present, a query may fire that cannot be closed without entry of a specific data point.
- When adding midstudy edit checks, consider any limitations the system might have. For example, edit checks added midstudy may only activate as a result of new or modified data. Therefore, the data manager should consider programming a listing to identify issues with existing data. In this case, sites should also be informed that they may be required to resolve issues identified in earlier visits.
- Edit checks that are more study specific (not standard across trials) may generate queries due to the timing of data entry. EDC systems using dynamic forms may cause such queries. Consider providing additional

training on this issue to sites, or program edit checks on the back-end CDMS rather than on the front-end EDC system.

### **Hard vs. Soft Edit Checks**

In addition to planning where edit checks should be programmed in the architecture of an EDC system, consideration should also be given to potential types of edit checks and corresponding user responses. Edit checks in EDC can be classified into two broad categories, “hard” edits and “soft” edits.

Soft edit checks are usually cross-panel or cross-item rules programmed to allow data to be entered into the system, but checks for consistency upon data entry. If an inconsistent or missing data item is identified by the edit check, a visual indicator (e.g., color change, iconography) indicates that a new query exists and on-screen text prompts site staff to address the query. For soft edit checks to be programmed correctly, the data manager should clearly identify fields on each form for which data must be entered.

Hard edit checks can be classified as “browser” checks or “system” checks.

- Browser checks prevent entry of data that is inconsistent with the data item being collected. If a user attempts to enter inconsistent data, submission of the form will be prevented until the inconsistency is addressed satisfactorily.
- System checks prevent entering data that do not match form and/or item property settings. For example, when a field requires a number with 2 decimals, a value of “3” cannot be entered. Instead, “3.00” must be entered to satisfy the property requirement. A system check does not produce a query or present an error message to the user. As they can disrupt the data entry flow at the site, system checks should be used only when deemed necessary.

## **Coding**

Decisions on how to handle coding of medications, adverse events, procedures, and other study data should be included in the specifications documents and in parallel with CRF development. The role of data managers in the coding process for an EDC study should be relatively unchanged from the coding process for a paper-based study. However, the process should be adapted to the technology used to perform coding.



The following best practices for coding on EDC systems should be followed:

- Data management should work with the pharmacovigilance and drug safety group to determine how data coding should be handled. For drug safety and clinical trials, coding should be handled centrally by data management. Alternatively, coding can be coordinated between drug safety and data management.
- During the CRF development process, all data fields to be coded should be identified.
- The capability of the EDC system to support coding should be understood. If the EDC system cannot handle coding, data management should establish a process to code study data on the back-end database.
- If the EDC system is capable of handling coding, the sponsor should decide whether the user should be able to see coded terms or only the reported verbatim terms.
- Ensure the clinical team understands who will be proposing terms for coding failures. For each study, it is recommended the data manager handles this activity.
- The coding team should review the design of electronic CRFs to ensure optimization for coding purposes. For example, CRFs frequently provide menus for the coordinator to enter terms. The medical coding team can assist with development of these menus so that available terms will code appropriately.

### ***Site Evaluation and Qualification***

The process for initiating an EDC study is not just a matter of the sponsor selecting an EDC vendor that meets certain business requirements. The sponsor must also consider the pricing model, management, deployment, and implementation of the vendor's EDC system. The sponsor is responsible for ensuring that sites are assessed and qualified to use hardware and software required by the EDC system. Site evaluation and qualification by the sponsor and EDC vendor must occur during start-up activities and must be seamless, especially when EDC is being implemented at a site for the first time.

The following sections detail criteria for evaluating and qualifying a site's readiness to implement an EDC system.

## **Evaluating Site Technical Capabilities**

Aspects of the site's technical capabilities may include the following:

- Presence of a wired or wireless network: If trials are run at remote site locations without Internet connectivity, installation of an offline data capture application may be necessary.
- Location
- Physical space (if required for deployed hardware)
- Language(s) spoken by site staff: Software configuration may be required to accommodate non-English speakers
- Compatible software, hardware and bandwidth
- Experience of site staff with EDC software: Staff may lack experience or exhibit hostility toward EDC. Users that are technically challenged may require personalized training. For a site that is new to EDC, it may be useful to identify an internal champion who can facilitate adoption of the new system.

## **Evaluating Site Connectivity, CRA Connectivity at Sites**

- Availability of a dial-up or high speed connection
- Proximity of the site's physical location to the EDC system's server
- Capability of the site to access and synchronize with the EDC system on a scheduled basis

## **Site Provisioning, if Necessary**

- Equipment and hardware (e.g., laptops, PCs, phones)
- Training
- Training manuals

- Access IDs (e.g., user account set up using secure access IDs)

### **Site Hardware and Broadband Provisioning if Necessary**

Good clinical practices advise that site assessments be completed, including current trends and practices. These assessments are performed to ensure that, prior to study initiation, sites selected for a study are completely prepared to enroll patients. Although site assessments are not required, not assessing a site could be very costly—a site may have qualified patients and learn that they do not have the equipment and/or knowledge to enter patients correctly within an EDC system. Learning this too late would be detrimental to the site, study, and sponsor. Hardware and broadband provisioning can be undertaken by the sponsor, CRO, or can be outsourced to a 3<sup>rd</sup> party provider that specializes in this area.

### ***End User Preparation***

This section concerns activities that should be conducted before the study begins to ensure staff are prepared to use the EDC system.

### **Setting System Rights Determined by Roles and Privacy**

Internet-based access to an EDC system presents additional challenges that must be clearly documented to ensure security and confidentiality of study data. The browser must ensure that all data connections cannot be breached or corrupted by an unauthorized user or external software. All audit trails detailing user access must remain unmodified and intact. User management begins with the sponsor's evaluation of the roles and responsibilities for each task within the system, based on criteria outlined in staff evaluation plans. Because the EDC system is used by different site staff and sponsor team members, access needs to be considered for all. Where input or review of data is required within the system, user roles and responsibilities should be defined and documented to identify specific access privileges or rights. Factors to be considered when defining these user roles include the following:

- Data entry rights by both site staff and sponsor team members. For example, dictionary coding requires that sponsor staff be able to enter or modify certain fields on a form. To ensure that the integrity and reliability of data is maintained, sponsors should carefully consider which fields will

be modifiable by the sponsor team. If sponsors will have such access, clear process documentation and a robust audit trail are also critical.

- Investigator signature rights
- Query generation—for example, in-house review by CDM versus reviews performed at the site by monitors
- Query resolution—for example, sites may only be able to resolve specific types of queries, while CDM can close queries after reviewing site responses
- SDV rights
- Read-only access—for example, blinding and patient privacy regulations may require user access to be limited to only certain CRFs
- Report creation, generation, or view-only access at both the site and by the sponsor should be considered. Some possible scenarios include limiting access so that each site can only generate reports for their subjects, limiting report generation across countries or regions, or limiting report creation to CDM staff who have received more advanced training.

## **User IDs and Passwords**

Conventions for user IDs and passwords need to be determined up front in study planning. In addition, processes for dissemination of IDs and passwords to users must be established. These processes should include tracking that users have been properly trained prior to receiving access to the system. The system should force users to change their password at first log-in.

Training or system documentation should educate users as to the rules and regulations on keeping user ID and password information confidential, as well as requirements for changing their passwords. Lastly, the training materials should instruct users on what to do should they lose or forget their ID and/or password.

## **Account Management**

Data management should participate in designing the account management process so they can train clinical staff on how they and their site coordinators

will obtain access to the EDC system. The process should minimize the number of manual steps that are included. Consideration should be given to linking the CTMS to the account creation and activation system, thereby eliminating the need to transfer user information between systems.

The typical account activation process is as follows:

1. A user is trained and authorized to be granted access to the system.
2. The user calls the help desk to request activation of his or her account.
3. The help desk confirms that EDC training has been completed by the user.
4. The help desk creates the account and assigns a temporary password to it.
5. The help desk guides the user through the process of logging on to the EDC system and selecting a new password.
6. The user confirms access to the EDC system.

### **Training Prior to System Access**

If study team members and site staff are not fully trained to use the EDC system, they are unlikely to use it properly. Therefore, users must complete required training before being provided access to the EDC system. If possible, a certification exam can be included at the end of training to certify competence. Certification forms should be given to trainees as appropriate.

User training on both the system and study setup within the system is important. There are various views on the extent to which these two components should be included in the training plan. At a minimum, each user with the ability to modify study data should have documented training on basic system functionality, such as logging on, opening a CRF, entering data, and responding to a query. User training can be provided through methods such as the following:

- Self-study: reading materials, e-learning materials, using sample forms in a training environment

- Training environments that provide training exercises with examples that are generic or customized to the study-specific workflow
- Web-based instruction or demonstration
- Face-to-face training: Conduct training for users in a central training facility, such as investigators' meetings or other centralized training meetings

### ***Recommended Standard Operating Procedures***

- EDC Design Specifications
- System Setup, Installation and Support
- EDC Training
- Medical Coding
- Data Collection and Handling
- Data Backup, Recovery, and Contingency Plans
- Data Review and Validation
- Prequalification Requirements including 21 CFR Compliance
- User Access Creation, Modification and Revocation
- Systems and Hardware Security
- Guidelines for Outsourcing with Vendors/Vendor Management
- Handling External Data
- Coding Medical and Clinical Terms

### ***References***

1. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice*. Geneva, Switzerland: Author; 1996.

2. Code of Federal Regulations, Title 21, Part 11, Electronic Records; Electronic Signatures. Washington, DC: US Government Printing Office; 1997.
3. US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Washington, DC: US Department of Health and Human Services; 2007.
4. Khosla R, Verma DD, Kapur A, Khosla S. Efficient source data verification. *Indian Journal of Pharmacology*. 2000;32:180–186.
5. Schuyt, Malcolm L, Engel, Thim. A review of the source document verification process in clinical trials. *Drug Information Journal*. 1999;33:789–797.
6. CDISC (Clinical Data Interchange Standards Committee) web site. Available at: <http://www.cdisc.org/>. Accessed on March 31, 2008.

### **Further Reading**

*Code of Federal Regulation, Title 21, Volume 1, Part 11: Electronic Records, Electronic Signatures*. Washington, DC: Government Printing Office; 1998.

*Data Basics*. 2001:Summer;8–9.

Drug Information Association. *Workshop on “Electronic Data Capture: Technologies, Implications and Future Requirements”*, November 1998. Baltimore, MD: Author; 1998.

eClinical Forum PhRMA EDC/eSource Taskforce (Formerly the Electronic Data Management Forum (EDM Forum). *The Future Vision of Electronic Health Records as eSource for Clinical Research, Version 1.0, September 14, 2006*

Electronic records; electronic signatures. *Federal Register*. 1997;62;13429–13466.

Hopp DI. Three topics integral to the use of the Internet for clinical trials: Connectivity, communication, & security. *Drug Information Journal*. 1998;32; 933–939.

Hyde AW. The changing face of electronic data capture: From remote data entry to direct data capture. *Drug Information Journal*. 1998;32;1089–1092.

Hyland ME, Kenyon CAP, Allen R, et al. Diary keeping in asthma; Comparison of electronic and written methods. *British Medical Journal*. 1993;306;487–489.

*IIR Symposia on “Automated Data Capture”, February 1998, Orlando, FL and September 1998, Washington, DC.*

Kubick WR. The elegant machine: applying technology to optimize clinical trials. *Drug Information Journal*. 1998;32;861–869.

Kush RD. Electronic data capture: A survey. *The Monitor*. 1999:Fall;37–40.

Lampe AJ, Weiler JM. Data capture from investigators’ & sponsors’ perspectives: Balancing quality, speed, & cost. *Drug Information Journal*. 1998;32;871–886.

Latham D. EDC: Establishing standards for electronic data capture. Interview of K. Vaillant. *Pharmaceutical Visions*. 1998:Autumn.

Ma, JM. *A Modeling Approach to System Evaluation in Research Data Management*. Unpublished doctoral dissertation. Chapel Hill, NC: University of North Carolina; (1986).

Marks, RG. Validating electronic source data in clinical trials. *Controlled Clinical Trials*. 2004;25:437–446.

Palm U. Controlling the increased complexity of electronic clinical data collection and management technology by data modeling and system engineering. *Drug Information Journal*. 2002;36;683–692.

Regalado A. Re-engineering drug development, II: Clinical data collection and management. *Start-up*. 1998:27.

Stokes T. Computer systems validation (6-part series). *Applied Clinical Trials*. September 1996, January 1997, February 1997, April 1997, June 1997, August 1997.



US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Washington, DC: US Department of Health and Human Services; 2007.

US Food and Drug Administration. *Guidance for Industry: Electronic Submissions of Case Report Forms (CRFs), Case Report Tabulations (CRTs) and Data to the Center for Biologics Evaluation and Research*. Washington, DC: US Department of Health and Human Services; 1998.

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2003	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
September 2008	Revised to reflect the orientation of chapter towards the concept and start-up phase of EDC. Content updated and organization of material revised. Study conduct and study closeout content moved to separate chapters.

**This page is intentionally blank.**

---

# **Electronic Data Capture—Study Conduct**

September 2008

## ***Abstract***

As electronic data capture (EDC) has become a more common and proven tool for clinical trials, understanding the principles and guidelines for EDC use has become more important for clinical data management (CDM) professionals. This chapter reviews processes and regulations that currently apply to EDC during the conduct of a study, and emphasizes the role CDM professionals have in properly maintaining an EDC system within an ongoing study.

## ***Introduction***

Electronic resources for clinical data management have developed over the last 30 years as a suite of processes and tools to enhance the management, quality control, quality assurance, and archiving of clinical trial research data. This development has led to a major paradigm shift in data management, with data capture now capable of being facilitated at investigator sites, and data transfer being expedited by Internet technologies.<sup>1</sup>

While pre-production activities and planning are crucial for a study employing EDC principles (see chapter entitled “Electronic Data Capture—Concepts and Study Start-up”), it is also vitally important to apply proper data management principles to the ongoing conduct of a study. Clinical research is a dynamic process, and clinical data managers must be prepared to adapt as needed to best serve the needs of a study.

## ***Scope***

This chapter provides information describing data management activities and processes that occur during the conduct of a study using EDC. It concentrates

on data reviews, trend analyses, communication, security, midstudy data requests, and various change control processes.

Many of the tasks described in this chapter may be joint responsibilities between different groups, just as there may be many different groups involved in the implementation of various tasks. However, clinical data managers need to be conscious of whether or not these tasks have in fact been performed in a satisfactory manner.

Detailed information comparing paper-based studies with studies employing EDC principles can be found in the chapter entitled “Electronic Data Capture—Concepts and Study Start-up,” along with detailed information describing pre-production activities and planning for an EDC-based study. Recommendations for proper study closeout principles for an EDC study are addressed in the chapter entitled “Electronic Data Capture—Study Closeout.”

### ***Minimum Standards***

- Work with the entire project team to decide which additional edit checks and listings are necessary during the study.
- Document all changes to edit checks and data review specifications.
- Maintain accurate and up-to-date system access, including documentation for changes to access (rights, revocation, addition, etc.).
- Keep training materials updated and readily available to study team members.
- Ensure proper training has occurred for all personnel involved in the conduct of the clinical study, and that all training is documented.
- Monitor changes in study team members to ensure new or reassigned members have been trained according to the study training plan.
- Provide sites with timelines for data entry and query responses.
- Make metrics reports available and review them on a regular basis.
- Where possible, utilize predefined metrics reports and develop new reports as needed to identify performance issues.

- Notify appropriate study team members of site performance issues.
- Monitor metrics reports and data query trends to identify when additional training is needed.
- Observe the frequency of automated and manual queries.
- Conduct additional training as needed to address any system and/or study specific changes.
- Monitor query status for both open and answered queries through reports and task summaries.
- Ensure continued review of data listings to identify any remaining data discrepancies that may generate queries.
- Track progress of investigators' signoffs on CRFs throughout the course of the study.
- Notify the project team of data trends.

### ***Best Practices***

- Document ongoing training activity throughout the life of a study.
- Use all available information to identify training gaps or needs (e.g., query trends, protocol deviations, monitor reports, help desk reports).
- Take advantage of opportunities to provide additional information and training at investigator meetings, study coordinator teleconferences, and monitoring visits, as well as through communications such as newsletters or a study Web site.
- Set metric goals and communicate expectations to study team and site staff.
- Enforce timely data entry and query resolution in order to take advantage of all EDC benefits.
- Program protocol deviation reports early in the study.

- CDM should either be responsible for closing all queries, or at a minimum, reviewing queries closed by other parties.
- Run compliance and safety reports early and frequently.
- Seek input into remedial actions from the project team.

### ***Data Reviews, Trend Analyses, and Remediation***

In EDC trials, the concepts of data review and trend analysis can be applied similarly to the way they are in paper-based studies. However, EDC systems offer a tremendous advantage over paper-based studies by changing the focus of CDM's data validation activities and enabling data trends to be detected more quickly. In EDC, traditional data management roles have changed, as site staff can enter data and have edit checks programmed to trigger queries at the time of entry or immediately after submitting the data. CDM no longer make data changes in the system based on site query responses, but rather site staff enters the data changes. CDM can now focus attention on performing data reviews using listings and reports, to fully ensure data are complete, consistent, and logical. Additionally, many EDC systems allow for sponsor staff other than CDM to perform tasks that have historically been performed by CDM, such as creating manual queries and/or closing queries. The business process, roles, responsibilities, and access rights established for a study will need to dictate how issues found during data reviews are identified and resolved, as well as who is responsible for their resolution.

When data are entered into an EDC system, it typically has not been source verified. Data validation and review activities can be performed before or after data are source verified. A decision must be made whether the project team will require specific data items to undergo source document verification (SDV) prior to data review and/or edit check and manual review activities. If the project team decides to perform data review and/or edit check activities after SDV, a method of communication between clinical research associates (CRAs) and data managers should be established. Some EDC systems have the functionality to indicate when SDV is completed.

One of the most efficient ways to see data trends is to produce a report showing the frequency of queries generated. Such a report will allow the

project team to react quickly and apply various solutions to address these issues. These solutions may include:

- Retrain the site concerning the protocol.
- Retrain the site in electronic case report form (CRF) completion guidelines.
- Retrain the site in system functionality.
- Explain issues in a newsletter to sites.
- Contact sites directly either during the next monitoring visit or more promptly via phone after an issue is identified.

The frequency of particular queries may also prompt the project team to examine edit check specifications. This review of the specifications may indicate such solutions as broadening ranges, rewriting specifications, or eliminating certain checks altogether.

The rapid availability of study data in an EDC system allows project teams to make decisions much earlier in the development lifecycle than in paper-based studies. This study data availability is especially beneficial in enforcing compliance, tracking protocol deviations, detecting safety concerns, and amending protocols if needed. Additionally, as part of the development of reports and listings, it is important for CDM to consider where the most efficient location/system is for generating this information. For example, if you are utilizing a back-end Clinical Data Management System (CDMS) to bring data collected from your EDC tool together with central laboratory data, it may be necessary to establish reports within the CDMS or a statistical analysis software package such as SAS<sup>®</sup> rather than directly from the EDC system.

Early notice of protocol noncompliance is crucial to study conduct, especially with more complex protocols. Site staff may have problems complying with medication requirements, procedure steps, correct device usage, or other protocol requirements. Reports programmed at the beginning of a study to capture protocol deviations should be run frequently to monitor compliance. An EDC system enables the project team to give immediate feedback to site staff so they can provide retraining as necessary. Sometimes, analysis of

protocol deviations and errors in clinical data may indicate the need to amend the protocol. Identifying problems early is vital for study conduct, as it allows the timely entry of correct data by sites. Amending the protocol immediately after an issue is discovered contributes significantly to patient safety and overall success of the study.

In some trial designs, such as adaptive design trials, the protocol may require changes in dosing, sample size, etc. during the course of the study. In this situation, early access to the data and the ability to observe trends in the clinical data as they occur is crucial. Reports with relevant treatment parameters should be produced and reviewed by the appropriate project team members.

A comprehensive safety data review will help identify trends and alert investigators immediately of patient safety issues during the study. Although serious adverse event (SAE) notifications can occur rapidly in paper-based studies by phone or fax, an EDC study offers a more thorough, all-inclusive approach. In addition to SAEs, non-serious adverse events and other pertinent patient information can be reviewed earlier in the study, ensuring that the Data and Safety Monitoring Board (DSMB) has a current, complete picture of the patient safety profile.

Interim efficacy and safety data reviews can also be performed earlier in an EDC-based study using the most current, near real-time patient information. Decisions by the DSMB to stop a study because of safety concerns or lack of efficacy can be made much more quickly than in a paper-based study. This ensures better subject protection and lowers costs if a study must be stopped.

## ***Communication Plan***

There are a variety of methods used for communication, including verbal, written, and electronic communication. In day-to-day business, the most common means of communication are e-mail, telephone, fax, or face-to-face meetings. To lay a solid foundation for an EDC study, effective communication is an absolute necessity.

Clear and comprehensible communication is an extremely important subject for all members of a clinical research team, including vendors involved in the



study conduct. Regardless the stage or component of a study in which an individual is involved, all parties should adhere to the following principles:

- Schedule regular and frequent meetings to keep everyone informed about key study issues and status and to nurture discussions.
- Provide good documentation in a timely manner as a follow-up to all meetings.
- Send a confirmatory message if decisions about the protocol, conduct of the study, or other study-specific matters were made via e-mail.
- Seek clarification if unsure or unclear of any study-related topics, and don't hesitate to discuss.
- Adopt a proactive approach in order to save time and energy.
- Correspond clearly and succinctly. Consider the native language of each participant; communicate clearly, concisely and unambiguously, and avoid phrases or terms that are dialect-specific.
- Use good judgment. E-mail may not be the best means of communication when phone calls or face-to-face meetings can clarify problems better or more easily.
- Maintain documentation of decisions and known issues related to the data in a location accessible by all study team members.

## ***Metrics Reports***

Very early after a study begins enrollment, an EDC system can provide unique opportunities for actively improving site performance as well as training materials. Metrics reports such as query response times, frequencies of queries by form or data items, and number of queries per site should be run early in a trial to identify potential problems. Remediation can then be taken to reduce or even eliminate these issues as the study progresses. Remediation may include revisions to CRF completion guidelines, retraining site staff, retraining CDM staff regarding query wording, special topic newsletters, and teleconferences.

Metrics reports from other data sources should also be considered. For example, if used in the study, a help desk can provide information regarding site performance trends and issues. Also, if a study is integrating data from automated equipment such as an electrocardiogram (ECG), personal digital assistant (PDA), or other electronic devices for electronic patient-reported outcomes (ePRO), additional metrics report opportunities and data interfaces with the EDC system may be considered.

## **Security**

Due to laws and regulations, such as the HIPAA privacy standards, ICH Guidelines E6 Sections 2.11, and 4.8.10, and Article 8 of EU Directive 95/46/EC<sup>2, 3, 4</sup>, access to an EDC system must be limited to authorized staff only. Maintaining appropriate access and system security is essential throughout the duration of a clinical trial. However, security cannot be ensured without user compliance. Therefore, all users must be informed and continually reminded about system access regulations. During monitoring visits, the sponsor and/or contract research organization (CRO) should reiterate to site staff the importance of confidentiality for each user's ID and password. Suspected noncompliance with access regulations should be reported to the assigned system administrator as appropriate.

## **Maintaining System Rights Determined by Roles and Privacy**

Throughout the course of a trial, roles and responsibilities may change, as may data management processes. CDM should manage any changes to documentation describing access rights. Because of the potential impact on specific access rights of individual users, any such changes should be communicated to all study team members. Additionally, role and responsibility definitions should be kept with user access documentation to assist auditors with understanding each user's role.

## **Managing Periodic System Access Review**

Managing user accounts and permissions is a time-consuming task, requiring diligence to ensure security and confidentiality are maintained throughout the duration of a trial. Open communication with clinical operations is necessary to keep track of site and CRO staff changes so as to activate or deactivate corresponding user accounts as needed. The data manager should also use

reports on EDC system activity to periodically review user access to the EDC system. Additionally, as part of this period review they should ensure the access rights are appropriate for each user. However, the length of time between review periods depends on the duration of the study. Standard operating procedures should state what the minimum review period is for the EDC system.

### **Managing Conventions for User Login IDs and Passwords**

Each user of an EDC system must have an individual account, consisting of a unique login ID and password. The sponsor should decide how the IDs and passwords are disseminated to users. Typically, the initial login ID and password can be sent to the individual user using his or her e-mail address, or through traditional methods such as mail or courier. The system administrator should only grant a user access to the system once the user's role-specific training has been completed and documented.

When a user first logs on, the EDC system should prompt the user to change their initial login ID and/or password. If the system is not capable of forcing the user to change their password on first entry, trainers will need to ensure this activity is discussed with all trainees. Users should be trained to keep their IDs and passwords confidential. Each login ID should uniquely identify the user within the EDC system's audit trail, and enable tracking of any information that the user enters, modifies, or deletes. Additionally, users should be instructed to log on to their account, complete data entry and review, and log out at the completion of review. Users should be instructed to log out of the EDC system when the personal computer (PC) used to access the EDC system is left unattended. Login ID and password requirements should include restrictions on re-use of accounts and passwords, minimum length of login IDs and passwords, required frequency of password changes, and automatic log-off when a PC accessing the EDC system exceeds a predetermined amount of inactive time.

### ***Managing User Access***

Turnover of site and study team members is likely, with the volume of turnover related to the size and duration of the trial. Therefore, management of user access will be an ongoing task throughout the course of an EDC study.

This will involve establishing processes for disabling accounts as well as for granting accounts to new users. The system should also be updated when study team members are reassigned to different roles during the course of a study. Monitoring user access will likely require both CDM and clinical operations resources to manage site and sponsor user access.

### **Disabling Access During a Study**

Procedures must be established to define processes for disabling or revoking access to the system as needed. These processes should clearly define who is responsible for communicating staff changes (both internal and external), documenting these changes, and executing these changes. Requirements for automatic deactivation of accounts should also be established in the event of security breaches or users who do not log in for extended periods, such as not accessing the study within 90 days or some other specified time frame.

The sponsor should define appropriate lock-out rules in the event of unauthorized access, whether attempted or successful. If a user enters an incorrect ID or password, an alternative method, as specified through established standard operating procedures (SOPs) or work instructions, should be employed to provide the user with system access.

### **Adding New Access During a Study**

Throughout the course of a trial, it will become necessary to add new users or modify access privileges for existing users. Procedures should be established to ensure these tasks occur without disruption of ongoing study activities. These procedures should detail training prerequisites, steps for requesting access, and the staff members who are responsible for ensuring all site staff and study team members have appropriate access. Documentation of completed training should be provided to system administrators so they know which users may be granted new or modified access rights.

## ***Ensuring Effective Software Support***

When available, reports (which may include surveys) detailing the responsiveness and effectiveness of software support (e.g., the average length of time the help desk takes to assist a user) should be reviewed regularly to ensure support is effective. Several factors are important to ensure assistance

is provided efficiently and expeditiously, including easy access to support staff, ability to address users' questions, and the availability of support when needed.

### **Providing Multiple Language Support**

Although language needs for the help desk should be determined during the pre-production phase of a study, CDM staff should be sensitive to complaints regarding communication problems during the study conduct phase. The problems may be, in part or in whole, related to an inability of the help desk to provide the language support needed, and may require a revision to the original translation needs of the study.

### **Providing 24 x 7 x 365 Support**

As with multiple language support, help desk availability must be determined prior to the start of a study. However, during the conduct of the study, CDM should evaluate feedback from users to ensure that the availability of support is adequate for the study. Reports detailing the responsiveness and effectiveness of software support should be reviewed regularly to ensure tier 1 software support is effective. Tier 1 software support is the lowest level of support needed and includes activities such as unlocking user accounts and resetting user passwords. Information gained from reports and feedback may involve reevaluating the original decisions regarding the level of support needed. For example, if 24x7x365 support was not originally set up, it may be necessary to reconsider it. If a vendor was contracted to provide help desk services, any changes to the contract will need to be considered and negotiated.

## ***Training***

EDC-related training should be provided to internal and external staff during the conduct of a study. Training is most effective when provided as close as possible to the time when the newly learned skills will be used. If a significant time lapse occurs between training and use of the learned skills, retraining should be considered.

## **Reviewing and Maintaining Training Materials**

EDC system training is an important part of proper study management. Training is dependent on the study and target audience, therefore training materials should be developed with these considerations in mind to make the training as effective and appropriate as possible. Moreover, training should be an ongoing process, not just a one-time event. An EDC system can provide the sponsor with the ability to identify a need for retraining users. Some EDC systems can also be used by the study team to deliver updated training materials and communications to users in a timely manner. For example, updated CRF instructions can be immediately provided to all sites and study team members, and newsletters can be provided through a dedicated Web site to communicate updates or changes.

Identifying users' needs for retraining is an important activity of both CDM and clinical operations team members who interact with the site regularly. CDM should be aware of situations at a site that may present challenges and a need for retraining, such as coordinator inexperience, isolation, turnover, or competing priorities. Available information, such as help desk reports, query frequency reports, and protocol deviation reports, can be used to identify materials that need to be updated or users requiring new or additional training.

## **Ensuring Site and Sponsor Staff Training During Turnover**

A common occurrence in clinical research is turnover of both site and sponsor staff. New staff must receive required training, and user accounts and permissions in the system should be updated to reflect staff changes. A plan should be established for new users to be trained in a timely manner so they will have the benefit of access to data on the EDC system. If new site staff are not trained and do not have access to the system, they cannot enter data, and study timelines can be negatively affected.

## ***Change Control***

Any EDC system may undergo changes during the conduct of a study because of changes in EDC software and/or changes in the study itself.

## Software Change Control

Because many clinical trials occur over the course of several years, software changes and upgrades will inevitably have an impact on EDC studies. These changes or upgrades are not just limited to core EDC software, but could also include upgrades to the operating system, back-end database software, or any auxiliary software integrated with the EDC system, such as reporting or extracting software. The differences in change control strategies and processes depend on whether the software is developed internally by the sponsor or purchased from a vendor.

If software is purchased, the sponsor may decide to rely on the vendor's system validation package for the software, including all releases or upgrades, and maintain the system as a "qualified" platform rather than performing system validation upon each release. However, "qualified" software platforms should not be customized by the sponsor unless validation of the customized platform will also be performed.

### **Controlling Changes to the System by Incorporating Software Development Life Cycle Principles**

Before making a decision to implement upgrades to the software system (whether it is a new release or a minor version update), CDM should make a complete assessment of the software changes and obtain input from other areas that may be impacted, including a thorough risk assessment. The first step in performing an assessment is to gain a clear understanding of all changes or additions that will be made to the software. For software purchased from a vendor, this task can be accomplished by ensuring that the software release notes are reviewed and well understood by appropriate staff. Release notes should include documentation of all changes, any known issues in the new release, and instructions for upgrading the software from previous versions.

For software produced internally by the sponsor, a well-developed change control process should be established. This process should include steps for reviewing change requests, grouping multiple change requests together as appropriate, updating requirements and design documentation, build, testing, and implementation.

To determine whether a software system should be upgraded, the sponsor should consider the following issues:

- Impact on data—Assess if any changes in software functionality could potentially impact data integrity. For example, if certain characters or functions will no longer be supported, the sponsor must make sure data integrity will be preserved after the software upgrade.
- Impact on existing code—The software upgrade may require you to make changes to existing programming code.
- Auxiliary systems—The sponsor should assess how related systems or programs will be affected by the software upgrade. Will other systems require corresponding upgrades or modifications?
- Impact on sites—Will the study be inaccessible during the software upgrade? Is the site required to perform certain tasks, such as installing software on their local PCs or changing browser settings? Will the site require additional training? How will the sites be notified of the impact?
- Comparison of cost and value—The costs of implementing and validating a software upgrade should be compared with the business value to be gained.
- The impact on ongoing studies—Considering the impact on the study database and remaining duration, is it worth upgrading software to a new version? Does the software for ongoing studies need to be upgraded simultaneously?
- SOPs and training materials—Will the software upgrade require revision of the sponsor's SOPs or training materials?

For internally produced or customized EDC software, new requirements documentation should be created. This effort is often led by data management. The requirements documentation should include new features and functionality, as well as changes to current features and functionality. The requirements documentation serves as the basis for design specifications. Creating design specifications is typically performed by the group who will be programming the changes.



In addition to the requirements documentation, data management will need to develop a test strategy that documents the testing and validation required for the new software. Depending on the type of upgrade, intensive testing is not always necessary. The following guidelines can be used to determine required testing efforts:

- For a minor version (bug fix or upgrade), limited testing is required.
- For a new release or major version upgrade, moderate to intensive testing is required.

For purchased EDC systems, the vendor should be able to provide and maintain testing plans and results. For internally produced EDC software, test scripts or test cases based on new user requirements should be produced. Before implementing changes in a production environment, all testing should be performed in a test environment. New features and functionality provided by the upgrade (as well as enhancements of existing features or functionality) should be tested. A problem log or Web-based error tracking system should be employed to track errors found during testing, so the status of these issues can be monitored through to their resolution.

If validation of the new release has been successfully completed, the new version or changes can be implemented in production. Please refer to the “Database Validation, Programming and Standards” chapter of *Good Clinical Data Management Practices* for more information about validation, including recommendations, minimum standards and best practices.

### **Training on Changed Software**

While a minor upgrade to software is likely to go unnoticed by users, a new release or major upgrade to software could require additional training. The sponsor should determine the required level of training, which users should receive training, and the method of providing the training.

Typically, sponsor staff (CRAs/CDM) and site staff will require training, which can be delivered in person, from a CD, using the Web, etc. Presentations using screen images can be particularly beneficial for training purposes, as they can be reused for later training sessions. Sponsor staff should be trained first on the software’s new or modified functionality and then the site staff.

### **Developing the Rollout Plan**

Before making new software available to staff, the impact of the revised software should be assessed. For example, if software revisions will require modification of approved CRFs, the sponsor should identify Institutional Review Board (IRB) issues to be addressed (IRB issues are not likely to apply to a software upgrade but may apply to CRF revisions). The sponsor should determine whether new software should be upgraded in stages or all at one time.

Sites should be informed of the rollout of new software with sufficient time given for necessary preparations. In case the upgrade does not occur as expected, a clearly defined rollback or fallback plan should be established prior to software implementation. For international studies, the time available to perform software upgrades can be limited and may require the upgrade to be completed during normal business hours.

### **Managing the Legacy Release**

Software vendors typically maintain all software versions for a defined period of time. The sponsor should be aware of the level of support provided by these vendors. When a vendor rolls out a new system, they may not continue to offer the same level of support for earlier versions of the software system and may eventually retire earlier versions. Typically, once a version is retired, the software vendor no longer provides support for that version.

Because of ongoing development of software systems, the sponsor should plan for future changes, and determine when it is appropriate to upgrade or retire an existing system. Some factors to consider include:

- Software upgraded during study conduct
- Vendor's support commitment to previous versions of software
- Software or hardware that becomes obsolete
- Decreased system performance
- Trial timelines

## **Study-Specific Change Control**

In contrast to software changes, a trial could also be affected by study-specific changes, such as protocol amendments or development issues. As a result, CRFs, edit checks, and/or reports may need to be modified. CDM should assess required changes to determine how they should be implemented in the system and deployed to sites.

## **Changes to CRFs**

Version-control software should be used for CRFs, and electronic files should be organized by study in a hierarchical directory structure. For each study, initial release of and subsequent changes to CRFs should be indicated by directory naming conventions or labels. Identifying the date and time that files were released must be possible so the release timeline is clear for regulatory purposes or to troubleshoot issues with CRFs. All installation qualification documents should be maintained for each version release. A log documenting the version release(s) should also be maintained and signed by all appropriate parties as required and defined by sponsor procedures.

Before rolling out the new version of a CRF, CDM needs to assess who the changes will impact. If the changes resulted in revisions to already approved CRFs, the reviewer should determine if these changes will impact the sites and IRBs, and inform all appropriate study team members and sites well in advance. For international studies, the time to deploy an updated version is more limited and may require deployment during normal business hours.

CDM should consult with clinical operations to determine whether to roll out the new version in stages or all at once. If changes are rolled out midstudy, the study team should first be notified when the changes will be available, and whether the study team will have an opportunity to review changes prior to deployment to sites. Site staff should be notified of training and rollout before changes are released to the production system. Once changes have been made, all parties should be notified. To ensure appropriate sites are moved to the new version of the CRF, CDM should create a log that will keep track of when each site was moved to the new CRF version. Proper logging will ensure no sites are missed in the process. Not all sites may require the new version, such as in cases where the changes are related to a protocol

addendum. Target dates should be set and tracked for upgrading sites to the new system, which should be closely monitored and tracked.

Data entered in the previous version of the EDC study should be made available to the sites and study team. Any new data that have not been entered into the previous version of the EDC study should follow the newly released CRF format. If edit checks were modified, CDM should review old discrepancies to determine if they are still valid, and if any of the discrepancies need to be closed due to the changes.

### ***Midstudy Requests for Patient Data***

A midstudy request for subject data can occur for many reasons, including, but not limited to:

- A scheduled interim statistical analysis based on study design and protocol, which typically focuses on efficacy data
- An interim review of data focusing on safety data, such as adverse events and other data that indicated safety issues in earlier studies (e.g., ECG data, lab panels)
- DSMB or Clinical Endpoint Committee (CEC) regularly scheduled meetings
- A submission package or other type of update (e.g., 120-day safety update) for regulatory purposes
- Any other planned or unplanned data lock

A major factor affecting delivery of midstudy patient data is whether the data are stored by the sponsor or a vendor. If data are stored by the sponsor, the data should be readily available, thereby reducing costs and resources needed. If a vendor's hosted system (Application Service Provider (ASP) model) is used, the timing and frequency of deliveries are more important, and planning will be required for the additional time and costs.

Whether a sponsor or vendor system is used, the required patient data should be clearly identified. Examples of prerequisite identification for exporting patient data include, but are not limited to:

- An interim analysis planned to occur at a particular milestone (e.g., the 100th randomized patient)
- A safety review planned to occur at a particular milestone (e.g., 25% patients enrolled, 50% enrolled)
- A midstudy efficacy analysis, based on statistical design of the protocol
- Regularly schedule DSMB/CEC meeting

In addition to determining which patients are to be included in an export, the sponsor should identify which records are to be included in the delivery. The simplest solution is to include all study data, regardless of its status. However, delivery could be restricted to data verified by the CRA or monitor, or to locked (clean) data, which requires close coordination with the CRA for scheduling monitoring visits. As is the case for paper trials, if data are to be used for an interim safety analysis, reconciliation of SAEs may require additional attention.

Any external data that must be integrated into the database prior to providing any subject data midstudy (e.g., laboratory data or ECGs) should be planned in advance of the study team's timeline for reporting. As necessary, the completeness and accuracy of such data should be ensured by reconciliation before the data delivery occurs.

The recipients of requested study data and the impact to study blinding must also be considered. For interim analyses, SAS<sup>®</sup> datasets are typically provided to a biostatistician or statistical programmer, who subsequently creates tables or listings from the raw data. Other delivery formats could include Microsoft<sup>®</sup> Access or Excel, but these formats are used less frequently and are generally less preferred. Timing of the delivery (e.g., planned or on demand) is also an important component to consider. If required data deliveries are scheduled, necessary procedures can be planned in detail. However, if ad hoc requests for data are anticipated, the process for exporting and delivering data should be robust and flexible to ensure timely delivery. When ad hoc requests are received, programs should be tested and validated to ensure timely delivery. Testing should include the complete extraction and delivery process, including checking that all required variables are available in the datasets and populated

with expected values. Errors or omissions noted during testing can be corrected until the data export operates as required.

### **Midstudy Requests for Notable Subject CRFs**

The Food and Drug Administration (FDA) requires CRFs from subjects to meet certain criteria. As required by CFR 314.50 (f), for any new drug application (NDA), individual CRFs must be provided for any subject who withdrew from the study due to an adverse event, or who died during the study.<sup>5</sup> Depending on the study and FDA center, the FDA may request additional CRFs for review of the NDA.

The sponsor should be prepared to transfer CRFs at any time during the study, for example, for an NDA periodic safety update or integrated safety summary. One possible solution is to provide electronic copies of CRF images. If the CRFs are to be used in a submission, the publishing software used to create the CRFs should be considered so electronic copies can be easily incorporated. When working with a vendor, the sponsor should factor the process for obtaining CRFs into the contract's study timelines and expectations (e.g., maximum number of requests).

### ***Recommended Standard Operating Procedures:***

- Data Review and Edit Checks for EDC Studies
- Data Management Plan
- System Maintenance
- EDC Training
- Study/CRF and Edit Check Change Control
- Software System Change Control
- User Management and Security

## References

1. Marks, RG. Validating electronic source data in clinical trials. *Controlled Clinical Trials*. 2004;25:437–446.
2. US Department of Health and Human Services. *Health Insurance Portability and Accountability Act of 1996*, (HIPAA) Public Law 104-191, as amended, 42 United States Code 1320-d. Washington, DC: US Government Printing Office; 1996. Available at: <http://aspe.hhs.gov/admsimp/pvcrec0.htm>. Accessed on March 27, 2008.
3. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Section 2.11 and 4.8.10. Geneva, Switzerland: Author; 1996.
4. Official Journal of the European Communities. *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995*, Article 8. No. L 281/31. 1995.
5. *Code of Federal Regulations, Title 21, Part 314, Applications for FDA Approval to Market a New Drug*. Washington, DC: US Government Printing Office; 1997.

## Further Reading

*Code of Federal Regulation, Title 21, Volume 1, Part 11: Electronic Records, Electronic Signatures*. Washington, DC: Government Printing Office; 1998.

*Data Basics*. 2001:Summer;8–9.

Drug Information Association. *Workshop on “Electronic Data Capture: Technologies, Implications and Future Requirements”*, November 1998. Baltimore, MD: Author; 1998.

eClinical Forum PhRMA EDC/eSource Taskforce (Formerly the Electronic Data Management Forum (EDM Forum)). *The Future Vision of Electronic Health Records as eSource for Clinical Research, Version 1.0, September 14, 2006*

Electronic records; electronic signatures. *Federal Register*. 1997:62;13429–13466.

Hopp DI. Three topics integral to the use of the Internet for clinical trials: Connectivity, communication, & security. *Drug Information Journal*. 1998:32; 933–939.

Hyde AW. The changing face of electronic data capture: From remote data entry to direct data capture. *Drug Information Journal*. 1998:32;1089–1092.

Hyland ME, Kenyon CAP, Allen R, et al. Diary keeping in asthma; Comparison of electronic and written methods. *British Medical Journal*. 1993:306;487–489.

*IIR Symposia on “Automated Data Capture”, February 1998, Orlando, FL and September 1998, Washington, DC.*

Kubick WR. The elegant machine: applying technology to optimize clinical trials. *Drug Information Journal*. 1998:32;861–869.

Kush RD. Electronic data capture: A survey. *The Monitor*. 1999:Fall;37–40.

Lampe AJ, Weiler JM. Data capture from investigators’ & sponsors’ perspectives: Balancing quality, speed, & cost. *Drug Information Journal*. 1998:32;871–886.

Latham D. EDC: Establishing standards for electronic data capture. Interview of K. Vaillant. *Pharmaceutical Visions*. 1998:Autumn.

Ma, JM. *A Modeling Approach to System Evaluation in Research Data Management*. Unpublished doctoral dissertation. Chapel Hill, NC: University of North Carolina; (1986).

Palm U. Controlling the increased complexity of electronic clinical data collection and management technology by data modeling and system engineering. *Drug Information Journal*. 2002:36;683–692.

Regalado A. Re-engineering drug development, II: Clinical data collection and management. *Start-up*. 1998:27.



Stokes T. Computer systems validation (6-part series). *Applied Clinical Trials*. September 1996, January 1997, February 1997, April 1997, June 1997, August 1997.

US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Washington, DC: US Department of Health and Human Services; 2007.

US Food and Drug Administration. *Guidance for Industry: Electronic Submissions of Case Report Forms (CRFs), Case Report Tabulations (CRTs) and Data to the Center for Biologics Evaluation and Research*. Washington, DC: US Department of Health and Human Services; 1998.

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2003	Initial publication as Electronic Data Capture Principles.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
September 2008	Revised to reflect the orientation of chapter towards the conduct phase of EDC. Content updated and organization of material revised. Study concept and start up, and study closeout content moved to separate chapters.

**This page is intentionally blank.**

---

# Electronic Data Capture—Study Closeout

September 2008

## *Abstract*

With the growing prevalence of electronic data capture (EDC) in clinical trials, the importance for clinical data management (CDM) to fully understand the impact of EDC upon all phases of the research process has grown. It is imperative that data presented to regulatory agencies come from a study that is not only planned and conducted properly, but also closed out in accordance with sound data management principles.

## *Introduction*

Many professionals in clinical data management (CDM) have been forced to reevaluate how they approach their work as electronic tools and systems have gained wider usage in the industry. These tools and systems have had a profound effect on all phases of clinical research, changing the way the CDM team approaches data collection, data transfer, data analysis, reporting, security, archiving, and storage.

Proper closeout activities for these studies are crucial, especially for clinical research presented to regulatory agencies such as the Food and Drug Administration (FDA). The time, energy, and resources employed to collect clinical data is wasted if data cannot be properly verified, validated, transferred, and stored.

The term “lock” may refer to not only locking a study or database, but may also refer to locking specific forms or casebooks. While not all topics discussed in this chapter actually occur during study closeout, they are components of the data lock process and closeout of a study.

## **Scope**

This chapter focuses on how to properly close out a clinical trial when electronic data capture (EDC) has been used. It examines the standards, practices, and procedures to close out an EDC study, including final source document review, database locks, audits, media generation, and hardware disposal.

Many of the tasks described in this chapter may be joint responsibilities between different groups, just as there may be many different groups involved in the implementation of various tasks. However, clinical data managers need to be conscious of whether or not these tasks have in fact been performed in a satisfactory manner.

With the huge impact electronic data capture is having on clinical research, separate chapters are devoted to EDC study startup (see chapter entitled “Electronic Data Capture—Concepts and Study Start-up”) and EDC study conduct (see chapter entitled “Electronic Data Capture—Study Conduct”).

## **Minimum Standards**

- Ensure completion of all required source document verification and data review.
- Ensure all investigator signatures (principal and sub) are in place at closeout.
- Ensure the procedures established for locking fields or forms in a CRF have been followed, including those with open queries or unreviewed and/or unverified status.
- Perform a final review of data listings to identify and resolve any remaining data discrepancies that may generate queries.
- Perform a final review of query status for both open and answered queries through reports and task summaries.
- Ensure defined procedures have been followed for locking the database, and for unlocking the database if necessary.

- Ensure defined processes have been followed for restricting user access once the database is locked, and for revoking access to the production database.
- Ensure adherence to definitions of the audit plan and postaudit data transfer process, as well as identifying audit team members well before study closeout.
- Define specifications for formatting subject profiles, as well as a process for generating and reviewing subject profiles.
- Ensure investigative sites have access to their CRF data after study completion. Once they have received the appropriate media for this data, their access to corresponding data in the EDC system can be revoked.
- Ensure any hardware provided to sites is retrieved according to organization standard operating procedures (SOPs).
- Determine requirements for creating additional media to represent the study database if needed.

### ***Best Practices***

- Ensure investigators and other site staff are educated in the signature-break process long before study closeout so there is no confusion on the topic at closeout. Signature-break may occur when data has been changed post investigator signature. Should signature-break occur, this information will help avoid confusion, and will ensure the investigator is available to re-sign if necessary.
- Implement a verification procedure to ensure data received or extracted from the database matches data entered in CRFs, especially in cases where additional output programming is conducted. This practice confirms integrity of the data being released for statistical analysis.
- Review and refine the source data verification timeline with monitors and clinical operations after the last subject visit occurs and data entry is completed, (In some cases these processes can also be performed prior to the last subject visit).

- Ensure all medical coding activities have occurred as required.
- Use an incremental form or casebook lock strategy to reduce the amount of data review and locking needed upon study completion.
- Ensure all tasks documented in the data management plan are complete, and coordinate with clinical operations personnel to ensure all site monitoring activities are complete prior to database lock.
- Use an established checklist of tasks to be completed prior to database lock in order to meet database lock timelines.
- In preparation to meet database lock deliverables, adjust timelines as needed for all queries to be answered by sites.
- Use an established communication plan between the clinical team, site staff, statisticians, and data management. This communication plan should ensure all data reviews are completed and queries are answered in time to meet database lock deliverables.
- Create a calendar of vacations or out of offices for all team personnel to ensure proper resources are available for study close out activities.
- Review current regulatory standards and guidelines for how data should be presented in the subject profile (e.g., headers, footers, and margins).
- Determine the appropriate media to use for reporting of subject profile data.

## ***Final Review***

### **Verifying Source Document Verification of All Records**

Before a form can be locked and the database closed, source documentation and data review must be completed as required for all forms and fields. It is beneficial if the EDC system can indicate source document verification (SDV) status and data review activity. Prior to locking a database, ensure all required SDV has been completed by clinical research associates (CRAs).

Frequent communication between the clinical study team, including CRAs and data management team, is critical. Changes to data in the CRFs are

possible as late as the day of database lock. A plan should be established to ensure CRAs are available to source verify any changes, if necessary.

### **Verifying All Queries Have Been Closed**

All reasonable efforts should be made to ensure all queries are answered or closed prior to database lock, particularly for those queries that may impact the analysis or outcome of study results. Depending on the details of the DMP, it may be acceptable to lock noncritical forms with open queries. During study startup, conditions for locking an CRF should have been defined; thereby ensuring the CRF cannot be locked with open query status. Prior to locking a database, CDM should ensure all queries have been answered or closed, including automatic queries and manual queries created by CRAs and data managers. This task may be accomplished through use of varied system reports and status indicators, according to the EDC system's features.

### **Verifying e-Signatures**

It is necessary to verify that all CRFs have been electronically signed by the investigators (principal and sub) responsible for a particular site. The investigators are responsible for reviewing each subject's CRF, and to confirm that data entered are complete and accurate. Sponsor organizations must determine the level of detail required for investigators' signatures. For example, some organizations will decide that the investigator's e-signature must be applied to every individual CRF, while others may decide one investigator signature on the final page of each subject's CRF is acceptable. While there are many studies using e-Signatures, some are also still working with paper-based signatures, even with those studies utilizing EDC.

Regardless of the final signature method being used, a process should be established for notifying sites that CRFs are ready for investigator signature. Policies and processes related to re-signing a CRF should also be defined and adhered to. If a site changes a data point on a CRF already signed by the investigator, the rules must be in place to decide whether the data change "breaks" the signature. If the signature is broken due to the change, the investigator must re-sign the CRF.

It is expected that the CRA will track the investigator's progress signing the CRFs. However, the data manager is responsible for conclusively verifying that all CRFs have been signed by the investigator prior to database lock. For any data changes that "break" the signature, the data manager must verify those CRFs are re-signed.

### ***Final Locking of Data and Database***

It is recommended that a checklist be established to ensure completion of required tasks prior to database lock. These tasks may include, but are not limited to:

- Identifying data management staff responsible for carrying out database lock
- Ensuring required medical coding of adverse events, prior and/or concomitant medications, and medical history verbatim terms has been completed and is accurate
- Ensuring issues identified in edit checks performed outside the EDC system and data listing reviews have been resolved
- Resolving and/or closing all outstanding queries
- Importing and/or reconciling all external data (and listing external data reconciled)
- Completing serious adverse event (SAE) reconciliation

Once all database lock tasks have been completed, the database can be closed and soft- or hard-locked according to the sponsor-approved definition. This step implies that all forms in the study have been locked according to defined procedures, all tasks are completed, all conditions have been met as defined in the data management plan, and the final data transfer has been received or extracted according to data extraction specifications defined at the start of the study.

After database lock, an audit may be performed on the final data. Based on findings from the audit of final data, further data corrections may be requested of the site. Once corrections have been made and verified, and additional investigator signatures have been obtained as necessary, another data transfer



or extraction should occur. It is recommended that a comparison program be run to determine if requested changes to the database were successfully executed, as well as any other changes that were made to data other than what was expected.

### **Soft Lock**

Typically, records that are soft-locked cannot be updated by sites, but the sites may still be able to respond to open queries. Many EDC systems support soft locks at a visit, page or data point level. This capability supports a process of “rolling” or gradual soft locking of data throughout the course of a study, reducing the effort required by the data manager to lock data at the end of the study. Incremental soft locks can also be an effective approach to supporting midstudy or interim data review activities.

### **Hard Lock**

Database hard lock typically occurs when all data have achieved soft lock status and all other study closeout activities have been completed. After a database has been hard locked, tightly controlled procedures for unlocking the database must be employed, and only a few privileged users should be able to modify data. Once the database has undergone a hard lock, the data are considered ready for final analysis and archiving.

### **User or System Access Revocation**

At the conclusion of a study, user access rights must be modified. During study closeout activities, rights should be modified to only allow the site to read data but not to enter or change data. Once the required media have been created and sent to the site, all access to the corresponding data in the EDC system should be removed.

User access (including view access) to a subject’s data should be completely revoked once a copy of the subject data is received and confirmed by the site. Prior to this revocation, the site must continue to have access to the study database in the event of a site inspection or audit.

## **Audits**

The process of auditing data for a CRF will be different than the process used for a paper CRF. However, the end result is the same—the data entered are the data presented for analysis. While the audit process may differ from one sponsor or contract research organization (CRO) to the next, the auditing process will be determined by how the sponsor or CRO extracts data from the EDC system. Any additional programming required to transform study data into SAS<sup>®</sup> datasets could affect how data are displayed. An audit ensures that the datasets received match data entered in the CRF. Additional EDC issues to consider for auditing include, but are not limited to, reconciling medical coding, data management plan comparison, external data import, query verification, and completion of all queries that required data changes.

A plan should be established in advance to identify approaches to be taken, including a sampling plan, acceptable error rates, and escalation in the event of audit findings.

## **Generating Archive Media for Sites**

At the conclusion of a study, media must be created that represent the data collected throughout the study.

### **Quality Review and Replacement of Subject Data**

Subject data should present the data collected in a CRF system organized by subject identifier, visit, and form. The data are typically presented in a manner that allows for effective navigation of the subject profile in a format such as a pdf file or a similar format. The subject data should also include an audit trail, electronic signature, and query information, which allow a reviewer the ability to see all data entry and modification that have occurred since it was created. Subject data should be provided on durable media, such as a CD-ROM. A master copy of the durable media should be created to contain all the subject profile data. In addition to this master copy, individual CD-ROMs or other durable media with site-specific data should be created and forwarded accordingly.

## **Archive Media**

After the study is complete and a copy of subject data have been generated and securely distributed to sites successfully, a complete copy of the study database should be created for archival purposes.

## **Ensuring Compliance with Applicable Guidance**

A final review of all study documentation should be performed to ensure defined processes have been adhered to such that end of study deliverables from the EDC system meet the requirements specified in applicable regulatory guidance documents. This review should ensure that the study documentation address expectations from regulatory agencies. Some of the guidance and specifications for this review include *Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications*<sup>1</sup>, and *ICH M2 EWG Electronic Common Technical Document Specification*<sup>2</sup>.

## ***Provisioned Hardware Disposition***

For studies in which hardware was used for data collection, a determination must be made once the study has concluded regarding hardware disposal. Any hardware that was provisioned to sites should be retrieved per organizational standards and requirements. The hardware may be refreshed and used in future studies, recycled for other uses within the organization, or retired.

### **Recycle**

If the hardware will be recycled, the responsible parties (e.g., sponsor, CRO, site, vendor, etc.) must ensure all data and applications are removed from the hardware.

### **Retire**

If hardware used to store information is to be retired and never used again for data entry purposes, the process for retiring the hardware should be determined. This process should consider how the hardware will be secured if

data are maintained on it, or how data will be removed and the hardware destroyed if it is not to be used again.

### ***Recommended Standard Operating Procedures***

- Study Closeout
- Database Lock
- Reconciliation of Electronic Lab Data
- Serious Adverse Event Reconciliation
- Study Close Audit (typically handled by the QA/QC department)
- Audit of Data Extraction and Output
- Generation and Review of Archive Media
- Maintenance of Coding Dictionaries
- Vendor Audits/Management

### ***References***

1. US Food and Drug Administration. *Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications*. Rockville, MD: US Department of Health and Human Services; 2006.
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, *ICH M2 EWG Electronic Common Technical Document Specification*, V 3.2, Feb 2004.

### ***Further Reading***

*Code of Federal Regulation, Title 21, Volume 1, Part 11: Electronic Records, Electronic Signatures*. Washington, DC: Government Printing Office; 1998.

Code of Federal Regulations, Title 64, Part 4432, *Providing Regulatory Submissions in Electronic Format-NDAs*. Washington DC: US Government Printing Office; 1999.

Code of Federal Regulations, Title 66, Part 57721, *Providing Regulatory Submissions in Electronic Format-ANDAs*. Washington DC: US Government Printing Office; 2002.

Code of Federal Regulations, Title 64, Part 4433, *Providing Regulatory Submissions in Electronic Format-General Considerations*. Washington DC: US Government Printing Office; 1999.

*Data Basics*. 2001:Summer;8–9.

Drug Information Association. *Workshop on “Electronic Data Capture: Technologies, Implications and Future Requirements”*, November 1998. Baltimore, MD: Author; 1998.

eClinical Forum PhRMA EDC/eSource Taskforce (Formerly the Electronic Data Management Forum (EDM Forum). *The Future Vision of Electronic Health Records as eSource for Clinical Research, Version 1.0, September 14, 2006*

Electronic records; electronic signatures. *Federal Register*. 1997;62;13429–13466.

Hopp DI. Three topics integral to the use of the Internet for clinical trials: Connectivity, communication, & security. *Drug Information Journal*. 1998;32; 933–939.

Hyde AW. The changing face of electronic data capture: From remote data entry to direct data capture. *Drug Information Journal*. 1998;32;1089–1092.

Hyland ME, Kenyon CAP, Allen R, et al. Diary keeping in asthma; Comparison of electronic and written methods. *British Medical Journal*. 1993;306;487–489.

*IIR Symposia on “Automated Data Capture”*, February 1998, Orlando, FL and September 1998, Washington, DC.

Kubick WR. The elegant machine: applying technology to optimize clinical trials. *Drug Information Journal*. 1998;32;861–869.

Kush RD. Electronic data capture: A survey. *The Monitor*. 1999:Fall;37–40.

Lampe AJ, Weiler JM. Data capture from investigators' & sponsors' perspectives: Balancing quality, speed, & cost. *Drug Information Journal*. 1998;32;871–886.

Latham D. EDC: Establishing standards for electronic data capture. Interview of K. Vaillant. *Pharmaceutical Visions*. 1998:Autumn.

Ma, JM. *A Modeling Approach to System Evaluation in Research Data Management*. Unpublished doctoral dissertation. Chapel Hill, NC: University of North Carolina; (1986).

Palm U. Controlling the increased complexity of electronic clinical data collection and management technology by data modeling and system engineering. *Drug Information Journal*. 2002;36;683–692.

Regalado A. Re-engineering drug development, II: Clinical data collection and management. *Start-up*. 1998:27.

Stokes T. Computer systems validation (6-part series). *Applied Clinical Trials*. September 1996, January 1997, February 1997, April 1997, June 1997, August 1997.

US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Washington, DC: US Department of Health and Human Services; 2007.

US Food and Drug Administration. *Guidance for Industry: Electronic Submissions of Case Report Forms (CRFs), Case Report Tabulations (CRTs) and Data to the Center for Biologics Evaluation and Research*. Washington, DC: US Department of Health and Human Services; 1998.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2003	Initial publication as Electronic Data Capture Principles.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
September 2008	Revised to reflect the orientation of chapter towards the closeout phase of EDC. Content updated and organization of material revised. Study concept and start up, and study conduct content moved to separate chapters.

**This page is intentionally blank.**



---

# CRF Completion Guidelines

June 2008

## *Abstract*

Accurate completion of case report forms (CRFs) is paramount to the quality of data that are captured during a clinical trial. This chapter covers guidelines for training sites to complete CRFs correctly, and includes discussion of the proper format, design, and content of completion instructions provided with CRFs. Suggestions for the given content of general instructions and CRF- or page-specific instructions are also given.

## *Introduction*

Case Report Forms (CRFs) should be completed as fully and accurately as possible, with the aide of complete, concise and logical guidelines. Well prepared CRFs should provide instruction and guidance on how the sponsors are expecting forms to be completed at the investigative site.<sup>1</sup> CRF completion guidelines should help ensure that all required fields are completed, and that the data provided within these forms are logical within the scope of the study protocol. The guidelines should not provide guidance or suggestions that could be considered leading the user.

The CRF completion guidelines document is a tool that should be available to all members of the multidisciplinary team participating in a clinical trial, and should be referenced to ensure accurate and consistent entry and interpretation of data. These guidelines help train site staff on proper form completion, and also aide Clinical Research Associates (CRAs) on how to review data on the completed forms.

A complete and accurate CRF will result in fewer queries being generated by data management for site staff to resolve. Accurate CRF completion and

review will also result in more meaningful data analyses<sup>2</sup>, quicker validation of data, and will ensure a timelier database lock.

CRF completion includes paper-based transcription as well as direct entry into an electronic system. Therefore, CRF completion guidelines should take into consideration the particular mode of data collection for the study, such as paper CRFs, electronic data capture (EDC) systems (both local electronic data capture systems and central Web-based systems), or interactive voice response systems (IVRS). It is important that the guidelines address each mode with appropriate instructions.

For traditional paper CRFs, the CRF completion guidelines are either printed as part of the CRF or as a separate document. For electronic CRFs or EDC systems, the guidelines may be provided as separate instructions on the screens, an online help system, or system prompts or dialogs generated relative to the data being entered.<sup>3</sup>

## **Scope**

The scope of this section is to describe how CRF completion guidelines are to be created and used to aid in the precise and logical capture of clinical study data.

## **Minimum Standards**

- Document the process by which CRF guidelines are created, reviewed, approved, updated, and distributed.
- Create CRF completion guidelines for at least every multiple site protocol.
- Provide site coordinators and CRAs with CRF completion guidelines, and train the users on the function of these guidelines prior to first patient visit or enrollment. Document training and forward this documentation to the appropriate study team member for retention.
- Provide data management, biostatistics, medical writing and other clinical research team members with CRF completion guidelines so these groups are aware of how sites are instructed to complete CRFs.

- Design CRF completion guidelines from the perspective of site coordinators and CRAs who will be using these guidelines, taking into account clinical treatment procedures at the site, such as the organization of medical records and methods being used to obtain measurements.
- Include a general instructions section and a page-by-page instructions section.
- Ensure guidelines are readily and easily available to the user. Ensure instructions are concise, easy to understand, and do not suggest answers to users completing the forms.
- Update CRF completion guidelines if any changes are made to the CRFs that affect CRF completion. Include version control on the updated documents.

### ***Best Practices***

- Develop guidelines in collaboration with representatives from clinical research, programming, data management, biostatistics, safety, and medical writing.
- Establish a formal written approval procedure for CRF completion guidelines consistent with or included as part of the actual CRF approval process. Document any changes and maintain version control of the document.
- Present CRF completion guidelines at an investigators' meeting (or similar forum) with data management team members leading the review and training. Provide site staff and CRAs with a correctly completed sample CRF and CRF completion guidelines at the time of training.
- Stress the importance of completing all mandatory fields—if a data item is unavailable or unknown, instruct users to enter an acceptable notation to account for the missing value (e.g., N/A or UNK). Clearly define notations to be used as well as the circumstances in which to use them (e.g., delineate between the use of UNK as opposed to N/A).
- Include a list of acceptable abbreviations (if any), with definitions that can be used in completing the CRF.

- Include detailed instructions on proper completion for every CRF page. For paper studies, printing CRF completion guidelines on facing pages of a blank CRF (the page's CRF completion guideline is on the back of the preceding CRF page) for a given protocol proves to be most beneficial.
- Review data quality periodically, re-educate site personnel as needed, and revise CRF completion guidelines as necessary, particularly for long-term studies.
- Make CRF completion guidelines for EDC studies available (for example, as an online file, a hard copy, or a printed version) even though CRF completion guidelines for EDC studies may be included as part of the programming and available on the screen.
- Develop standard CRF completion guideline modules that can be used across studies.

### ***Format and Content of CRF Completion Guidelines***

CRF completion guidelines can be instructions within a specific section of a given CRF page, such as “check only one box”, “record all medications taken within the last 7 days”, etc. Additional instructions can be included within the CRF, such as instructions printed on facing pages, or they can be maintained in a separate document providing detailed instructions (e.g. CRF completion manual).

Following is a suggested format for CRF completion guidelines which are created as a separate document. CRF designers should determine the format of CRF completion guidelines that are integrated throughout the actual CRF pages.

### ***General Instructions Section***

The general instructions section of CRF completion guidelines should include information that applies to completing the entire CRF and submitting completed CRFs.

General instructions for completing CRFs include, but are not limited to the following:

### **EDC and Paper CRF Studies (all CRFs):**

- Ensure that all required fields on each CRF are completed.
- Provide contact information if questions arise while completing the CRF.
- Describe study convention for visits or assessments that are not performed.
- Ensure that all free text entries are spelled correctly and are clinically appropriate.
- Provide a list of acceptable abbreviations, which may vary between studies or indications.

### **Paper CRF Studies Only:**

- Ensure the use of permanent media (blue or black ink).
- Ensure that all items captured in the CRF are legible.
- Specify procedures for making corrections to data. For example, “Corrections to data should be made by drawing a single line through the incorrect entry, writing the correct response above or near the original entry, and initialing and dating the change. Scratch outs and/or correction fluid or tape should never be used.”
- Provide instructions for the process flow of completed documents, including shipping address, which copies of the CRF to ship, courier service to be used, etc.

### **EDC Studies Only:**

- Do not share user IDs or passwords with anyone.
- Do not record and/or store user IDs and/or passwords in non-secure locations. Try to remember user IDs and passwords without recording the information on paper.

## ***Page/Screen-specific Instructions***

Every page or screen should have specific instructions on how data should be captured according to the protocol. Keep these instructions brief and focus on critical fields or those that may be interpreted in a variety of ways.

Page-specific instructions include, but are not limited, to the following:

- Indicate mandatory fields and appropriate notation for information that is not available.
- Note any instructions for subject-completed forms, such as assessments completed at visit by the patient.
- List and explain procedures for clearly reporting:
  - ❑ any visits that a subject fails to make (e.g., specific instructions on completion of information on blank visit pages)
  - ❑ tests that are not conducted, examinations that are not performed
  - ❑ all withdrawals and dropouts of enrolled subjects from the trial
- Provide any special instructions for completing unscheduled visit pages.
- Provide instructions for recording Adverse Events and Serious Adverse Events (e.g., record diagnosis instead of symptoms whenever possible).
- Instruct personnel to only capture data in specified CRF fields and to not write in margins.

## ***Recommended Standard Operating Procedures***

- Preparation, Review, Revision and Distribution of CRF Completion Guidelines
- Study Start and Other Investigators' Meetings
- Training Records
- Study Initiation Process

## References

1. Spilker B. Guide to Clinical Trials and Developing Protocols. New York: Raven Press; 1984.
2. International Conference on Harmonisation. *ICH Harmonized Tripartite Guideline for Good Clinical Practice*. 2<sup>nd</sup> ed. London: Brookwood Medical Publications; 1996.
3. Spilker B, Schoenfelder J. *Data Collection Forms for Clinical Trials*. New York: Raven Press; 1991.

## Further Reading

McFadden E. 1997. *Management of Data in Clinical Trials*. New York: John Wiley & Sons.

## Chapter Revision History

Publication Date	Comments
September 2002	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
June 2008	Revised for content, style, grammar, and clarity.

**This page is intentionally blank.**



---

# **CRF Printing and Vendor Selection**

May 2007

## ***Abstract***

Planning for the printing of a study's case report forms (CRFs) is essential to the study's conduct. This chapter provides insight and guidance for this critical component. Guidelines for the evaluation and selection of CRF printing vendors are provided. The chapter also covers the process by which a clinical data manager plans for the production of printed CRFs and their timely delivery to sites, with both tasks completed by a third-party vendor. Guidelines for the CRF binder, the paper used for printing, and tabs banks are discussed in regard to the specifications that should be provided to the printing vendor. Recommendations are made for binding, packaging, and shipping the CRFs, with an emphasis on the importance of timetables. Guidelines for the evaluation and selection of CRF printing vendors are provided. An example of a CRF printing specifications checklist is included.

## ***Introduction***

The case report form (CRF) is a critical document for capturing relevant data in a clinical trial. The development of a study's CRFs is addressed in the Data Acquisition chapter. However, the production aspects of the CRF, including printing CRFs and ensuring their delivery to sites, must also be addressed. The selection of a vendor for these tasks should also be carefully considered.

## ***Scope***

This chapter will review considerations for outsourcing the printing of CRFs. Use of the following guidelines will help ensure the same quality and service from the contracted print vendor that the Clinical Data Manager expects to receive.

## **Minimum Standards**

- Establish specifications outlining CRF printing and distribution requirements. These specifications should include:
  - the complete list of items in the CRF binder
  - the total number of each item to be printed
  - the type of paper
  - the type of binding
  - the collation order
  - the type and number of tab banks
  - the number of tabs per bank
  - the images to be printed
  - the instructions for printing
- Provide packaging instructions to the printer.
- Submit new printing specifications (including printing and shipping timetables) to the printers whenever significant modifications are made to the CRF or to any item outlined in the specifications.
- Obtain approval by appropriate team members of the final print-ready CRF, the CRF printing specifications, and the shipping/distribution timetable prior to the submission of the final printing specifications to the printer.

## **Best Practices**

- Use a vendor qualification program to select a vendor.
- Ensure that other study materials such as pocket cards, study schedule posters, pre-printed return envelopes, and study contact information are printed to compliment the CRF and associated materials and are distributed simultaneously.

- Obtain a prototype of the CRF book from the printing vendor for review and approval before the final print run. The prototype should include all pages, tabs, spine label, and cover.
- Use a vendor evaluation program throughout the vendor relationship.

## **CRF Binder**

Prior to submitting the final printing specifications to the printer, the final print-ready CRF, printing specifications, and shipping/distribution timetable should be approved by appropriate project team members. CRF binder specifications should include all of the information the vendor needs to produce the CRF binder and associated materials.

To determine the total number of CRFs, diaries or other required pages to be printed, consider the number of evaluable patients required per the protocol, the expected drop-out/replacement rate, and the possible need for a back-up supply. The back-up supply should be 10–15% of the total number of patients enrolled. If materials are distributed in packages, overage estimates should take into account the extra items that are in the pack. For example, if SAE forms are printed on a pad of 100 forms, they will be distributed in allotments of 100. Generally, a site that requires 101 pages will actually use 200 printed forms.

Also estimate the number of CRF papers with a breakdown of the number of no-carbon-required (NCR) pages, non-NCR pages, and other pages (e.g., diary or quality of life pages).

## **Paper**

Specify the paper to be used for printing the CRFs. Include information on the type of paper, color, page weight, hole-punch, perforation, and gum for each page or section. For example, conventional three-part, NCR paper comes in many colors and weights. Many organizations use a white, yellow, pink combination or a white, yellow, heavy card stock combination. The type and number of NCR pages required depend on the workflow and system used. Traditionally, white is the original copy, yellow is the working data management copy, and pink is the site copy. Scanning or fax-based systems

may require only two copies (the original white copy for scanning and the site copy).

There are other special considerations with the use of NCR paper. Printer specifications should include a piece of cardboard or other provision for the site to protect unused pages while completing a CRF page. When using a new vendor or a new paper supplier, it is advisable to test the NCR paper. The copy quality on the second or third ply is dependent on the quality of NCR paper. The weight of the paper should also be specified depending on your workflow. Paper of certain weights has been known to work more efficiently when faxed or scanned. If evaluating the paper supplied by a vendor, test the paper's quality when used to fax or scan printed material.

Consideration for collection of adverse events (AEs) and concomitant medications must be taken. If AEs and medications are collected at each visit and then harvested at every monitor visit, a pull-page system may be used. For example, a clinical data manager (CDM) may use four-part NCR paper in which the fourth page is harvested first (a pull page), thereby enabling the data to be collected earlier. In an alternative approach, the fourth copy could be non-NCR so the next copy of the document reflects only the changes to the data.

## ***Tab Banks***

Tab banks are very helpful to the sites in navigating the CRF during the clinical trial. Specify the number of tab banks and number of tabs per bank. Organizing the printing specifications by tabs can effectively communicate the collation order to the printer. Also, specify the paperweight of the tabs (usually card stock), the type and color of Mylar dip or other laminate on the tabs, and the text to be printed on each tab or tab page.

## ***Binding, Packaging, and Shipments***

Specify the type of binding, binder color, width, number of inside pockets, cover text or art, and spine label.

Specify the packaging instructions and include a packing list of the items that each site should receive. For example, special forms such as drug accountability logs, screening logs, SAE forms, subject diaries, and

questionnaires may be bound separately in books or pads. Special forms may also be conveniently shrink-wrapped in appropriate numbers for each site.

If the printer is shipping materials to sites, provide shipping instructions. Specify the number of sites and the number of items per site, the shipping company, and the shipping method (e.g., ground or air). When finalizing timelines, the location of sites should be considered. Shipping to international sites may require additional time. With the shipping timetable, provide process instructions for tracking the shipment, checking the inventory of the shipment, and notifying the sponsor of the shipment's status.

### ***Information Commonly Provided With Printing Specifications***

If applicable, the following information should be provided to the printer in addition to the printing specifications:

- The final camera-ready artwork of the CRF, the diary, and other pages in electronic files. The format of any electronic files should be discussed and agreed upon with the printing vendor.
- The specifications for CRF layout (e.g., layout of the CRF identifying location of tabs, instructions on the back of tabs, collation of pages, etc.).
- A list of tabs, including the breakdown by bank and color.
- The camera-ready artwork of instructions to be printed on the tab backs.
- The company logo and text for the spine label.
- If the printer is shipping to the sites, a list of sites and their mailing addresses. Moreover, shipping instructions should include details on how the printer will know when the site is approved to receive study materials.
- The priorities and specifications for printing the barcode, if applicable.
- The tentative timetable for sending the final-master copy to the printer, for reviewing the materials prior to the final printing run, and the deadline for the arrival of the shipments at the sites.

The printer should provide a complete prototype of the CRF book for review and approval before the final print run. The prototype should include all of the book's pages and tabs, the spine label of the book, and the cover of the book.

New printing specifications (including printing and shipping timetables) should be submitted to the printers each time significant modifications are made to the CRF or to any item outlined in the specifications. An example of a CRF printing specifications checklist appears in Appendix A.

### ***Selection & Evaluation of CRF Printing Vendors***

Print vendors should be qualified. Select print vendors who specialize in CRF printing and have an understanding of the clinical trial process and CRF design. The print vendor should understand the importance of maintaining timelines in the printing and shipping of CRFs before the first patient is enrolled at each site. The printer should be knowledgeable regarding time-to-ship internationally and customs regulations.

Evaluation criteria should include the following: accuracy of printing, quality of service, turnaround time (turnaround time on initial print job and additional requests for extra pages), pricing, CRF design experience, digital or rotary printing, bar-coding capabilities, changes for re-setup, and storage charges. Other criteria to consider is whether the printer out-sources parts of each job such as printing Mylar tabs, separate charges for printing on tab backs, volume discounts, international shipping capabilities, and turnaround times.

### ***Recommended Standard Operating Procedures***

- CRF Design
- CRF Production Guidelines
- CRF Printing Specifications
- Vendor Selection

### ***References***

N/A

### ***Further Reading***

N/A

### ***Chapter Revision History***

<b>Publication Date</b>	<b>Comments</b>
January 2002	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.

## Appendix A: Sample CRF Printing Specifications Checklist

Total # of CRF binders to be printed \_\_\_\_\_  
Total # of diaries to be printed \_\_\_\_\_  
Total # of CRF pages per binder \_\_\_\_\_  
    # of NCR pages per binder \_\_\_\_\_  
    # of non-NCR pages per binder \_\_\_\_\_  
    # of diary pages per binder \_\_\_\_\_

Page formats: 2-part NCR with 2<sup>nd</sup> part cardstock, or specify other (The first part NCR should be white paper of weight 26):  
\_\_\_\_\_

Specify page format for diary pages and diary covers (ex. Tri-fold):  
\_\_\_\_\_  
\_\_\_\_\_

Tabs: specify # of banks, # tabs/bank, #tabs with printed instructions on back, mylar-laminated or not and Mylar color:  
\_\_\_\_\_

Does printer need to add page numbers? :    Y        N

Binders (specify):

Color: \_\_\_\_\_ Attach spine label              
Width: \_\_\_\_\_  
# inside pockets: \_\_\_\_\_ Attach cover artwork           

Timetable:

Final master copy to printer: \_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
Prototype to XXX/XXX for review & approval \_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
Shipment to arrive on site or at XXX/XXX \_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_

Packaging instructions:  
\_\_\_\_\_  
\_\_\_\_\_

Shipping Instructions: # of CRFs to be shipped per site: \_\_\_\_\_

# of diaries to be shipped per site \_\_\_\_\_

Will printer store backup supply?    Y        N

Specify ground or air and # of days to arrive: \_\_\_\_\_

Attach bar code specifics \_\_\_\_\_

Attach list of sites, addresses, and shipping priorities \_\_\_\_\_

Attach final camera ready artwork for the CRF, tab back instructions, diary \_\_\_\_\_

Attach specifications of layout

Additional comments: \_\_\_\_\_



---

# Database Validation, Programming and Standards

March 2009

## *Abstract*

Success of any clinical study depends on the quality and integrity of its final database. Validation of the software system and database used for a study are crucial risk-focused quality processes for assuring and ensuring quality and integrity. This chapter discusses principles and types of validation, as well as common validation risks. Although system validation is discussed, the primary focus of the chapter is on study-specific validation, which has a greater direct impact on clinical data managers.

## *Introduction*

The clinical data management system (CDMS) used to conduct a clinical study "...should be designed to...prevent errors in data creation, modification, maintenance, archiving, retrieval or transmission...".<sup>1</sup> As required by 21 CFR Part 11 and the predicate rule(s) applicable to the drug, device, or biologic in development, thorough documentation should exist at all levels of a clinical study's data collection and management. Given the multifaceted responsibilities of a CDMS, the validation process is necessary, ongoing and often complicated.

The term "validation" may refer to validation of the CDMS itself or validation of programming related to the development of a study- or protocol-specific database. Although both types of validation are crucial to the success of a study, the details of CDMS validation tend to be the responsibility of programmers or information technology (IT) personnel, although clinical data management (CDM) personnel are responsible for verifying the CDMS has been validated properly and is fit for its intended purpose.

## Scope

This chapter addresses CDM validation activities that should accompany the installation of a CDMS and its patches or upgrades, as well as the testing and validation necessary when designing study-specific databases on that system. Although this chapter briefly discusses validation associated with software application development, a full description of software development validation is outside the scope of *Good Clinical Data Management Practices*. The validation measures necessary for software development and proprietary systems design are very different and more complex than the process of validating study-specific applications. This chapter also does not address validation of external data such as laboratory data. Recommendations for validation of these data are addressed in the chapter entitled “Laboratory Data Handling.”

The software development life cycle (SDLC) validation approach advocated in the device and Good Laboratory Practice regulations is also appropriate for application development. The same general principles offer guidelines on the setup of individual protocols within a validated CDMS, although direct application may not always be appropriate or practical.

Although some of the specific topics addressed by this chapter may not be the direct responsibility of CDM personnel, CDM must have an ongoing awareness of the requirements and ensure these tasks have been completed in accordance with the principles and standards of their organization, regulatory bodies and good clinical practice.

## Minimum Standards

- Generate a validation plan defining the testing methodology, scope, problem reporting and resolution, test data, acceptance criterion and members of the validation team.
- Ensure the CDMS meets user/functional and regulatory requirements and continues to meet these requirements through the course of its use.
- Implement the CDMS carefully, testing according to specifications, documenting all testing and issues, and ensuring objective evidence of testing is generated.

- Define processes for handling change control issues, with a clear determination of when revalidation will be required due to changes.
- Document all validation details prior to implementation in a summary document (e.g., validation report), including all applicable review and approval signatures.
- Ensure documentation remains complete and current.
- Ensure that only qualified staff develop, maintain and use the system.
- Approval of validation plan and documented results from an appropriate level of independent quality resource(s).

### ***Best Practices***

- Identify all intended user requirements of study-specific programming.
- Use organization standards, as available, to prepare study-specific programming.
- Use organization standards to document programs.
- Use code libraries wherever possible.
- Confirm that study-specific programming applications perform as intended based on the user requirements (data management plan requirements, CRF requirements, database specifications, edit check specifications, validation plan, etc.).
- Document performance during validation.
- Ensure documentation remains complete and current for live use, and is indexed for ready retrieval when it is retired or archived.
- Confirm accuracy, reliability, performance, consistency of processing and the ability to identify invalid or altered records. Confirm through testing and document.
- Ensure the system has an appropriate traceability matrix linking test cases to requirements.

- Confirm that the study-specific application has been configured properly.

## **Validation**

“Validation” is a term applied to different processes, and is sometimes misused or used in a context that may not always be clear. Even when the term “validation” is used clearly and correctly, clear distinctions exist between validation of different systems, processes and contexts. The following descriptions distinguish between different types of validation and processes associated with validation.

- Validation versus user acceptance testing (UAT)—In *Guidance for Industry: Computerized Systems Used in Clinical Investigations*, the Food and Drug Administration (FDA) defined software validation as “Confirmation by examination and provision of objective evidence that software specifications conform to user needs and intended uses and that the particular requirements implemented through the software can be consistently fulfilled.”<sup>1</sup> UAT is one element of the examination, and documented UAT results serve as one component of “objective evidence” supporting the validation process. UAT is performed by users of the database or CDMS, and should test for both false positive and false negative results in all fields and functions. UAT does not constitute validation by itself; other elements of validation include, but are not limited to, the validation plan, requirements specifications, a traceability matrix, a UAT summary and a validation summary.
- Core CDMS validation—CDMS end users must confirm that the system has been appropriately validated prior to its release for operational use (e.g., creating individual studies). This validation is conducted using a SDLC methodology and is typically a collaboration between IT, quality assurance (QA) and end user personnel. Expected system functionality will be defined in a system requirements specification (SRS) document describing the processes followed and testing performed to ensure the product installs the way the manufacturer intended (sometimes known as installation qualification or IQ), that the system is designed according to the manufacturer’s design specifications (sometimes known as operational qualification or OQ), and that the system functions according to stated requirements and the system’s intended use (sometimes known as performance qualification or PQ). Primary system users should review the

results of this testing to determine if the testing has adequately demonstrated the validity of the system. Descriptions of the more prevalent types of CDMS validations are provided below:

- Commercial off-the-shelf (COTS) products—Most software developers are not directly responsible for compliance with regulatory bodies, leaving the sponsor with the ultimate responsibility for this compliance. End users should investigate and assure that the software vendor has developed and maintains the CDMS using SDLC methodology, including design level testing. This assurance can typically be provided by conducting an audit of the software vendor’s development and design level validation.
- Internally developed CDMS validation—The primary distinction for an internally developed CDMS is that internal staff are responsible for developing and maintaining the CDMS. Those staff developing the CDMS should follow SDLC methodology and be held to the same standards as any vendor providing a CDMS. End users should conduct the same UAT and validation activities described in this chapter.
- Prospective CDMS validation—According to the FDA, “Prospective validation is conducted before a new product is released for distribution or, where the revisions may affect the product's characteristics, before a product made under a revised manufacturing process is released for distribution.”<sup>2</sup> This is the type of CDMS validation most frequently performed.
- Retrospective CDMS validation—According to the FDA, “Retrospective validation is the validation of a process based on accumulated historical production, testing, control, and other information for a product already in production and distribution. This type of validation makes use of historical data and information which may be found in batch records, production log books, lot records, control charts, test and inspection results, customer complaints or lack of complaints, field failure reports, service reports, and audit reports. Historical data must contain enough information to provide an in-depth picture of how the process has been operating and whether the product has consistently met its specifications. Retrospective validation may not be feasible if all the appropriate data was not

collected, or appropriate data was not collected in a manner which allows adequate analysis.”<sup>2</sup> Any time a CDMS must be validated while in active use, validation will be more difficult and the validation plan will be more detailed than expected for prospective validation.

- Legacy CDMS validation—Although there is no formally accepted definition for a legacy system, the term is often used to refer to a CDMS that is currently in operation but does not comply with current regulations.<sup>3</sup> Some may consider a legacy system one which was operational prior to the release of 21 CFR Part 11. The first step of validating a legacy system should be to perform a detailed evaluation of risks and gaps between the system and current regulations. After conducting this evaluation, CDM personnel may find the best solution is to move to a different CDMS. If the decision is made to validate the legacy system, the validation should follow the same processes and procedures as retrospective validation.
  
- Validation of an externally hosted CDMS—Although very similar to validation of a commercially available CDMS, validation of an externally hosted CDMS differs in that the vendor’s documentation should also include information relating to infrastructure qualification, networks, servers’ maintenance and logical/physical security measures.
  
- Study-specific validation—After a CDMS has been validated, study-specific validation must be performed to demonstrate that the requirements for the implementation of a given study using the CDMS have been successfully developed, tested and documented. The FDA states that the “Protocol should identify each step at which a computerized system will be used to create, modify, maintain, archive, retrieve, or transmit source data.”<sup>1</sup> The processes involved with study-specific validation will be discussed in greater detail later in this chapter.

## ***Importance of Validation to CDM***

CDM plays a key role in providing high quality databases that meet both clinical and regulatory requirements. Because clinical data managed through a CDMS is the basis for the marketing approval of new drugs, devices, and biologics, it is imperative that companies seeking to market their products be

able to demonstrate the quality, reliability, reproducibility and integrity of data managed during the conduct of a clinical study. Validation provides evidence that the data management system or study-specific database meets its specifications and requirements and is therefore suitable for its intended purpose.

A clinical data manager's objective is to finish a study with a database that is accurate, secure, reliable, and ready for analysis. Any errors leading to assessment of inaccurate data may detrimentally impact the confidence of a study's results and conclusions. As stated by the FDA, "The computerized system should be designed to...prevent errors in data creation, modification, maintenance, archiving, retrieval or transmission..."<sup>1</sup>

## ***SDLC and Validation***

Principles of SDLC are applicable to all types of validation. One can expect that the details of each step will vary between prospective, retrospective, commercially available, internally developed, CDMS and study-specific validation, although the same general principles can be applied to each. The following are phases of SDLC and how they may apply to validation within a clinical study.

### **System User/Functional Requirements**

Before a CDMS is designed or purchased, the requirements of the system should be clearly defined. Every organization conducting clinical research should have an SRS template listing basic IQ, OQ and PQ requirements, as well as system requirements relating to electronic records and electronic signatures as per 21 CFR Part 11.

### **Design and Build**

For either a CDMS or study-specific database design, the process begins with a design of the program or database, which may be presented as a flowchart. Thorough documentation should be made of what the CDMS or database aims to achieve and how it will be achieved. All algorithms and programming codes should also be clearly documented.

## **Testing**

The testing phase of the SDLC is the phase most commonly thought of as validation, although every phase is important to help ensure testing is adequate and effective. Testing should be performed at each step of development, and integrated testing should be performed to ensure all parts work together correctly once the database or CDMS is completed. A traceability matrix should be used to log which tests correspond with each SRS and to document when each test is completed.

## **Implementation**

A database or CDMS should be put into production only after all validation activities have been completed and thoroughly documented. Once validation is completed, a final validation summary report should be produced and signed by all responsible parties.

## **Operation and Maintenance**

Following implementation, CDM should make certain that the system continues to do what it is expected to do. This may be accomplished by maintaining thorough, appropriate documentation of records relating to training, change control/revalidations, protection of programs and data, recoverability, review of use and performance of the system, etc.

## ***Validation Standards***

Validation standards help ensure reproducibility of validation results, enhance system reliability, and ultimately increase quality. Validation standards can simplify executing the validation process by providing an assurance of the accuracy, completeness, and reliability of the CDMS or study-specific database. Validation standards ensure that each iteration of the validation process is performed consistently, thus ensuring the same level of confidence in the ongoing performance and integrity of a CDMS or study-specific database. Although standards vary between organizations, published standards from entities such as the International Organization for Standardization (ISO) and Good Automated Manufacturing Practice (GAMP) can provide a foundation for developing an organization's validation standards.



In spite of the numerous benefits validation standards provide, they should be used in conjunction with a thorough risk assessment. Validation standards can become onerous and difficult to follow if they are inappropriately focused or scaled, but a risk assessment can help prevent CDM from doing far more work than is needed.

## **Validation Plan**

A validation plan gives an overview of the entire validation project, describing the scope of work to be performed and processes and test procedures to be employed. The plan also describes responsibilities of different members of the validation team, which will typically consist of members of various departments, including IT, QA and CDM.

In addition to the validation plan, a validation protocol may be needed, which would be employed for software patch updates, minor software revisions or small software packages that do not warrant a full validation plan. A validation protocol will typically incorporate features of a validation plan, IQ, OQ, PQ and a traceability matrix displaying the test steps that validate each specific function.

## **How to Develop a Validation Plan**

A validation plan clearly describes all validation activities in a manner that elucidates the plan's compliance with company, industry and regulatory standards. Some fundamental elements that should be addressed include an overview of the plan, document approval, document history, system description, roles/responsibilities, validation strategies and approaches, documentation practices, deviation/response forms, discrepancy logs and reports, a traceability matrix, a script error log, and references.

## **Components and Processes of a Validation Plan**

A validation plan should contain the following components:

- Purpose of the validation plan
- Scope

- Project documentation development and reviews
- Schedule/timeline
- Risk analysis
- Development and test tools
- Team resources and responsibilities
- Development and test environments
- Test data sets
- Validation tasks
- Test documentation
- Test definition and execution
- Traceability matrix
- Metrics for project progress tracking
- Criteria for release into production
- Release procedures
- Required approvals
- Reporting

In addition to the components described above, the following processes should be considered:

- Validation testing
  - Test environment, test data or combination of the two
  - Manual
  - Automated
  - Metrics or quantification of validation quality criteria

- Data migration
  - Moving from one data capture system to another
  - Moving from one database to another (e.g. Access to Oracle)
  - Moving to a newer version of the same application and the appropriate revalidations that are required
- Documented processes should define when change control is appropriate
  - SOPs should say when change control is appropriate
  - SOPs should say when revalidation is appropriate
  - When and how much regression testing is appropriate?
- Validation-related risks
  - Business risk (likelihood that the system contains quality problems)
  - Audit risk (impact of negative findings from any sort of audit)

### ***Study-Specific Validation***

After a CDMS has been validated and approved for use within an organization, validation then focuses on study- or protocol-specific database design and implementation. Validation at this phase can be addressed in three major categories: database design, data entry or capture, and other study-specific programming.

Database design should be based on standard data architectures within an organization, as well as on regulatory requirements and industry standards. Utilizing standard ways of designing study databases and maintaining study data allow validation efforts and results to be easily documented, maintained, and leveraged across many projects. If data structure libraries are available, the templates should be accessible and adaptable where necessary to accommodate specific and unique project requirements. When standards are not available, efforts should be made to keep database design and data structures as consistent as possible within projects and, wherever possible, across projects. For example, data structures developed for Phase I studies

should be used throughout Phase II and III studies wherever appropriate. If use of standards is not possible, as in the case of a contract organization designing a database according to sponsor specifications, the specifications are sufficient. When designing a database according to sponsor specifications, every effort should be made to be consistent, particularly if multiple databases are designed for the same sponsor.

At a minimum, database specifications should provide the following information for each variable:

- Name and label
- Dataset label, panel, or other logical group to which the data belongs
- Type (e.g., numeric, character, integer, decimal, date)
- Length (including number of characters before and after the decimal point, where applicable)
- Definitions for all coded values included in code lists
- Algorithms for variables derived or calculated within the database or CDMS

Use of standards simplifies the specification process by providing a shorthand way of indicating standard items that are obtained from existing documentation. Some examples of standards commonly used in clinical research include those published by the Clinical Data Interchange Standards Consortium (CDISC). For more information about CDISC standards, please see <http://www.cdisc.org>.

When testing a study's data capture system, the most important considerations are to ensure that data entered through a data entry screen or captured via some other transfer process (e.g., electronic lab data transfers) map to the correct variables in the clinical study database and that the parameters for the variable correctly house the data provided. Useful validation measures include entering test or "dummy" data into the screens or loading test data transfer files so that output data listings and data extracted from the database can be reviewed to ensure that the variables were correctly added and saved within the database structure. Testing should be performed on all data, regardless of whether the data do or do not meet defined data structures. It is critical to

validate the data field definitions in terms of length and type. Will all study data be accepted by the database? Are variable lengths sufficient to prevent truncating or rounding? Do character and numeric formats provide the necessary output for analysis files, query management software and other modules within the sponsor's overall CDMS? If the database is programmed to flag out-of-range data, are flags appropriately triggering at data entry or import?

Database entry or capture validation testing should help identify key records management issues. For example, the database should not accept entry of duplicate records, and primary key variables should be appropriately assigned and managed by the definition of the database's structure. When discrepancies between the first and second passes of data entry are resolved for double data entry systems, validation should ensure that one record with the correct data is permanently and correctly inserted into the study database and can be extracted. Most importantly, the audit trail for the study should be validated and protected so that all manipulations of the study database or external files are recorded by date, time, and user stamps in an unalterable audit trail that can be accessed throughout the life of the data.

Other examples of study-specific programming are data loading or transfer programming (e.g., loading adverse event coding variables or loading central lab data), and programming written to validate the data (e.g., edit checks, query rules, procedures). This programming includes any code written to check the data and can occur at the time of entry or later as a batch job. This programming must be validated if action is taken regarding clinical data intended for submission as a result of the programming. Examples include programming that identifies data discrepancies such that queries are sent to clinical investigators or in-house data-editing conventions followed for items identified by the programming.

Best practices include identifying all intended uses of study-specific programming and testing each logic condition in the programming based on a validation plan. Algorithms for variable derivations occurring within the database must be validated.

Practical suggestions include utilizing organization standards to document as much of the programming specification and validation plans as possible and code libraries to reduce the amount of new code generated for a protocol. The

entire validation plan can be a standard operating procedure containing testing methodology, scope, purpose, acceptance criterion, approvals and the format for test data and problem reporting.

## **Validation Risks**

The ultimate risk in validation is ending a study with incorrect or unreliable data, which could have a negative effect on patients' safety. There are also risks relating to relevant regulatory bodies such as the FDA. For example, regulatory bodies may not accept positive study results due to inadequate validation or validation documentation.

### **Validation Risks**

- **Scope inappropriate**—Many software packages may have extraneous functionality that is not needed for the study in which the CDMS is used. Timelines and costs may dictate that only components and functions of the CDMS that will be used in the study be validated, however, any components affecting data and outcomes must be validated.
- **Testing inadequate**—All functional requirements must be thoroughly tested. If testing is inaccurate or incomplete, validation may not be considered successful, increasing costs and timelines by necessitating a repeat validation be performed.
- **Evidence insufficient**—Poor documentation is just as much a risk as inadequate testing. If auditors or inspectors are not provided with sufficient evidence to prove an adequate validation occurred, they must assume that an adequate validation did not occur. Examples of insufficient evidence include a lack of change control processes, incomplete UAT documentation or having a pass/fail checkbox without a section properly documenting the results in greater detail. In the case of validation done by a CRO for a sponsor trial, an audit should contain review of validation documentation at the CRO and a confirmation of the validation should be provided to the sponsor as part of their study documentation.

### **Study-Specific Validation Risks**

Because study-specific programming may be perceived to have less impact than programming in a CDMS, study-specific validation may be taken for

granted by some. However, no matter how miniscule the amount of programming performed, any type of validation failure can potentially cause harm to patients' safety or the organization's bottom line. Following are some study-specific validation risks.

- User requirements not clearly defined or documented
- Incomplete testing
  - Thorough program design testing not performed prior to UAT
  - All study-specific requirements not tested
  - All edits/error messages not tested
  - All data points not tested
  - Workflows not tested
  - Challenges not robust or not performed
- Testing is inadequately documented
  - No traceability to requirements
  - Review not evident
  - Anomaly resolutions not clearly documented
  - Lack of objective evidence (e.g., screen prints) to show that the system works as intended
  - Poor organization of documentation
- Staff qualification or training not appropriate
  - Not well trained in testing protocol
  - Not familiar with business process
  - Not familiar with system

- Not familiar with applicable SOPs, testing principles, standards or conventions
- Process roles and responsibilities not well defined
- Inadequate change control processes
- Actualized risk results in financial loss (e.g., responding to inspection/audit findings, loss of clients, repeating study processes, rejected submissions)

### ***Regulatory Impact on Validation***

Those responsible for validation must be mindful of how their validation activities and documentation would be perceived in an audit or inspection by regulatory bodies. Although referring specifically to software, the following statement by the FDA could just as easily apply to study-specific validation. “Software verification and validation are difficult because a developer cannot test forever, and it is hard to know how much evidence is enough. In large measure, software validation is a matter of developing a ‘level of confidence’ that the device meets all requirements and user expectations for the software automated functions and features of the device. Measures such as defects found in specifications documents, estimates of defects remaining, testing coverage, and other techniques are all used to develop an acceptable level of confidence before shipping the product. The level of confidence, and therefore the level of software validation, verification, and testing effort needed, will vary depending upon the safety risk (hazard) posed by the automated functions of the device.”<sup>4</sup>

Although the preceding quote acknowledges some of the difficulties of validation, an external audit or inspection will never say there is too much information or documentation related to system or database validation. Providing auditors or inspectors with a thorough, well-designed validation plan can help impart a comfort level that the validation has been complete and accurate.

### ***Recommended Standard Operating Procedures***

- Study-specific Database Design



- System Validation
  - UAT
  - Validation Documentation

The preceding SOPs are intended to augment the following SOPs recommended in the FDA's *Guidance for Industry: Computerized Systems Used in Clinical Investigations*, which states, "The SOPs should include, but are not limited to, the following processes.

- System setup/installation (including the description and specific use of software, hardware, and physical environment and the relationship)
- System operating manual
- Validation and functionality testing
- Data collection and handling (including data archiving, audit trails, and risk assessment)
- System maintenance (including system decommissioning)
- System security measures
- Change control
- Data backup, recovery, and contingency plans
- Alternative recording methods (in the case of system unavailability)
- Computer user training
- Roles and responsibilities of sponsors, clinical sites and other parties with respect to the use of computerized systems in the clinical trials"<sup>1</sup>

## References

1. US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Investigations*. Rockville, MD: US Department of Health and Human Services; 2007.

2. US Food and Drug Administration. *Guidance for Industry: Medical Device Quality Systems Manual*. Rockville, MD: US Department of Health and Human Services; 1999.
3. ISPE GAMP Forum. GAMP Good Practice Guide: The Validation of Legacy Systems. *Pharmaceutical Engineering*. November/December 2003.
4. US Food and Drug Administration. *General Principles of Software Validation; Final Guidance for Industry and FDA Staff*. Rockville, MD: US Department of Health and Human Services; 2002.

### **Further Reading**

Association for Clinical Data Management and Statisticians in the Pharmaceutical Industry. *Computer Systems Validation in Clinical Research*. Macclesfield, UK: 1997.

International Conference on Harmonisation. Good Clinical Practice: Consolidated Guideline. *Federal Register*. 1997; 62(90): 2571.

*Code of Federal Regulations, Title 21, Part 11, Electronic Records; Electronic Signatures*. Washington, DC: US Government Printing Office; 1997.

### ***Chapter Revision History***

<b>Publication Date</b>	<b>Comments</b>
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
March 2009	Revised for content, style, grammar, and clarity.

**This page is intentionally blank.**

---

# Laboratory Data Handling

October 2009

## *Abstract*

The vast majority of clinical studies use laboratory data, which should be treated with the same rigorous attention to detail and data quality as any other clinical data. This chapter describes different types of laboratories, different types of laboratory data, and important elements of laboratory data handling. In particular, the chapter discusses the importance of standards and reference ranges for laboratory data, as well as principles and processes to help ensure the accuracy and integrity of all laboratory data.

## *Introduction*

The word “lab” (or “laboratory”), is defined by Merriam-Webster as “A place equipped for experimental study in a science or for testing and analysis.” Within the context of clinical data management (CDM), labs are where biologic samples such as blood or urine are sent for analysis, or diagnostic images or data such as electrocardiograms or Holter monitors are evaluated or interpreted. Because the results of these tests do not originate from a case report form (CRF) at a study site, these types of external data are often transferred as electronic files.

Lab data are used in most preregistration clinical studies and proper handling of these data is crucial to the success of a study. CDM personnel are responsible for data integrity throughout all lab data transfer and cleaning activities. CDM personnel may also be involved with setting up standards and processes for their organization to help ensure the integrity of all data, including those from labs.

## **Scope**

This chapter describes differences between various types of labs and lab data, as well as how CDM practices may vary in different situations. For the purposes of this chapter, the term “lab” generally refers to lab vendors, as opposed to lab tests, which will be referred to as “tests” or “lab tests.” Although local and central labs are not the only lab types discussed, the distinctions between local and central labs can also apply to specialty labs, core labs, and virtual central labs. Specialty labs and core labs may operate as either central or local labs, while virtual central labs operate as central labs. Also, most CDM processes relating to lab data handling primarily vary between local and central labs. As such, the main focus of this chapter will be on local and central lab data handling.

Some of the tasks described in this chapter may be joint responsibilities between different groups, just as there may be many different groups involved in the implementation of various tasks. However, clinical data managers need to be conscious of whether or not these tasks have been performed in a satisfactory manner.

## **Minimum Standards**

- Maintain standard operating procedures (SOPs) for all processes relating to lab data collection, transfer, and validation of data loading and data feasibility.
- Identify labs involved with a study as early in study setup as possible.
- Use standardized names for lab tests and units.
- Ensure reference ranges are defined prior to first data receipt when using a central lab.
- Where possible, ensure reference ranges are defined prior to first data receipt when using a local lab.
- Ensure updates to reference ranges are obtained and implemented in a timely fashion.
- Document all data transfer specifications thoroughly when using labs transferring data electronically.

- Determine software/hardware required to access data prior to a test transfer and ensure the format of the data medium is compatible.

### ***Best Practices***

- Use accepted standards such as those from Clinical Data Interchange Standards Consortium (CDISC) when possible.
- Define all lab data standards prior to beginning data collection.
- Ensure reference ranges are defined for population subgroups (e.g., ethnicity) that differ significantly from other defined groups or subgroups.
- Implement a standard process to collect and archive reference range data.
- Use a standard method of data review for local lab data and reconciliation of central lab data.
- Develop a data transfer agreement for electronic transfers and perform quality control of the test transfer.
- Document and confirm all lab variables prior to signing off on data transfer specifications.
- Implement a conversion factor table to standardize conversion of conventional units to the International System of Units (SI).
- Define edit checks for inclusion/exclusion criteria based on lab data and route to appropriate team members to review.
- Use standardized units so that performing edit checks on converted data produces a more consistent review of results.
- Send requests for central lab data corrections using a formalized process, for example, on a correction log sent to the lab vendor to update and return after correcting and resubmitting the lab data file.
- Implement a system to manage data collected outside protocol parameters.

## ***Distinctions Between Types of Labs***

Although data managers most frequently work with central labs or local labs, other kinds of labs include virtual central labs, specialty labs and core labs. These types of labs tend to fall under the categories of local or central in regard to many processes and characteristics. This section details each type of lab and defines which tests and processes they support. Table 1 details advantages and disadvantages of each type of lab. Advantages and disadvantages may vary geographically, due to regional variations in definitions of various types of labs.

### **Central Labs**

A central lab processes lab samples from multiple clinical sites or studies at one central location. These labs often support multicenter and international studies. Central labs can process many types of samples but most commonly process and report clinical chemistry, hematology and urinalysis. Central lab data are typically transferred electronically from the lab to the sponsor or contract research organization (CRO) throughout the course of a study, resulting in rapid and continuous data transfers and improved safety review and study management. Most central labs have their own file formats but are willing to work with sponsors or CROs at the beginning of a project to define data transfer specifications. Establishing these specifications up front streamlines the process of data transfers.

### **Local Labs**

Local labs are labs in close proximity to individual clinical study sites or patients and are most often used when timely results are needed. Local labs may also be known as “regional” or “preidentified” labs in some locations, such as parts of Europe. Local labs are commonly used in oncology studies, where lab results could be the deciding factor on dosing or not dosing a subject. Each local lab must provide a set of reference ranges to the sponsor or CRO, which increases the work needed for all aspects of lab data collection and integration with study databases. Local labs are typically not able to perform electronic data transfers, so sites become responsible for entering this information onto CRFs. This process can be very time-consuming and error prone, resulting in an increase in the number of queries to the site for clarification or correction.



## **Virtual Central Labs**

The virtual central lab (VCL) is typically a group of labs located throughout the world that are under the umbrella of one company (or partnership). The VCL is based on a central calibrator that runs in parallel with lab samples from all labs participating in a clinical study. The calibrator and lab sample results are compared and results are adjusted based on the calibrated value used by all the labs participating in the study. This process reduces the logistics and costs of shipping lab samples.

## **Specialty Labs**

Specialty labs are used to analyze samples or run assays for nontraditional (or esoteric) tests, which are typically tests that take a considerable amount of time and effort to produce. The amount of time needed is typically outside the control of the lab, although the longer timeframe for these test results must be considered when planning a clinical study. Examples of these tests include biomarkers, genetic testing, pharmacokinetics, and isolation of cancer genes. Specialty tests may be conducted by one lab to ensure standardized results, which is vitally important because these test results are often used as primary efficacy variables.

## **Core Labs**

For the purposes of this chapter, core labs are labs that specialize in a particular therapeutic area or body system. Examples of core labs include stem cell core labs, electrocardiogram (ECG) core labs, imaging core labs, cardiovascular core labs, hematology core labs and oncology core labs. Core labs are vitally important in large clinical studies for their accurate results, which may be used to interpret or support primary or secondary endpoints.

**Table 1. Advantages and Disadvantages of Lab Types**

Type of Lab	Advantages	Disadvantages
Central Labs	<p>Uses one set of analytical equipment, methodologies, kits and reagents</p> <p>Provides training and instructions for collection and shipping of samples, as well as safety alert notifications</p> <p>Standardized results from one set of reference ranges and units</p> <p>Access to lab results in near real time once samples are received and analyzed</p>	<p>Logistical support and costs for shipping lab samples</p> <p>The turnaround time needed to receive central lab data may be too long when immediate results are needed</p>
Local Labs	<p>Lower costs and shorter turnaround time due to not having to ship samples</p> <p>Local lab experience with processing samples from their subject population</p>	<p>Greater potential for errors due to paper-based data transfers and differences between reference ranges from one lab to another</p> <p>Variability in the methods used to perform tests</p> <p>Variability in reference ranges and units used for measurement</p> <p>Reference ranges may be more difficult to obtain</p>
Virtual Central Labs	<p>Reduced shipping costs</p> <p>Decreased need for resampling due to samples becoming compromised during shipment</p> <p>Simpler data processing due to having a central calibrator</p>	<p>Requires detailed process and quality control (QC) measures to ensure lab results are reproducible with minimal variance from site to site</p>
Specialty Labs	<p>Highly experienced and qualified for performing specialty tests</p>	<p>Many specialty tests require more time to generate test results</p>
Core Labs	<p>More focused quality control, more accurate results and a higher degree of standardization and specialization within a designated area</p>	<p>Additional time may be incurred for centralized processing</p>

## **Lab Data in Clinical Studies**

Lab data usually fall under the categories of safety, efficacy or specialty data. There are, however, instances where data may fall into more than one of these categories, such as efficacy data that also relate to a safety parameter.

- **Safety Data**—Lab data can be used to identify or quantify deleterious biological processes occurring in a subject. One of the main purposes of safety data is to provide a baseline at screening of a standard battery of tests that can be repeated during the study to ascertain if there are any detrimental changes to a single parameter or panel. Examples include cardiac biomarkers released into the blood when heart tissue is damaged, or glucose levels in a diabetic population. Many lab tests performed in preregistration studies are performed for safety testing. These tests provide data for a warning system to detect potential safety concerns before they are observable as signs or symptoms.
- **Efficacy Data**—Efficacy data are typically lab data relating directly to the effectiveness of the study treatment. For example, in a study of a new drug intended to battle high cholesterol, one of the primary measures would be lab results of the subject's cholesterol levels in the bloodstream.
- **Specialty Data**—Specialty data may consist of genomic, proteomic or pharmacokinetic data from a specialty lab. These data do not always relate directly to safety or efficacy, but may be very informative with regard to underlying biologic or genetic processes. The following types of data are those most commonly collected by specialty labs.
  - **Genomic**—Genomics is the study of the genes of an individual at the DNA (genotype), mRNA (transcriptome) or protein (proteome) levels. Another variant of the study of genomic data is pharmacogenomics, which is the study of how an individual's genome affects the body's response to drugs. Pharmacogenomics may be instrumental in personalizing treatments for greater efficacy and safety.
  - **Proteomic**—Proteomics is the study of proteins produced by an organism or system, particularly the proteins' structures and functions. The proteome is the entire complement of proteins, including modifications made to a particular set of proteins. Proteomics is often

considered the next step after genomics in the study of biologic systems. Proteomics, however, is much more complicated than genomics, because while an organism's genome is constant, the proteome differs from cell to cell and over time.<sup>1</sup>

- Pharmacokinetic/pharmacodynamic—Pharmacokinetics studies drug absorption, distribution, metabolism, interaction and excretion. Drugs exist in a dynamic state within the body, and different drug events often occur simultaneously. To describe a complex biologic system, simplifying assumptions are often made concerning the movement of drugs. A pharmacokinetic model is conceived using mathematical terms, which are a concise means of expressing quantitative relationships. The intensity of the pharmacologic or toxic effect of a drug is often related to the concentration of the drug. For example, monitoring the concentration of drugs in the blood or plasma confirms that the calculated dose actually delivers the plasma level required for therapeutic effect. Pharmacokinetic models allow more accurate interpretation of the relationship between plasma drug levels and pharmacologic response.<sup>2</sup>
  
- Biomarkers—Biomarkers are substances that are objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention. According to some experts, to be defined as a viable biomarker, the biomarker should meet the following conditions:
  - ◆ Highly sensitive and specific in detecting a desired characteristic
  - ◆ Validated in postmortem confirmed cases
  - ◆ Standardized with sound bioinformatics
  - ◆ Specific for the desired characteristic compared with related disorders or biologic states
  - ◆ Reliable in many testing environments and labs
  - ◆ Minimally invasive
  - ◆ Simple to perform
  - ◆ Inexpensive<sup>3</sup>

## **Standards**

The more standardized lab data are, the easier they will be to collect, process, combine, analyze and submit. Although standardization during study setup is optimal, standardization may also be performed during lab data collection or analysis of final results. A number of data standards have been published or are in development by Clinical Data Interchange Standards Consortium (CDISC), including a standard specific to lab data (LAB). For more information on CDISC standards, visit <http://www.cdisc.org>.

## **Test Names**

Test names are the easiest and most common part of lab data to standardize. If using a central lab, a list of test names should be provided by the lab at the inception of the study. If using a local lab, test name standards would be applied when setting up the clinical database and CRF entry screens.

CDISC terminology for lab test names and test codes can be used to standardize results for a local or central lab. The CDISC controlled terminology model consists of an alphabetical accounting of the most common test names (long name) and test codes (short name). By utilizing CDISC controlled terminology for test names and test codes, sponsors and CROs can reap the benefits of less conversion time when preparing for submissions to regulatory bodies. In addition, if multiple studies are being conducted, the format of data has been established, and table templates and analysis dataset structure can be predefined and programmed earlier in the process.

## **Units**

Although not encompassing all potential analytes, the most universal format to capture lab data is the International System of Units (SI). The following quotation describes SI units and the history of their development.

‘SI units’ is the abbreviation for le Systeme International d’Unites. These units are the result of over a century of international cooperation to develop a universally acceptable system of units of measurement. The SI is an outgrowth of the metric system that has been widely used throughout most of the world, but which has had little impact outside scientific fields

in the United States, even though Congress passed the Metric Conversion Act in 1975, which endorsed the SI.

The SI is a uniform system of reporting numerical values permitting interchangeability of information between nations and between disciplines. The SI not only provides a coherent system of units, but also ensures that units are uniform in concept and style. A coherent system is one in which interconversions between the units for different properties requires the factor 1 only. With the SI, quantities can be more easily compared by means of the reduction in the number of multiples and submultiples in common use.<sup>4</sup>

SI units have almost complete worldwide acceptance and do not need any further conversion. In addition to an SI unit, most tests in the US are also associated with a conventional unit, which is typically based on US measuring methods. When lab test results are collected, the data must be standardized and converted to one common unit before analysis can begin. This can be a time-consuming task, especially when working with multiple local labs, each using a variety of conventional units.

One way to make unit conversion easier is to develop an internal conversion factor table using publicly available references. A table can be created for all tests, listing the most common conventional units as well as the conversion factor to transform to SI units. This conversion table will take significant effort up front; however once completed and verified it will save an enormous amount of time by being applied to subsequent studies.

### ***Unexpected/Unscheduled Lab Data***

During the course of a clinical study, lab tests are performed according to the schedule of the protocol. Sometimes an investigator decides to order a lab test outside protocol parameters, usually when a subject is experiencing adverse events or exhibiting symptoms of another disorder. When these lab tests are performed, they are considered unexpected or outside the protocol.

When the results of these tests are received from a central lab, they may be kept in a separate dataset from protocol-specified tests or flagged to ensure they are easily recognized and are not part of eventual study analyses. When these tests are received from a local lab, the CRF should be designed to

capture results from these unexpected tests. Because these tests will not usually be known in advance, the CRF should be as generic as possible to accommodate study-specific variations, but should include the following fields.

- Test name
- Test result, reference range upper and lower limits, high and low values, and units
- Lab name
- Sample collection date
- Comments section for capturing why tests were ordered and to describe results of the tests

Unscheduled lab data, on the other hand, refers to tests that are within the scope of the protocol but are not performed according to the time and events schedule. This may occur for a number of reasons, including follow-up tests due to previous abnormal values, a subject's unavailability for sample collection at a specified time, or damaged samples (which may be classified as repeat lab tests by some organizations). These tests are captured in the same manner as scheduled sample collections, but must be identified as unscheduled data. For unscheduled results from a central lab, the lab should have a way to differentiate unscheduled sample collections from those that are scheduled. One convention is to have the visit number left blank and the visit name labeled as "Uns" or "U" for unscheduled, although some organizations may design a numbering convention in advance for these circumstances. The sample collection date will then be used to sequence the sample collection among others for that subject. For local labs, the CRF should capture the lab name, sample collection date and unscheduled status.

### ***Lab Reference Ranges***

Lab results are of little value without the ability to analyze the results in comparison to other values. Lab results are typically either compared with other samples taken from the same subject at a different time point (e.g., baseline values), or are compared with a reference range. Reference ranges

can also be known as “normal ranges,” although not all populations can be considered truly “normal.” Reference ranges are established by analyzing a large number of samples and statistically determining the appropriate reference range. Because values may differ according to variables such as age, gender, disease processes, or regional variations, multiple ranges are often established for a given test. Labs may either establish their own set of reference ranges or obtain ranges from published sources. Reference ranges typically consist of a high value, a low value, the unit of measurement, and an effective date. Reference ranges can also be age- and gender-specific, necessitating identification of these parameters. These values need to be collected only once per study unless there are changes to the specimen collection, instrumentation or methodology. Lab relicensure may also trigger the need to update documentation of reference ranges.

### **Use by Clinicians During a Study**

In clinical studies physicians use lab results to determine if a subject meets study enrollment criteria and to monitor the subject’s safety profile or efficacy effects, which may be attributable to the treatment received or from existing or new conditions. Physicians may use other tests to confirm a diagnosis or eliminate error due to false-positive results. They are aware that the reference range provided by a lab has confidence limits and that some normal individuals will have a value outside the reference range. Therefore, most physicians will consider a result normal if it is within the reference range, suspicious if it is slightly outside the range, and abnormal if it is considerably outside the range. Ultimately, the clinical assessment will determine if a particular analyte has clinical significance.

### **Use by Statisticians in Data Analysis**

Biostatisticians view lab values through summaries of data, often comparing the proportion of subjects with out-of-range values to the proportion of subjects with values within the expected range. Biostatisticians also look at changes within subjects and summarize and compare those changes between treatment groups. Shift tables are used to present categories of test results before and after an action, such as study treatment, presenting classification comparisons such as “High-High,” “High-Normal,” “High-Low,” “Low-Normal,” “Normal-Normal,” etc. Biostatisticians also use flags present in the



lab data as cut points to identify out of range values, such as “H” for an abnormal high value or “C” for a critical value.

### **Collection of Reference Ranges**

ICH 8.2.11 requires that “...normal value(s)/range(s) for medical/laboratory/technical procedures(s) and/or test(s) included in the protocol...” be located in the files of the investigator/institution and sponsor.<sup>5</sup> Also, the Clinical Laboratory Improvement Amendments (CLIA) require that labs have reference ranges for all test results produced. The collection of reference ranges is imperative to appropriately handling lab data.

### **Changes in Reference Ranges**

ICH 8.3.7 requires that “Updates to normal value(s)/range(s) for medical laboratory/technical procedures(s)/test(s) included in the protocol...” be located in the files of the investigator/institution and sponsor.<sup>6</sup> Reference ranges are generally not changed or revised unless a new methodology is adopted, primary reagents are modified, or new instrumentation is introduced into the lab. Minor changes in the reference ranges of an analyte may not be significant due to the precision of the method. However, if there is a change in units or a large shift in the reference range, the new range should be used for any results after the effective date of the change. Changes to reference ranges and the effective date of the change(s) should be quickly communicated by the lab and/or investigator to the sponsor or CRO, and all changes should be clearly documented.

### **Importance of Population-specific Ranges**

Many variables complicate establishing reference ranges, including sex, age, ethnicity, weight, geography, or time of specimen collection. Reference ranges should be defined for each subgroup that differs significantly from another subgroup. When ranges are not divided into subgroups, there may be a broadening of the reference range and loss of discriminatory power. Variations in reference ranges are most commonly seen between different sex and age groups.

## **Lab Processes in Studies**

### **Local Labs**

When using local labs, more responsibility is placed on the site to record information. The process begins with obtaining and identifying a sample, then sending it to the local lab for analysis. Once the sample is tested and the report is received at the site, it is the responsibility of the primary investigator or subinvestigator to assess the lab report and determine if out-of-range values are deemed clinically significant (CS) or not clinically significant (NCS). If out-of-range values are deemed clinically significant, the site investigator(s) must then determine if these values are due to an underlying disease state or constitute an adverse event (potentially even a serious adverse event).

The presence or absence of clinical significance is recorded on the hard copy lab report, which becomes the source documentation. In order to incorporate this information into the clinical database, the reported information can be entered into the database from the lab reports or transcribed onto a CRF and entered with the same processes applied to all other CRFs. Although more labor-intensive, the latter solution is cleaner and more consistent with other overall study processes. If CRFs are not used, the database should be set up to minimize transcription errors by mirroring the lab report, and may contain some of the following items.

- Local lab name
- Sample collection date (and time, if collected more than once during a visit or for pharmacokinetic analysis)
- Result field for each analyte
- A single “not done” box for the full panel, as well as “not done” boxes for each analyte
- CS/NCS check boxes (evaluation of which is typically the responsibility of clinical reviewers)

Reference ranges (high value, low value and units) and effective dates are collected at the beginning of the study and if reference ranges change. The corresponding lab name should also be collected on a CRF so that during reconciliation between local lab reference range data and the clinical database,

as well as for statistical analyses, the results and reference ranges can be merged to create a complete file.

## **Central Labs**

When using a central lab (or any lab that transfers data electronically), the lab and sponsor will complete a data transfer agreement (DTA) during study setup. The DTA defines the format of files, frequency of data transfer, file naming conventions, encryption levels, method of transfer, type of transfer (complete versus partial), recipient, test names, formats, high and low value flags or alerts, and any additional information concerning the lab data. A very important part of the DTA is the definition of data that need to remain blinded. If the result of a certain test could potentially identify which treatment a subject is randomized to or if the subject is responding to treatment, these results need to be blinded. Typically, blinded results remain blank in the file until the clinical database is locked and an unblinding memo is provided. Once this unblinding memo is supplied, the lab releases the information and analysis can occur. The DTA should also include range or data checks being performed by the lab, as well as reconciliation processes.

## ***Cleaning Lab Data***

### **Typical Types of Errors**

The most common types of errors from central lab data are demographic errors. When a sample is sent to the lab, a requisition form is completed to identify the subject number, site, sample collection date and time, birth date and gender of the subject (optional) and visit number. If an error is made on the requisition form, this information may differ from the clinical database and prompt a query to be sent to the site or lab.

Careful review and tracking can be used to identify data errors. Review each subject record for values outside defined reference ranges, as well as for consistency of values and units of a given test across multiple visits. If reference ranges are lacking, they should be carefully tracked to ensure all values are associated with the correct reference ranges.<sup>7</sup>

For local lab data not received electronically, the most common errors occur when transcribing results from the printed lab report to the CRF. These errors

should be caught by the monitor when reviewing site data, and if caught by the monitor, will not directly impact data management personnel.

Other types of errors encountered may include:

- Interchanged values—Certain values are particularly susceptible to these errors, such as dates, which may be presented differently in the US and Europe.
- Errors in decimal placement—One example would be specific gravity values, which typically have three decimal places (e.g., 1.014). However, sometimes the decimal may be missing, leading to the value being incorrectly recorded as 1014.
- Errors in units—The majority of errors seen in lab data involve inconsistent units. This may happen if different labs are responsible for performing the test for different visits, if the reference ranges and units change during the study, or if the results are recorded in a unit of measurement that differs from that of the reference ranges.
- Misinterpretation of written values, symbols and units—Handwritten numerals, such as 1 and 7, may be misinterpreted due to illegible handwriting on the CRF.

### **Self-evident Corrections**

Self-evident corrections (SECs) are not applicable for electronically transferred data (typically central lab data) but can be used for local lab data if agreed to by all responsible parties. When using local labs, reference ranges should be collected at the beginning of the study for each local lab used at each site. The corresponding lab name should also be recorded on the CRF so that during reconciliation between local lab data and the clinical database, as well as for statistical analyses, the results and reference ranges can be merged to create a complete file. If the lab name on the CRF has been entered incorrectly or misspelled, an SEC can be performed to enter the correct lab name. In order to apply an SEC, the data manager should carefully examine the data to ensure there are no doubts as to the correct information. This will ensure that the correct reference ranges are merged with the corresponding results.

## **Cleaning Local Lab Data**

Lab data recorded on paper CRFs should be subjected to the same data cleaning and edit check specifications as other CRF data, but extra attention should be devoted to verifying subject and lab vendor identifiers. If a local lab transfers data electronically, the measures described in the following section on central lab processes should be adopted.

## **Cleaning Central Lab Data**

Once a test transfer is received, the sponsor or their designee should perform a quality control check of the data against the DTA to ensure completeness and adherence to the defined structure. If the test transfer is acceptable, regular transfers can begin and reconciliation with the clinical database can commence. The key parameters for reconciliation are information such as subject ID, subject initials, visit or collection date, visit number, visit name, sex, date of birth or age, and test or panel name, although some of these parameters may be optional.

If discrepancies are observed during reconciliation, a query should be sent to the clinical site to verify or correct the information in question. If the query is returned from the site indicating data in the clinical database are correct, the lab data need to be updated according to agreements made with the lab. Some organizations may reverse the order of this process by querying the lab prior to querying the site. When having central lab data corrected or updated, the information should be sent to the lab on a correction log and the lab should update the log once the correction to the data file has been made. This log not only serves as internal documentation during an audit, but also provides the lab with documentation as to why the change was requested and who requested the change. When the changes are made at the lab, a newly updated data file should be sent and reconciliation programs run again. This cycle should occur after every lab data transfer until the data are clean and the clinical database is locked.

## **Edit Checks for Lab Data**

Some standard edit checks that can be applied to lab data include:

- Invalid specimen dates or times

- Blank data, including lab names
- If collecting clinical significance, flagged or out-of-range lab data should be appropriately identified and an associated adverse event should be recorded, when applicable.
- Instances when one test value requires another test value to be provided. For example, if the total bilirubin is greater than 1.0 mg/dL, a direct bilirubin value should be provided.
- Inclusion/exclusion criteria involving lab data can be programmed into edit checks, where appropriate, for flagging when values exceed protocol-defined criteria.
- Listings should be used to compare abnormal results to medical history, adverse events, or other appropriate data.

### **Lab Accreditation/Certification**

According to the International Organization for Standardization (ISO), accreditation is determined as “a procedure by which an authoritative body gives formal recognition that an organization or a person is competent to carry out specific tasks,” whereas certification is defined as “a procedure by which a third party gives written assurance that a product, process, or service conforms to specific requirements.”<sup>8</sup>

### **Clinical Laboratory Improvement Amendments (CLIA)**

In the US, the term "accreditation" refers both to authorization of labs and to certification of procedures and processes. In 1988, Congress passed CLIA to establish quality standards for lab testing regardless of where the test was performed. The requirements are based on test complexity rather than the type of lab where the testing is performed and are intended to ensure the accuracy, reliability and timeliness of subject test results.

CLIA requires all facilities that perform even one test on “materials derived from the human body for the purpose of providing information for the diagnosis, prevention, or treatment of any disease or impairment of, or the assessment of the health of human beings” to meet certain federal requirements. If a facility performs testing for any of these purposes, it is

considered a lab according to CLIA and must obtain a certificate from the CLIA program. CLIA also requires an inspection by the state Department of Health or an accreditation organization such as the College of American Pathologists.<sup>9</sup>

## **International Accreditation/Certification**

The development of quality systems in medical labs of the European Union is based on adherence to the requirements of ISO standards (primarily ISO 15189:2007). The process of accreditation in most European countries is carried out by cooperation among national accreditation bodies, medical experts appointed by scientific associations and health departments. This collaboration has proven successful in the UK, Germany, Hungary, France and Croatia.

## ***Regulatory Agencies***

Although it is not a legally binding document, *ICH Guidelines for Good Clinical Practice* provides a solid framework for determining what lab-related documentation should be retained for a study. The regulatory requirements of individual countries will in most cases be very similar to these guidelines, and in some cases the regulatory agencies may be less stringent. Although the ICH guidelines are a great resource, CDM personnel should always consult the regulations of the country in which the study is being conducted. Information regarding regulations from various countries can be found at <http://www.hhs.gov/ohrp/international/HSPCompilation.pdf>.

For all studies using lab data, *ICH Guidelines for Good Clinical Practice* recommends the following information be kept in the files of the investigator/institution and sponsor.

- Reference values or ranges for all medical/lab/technical procedures or tests
- Changes or updates to reference values or ranges for all medical/lab/technical procedures or tests

- Documentation of certification, accreditation, established quality control, or other validation (where required) of all medical/lab/technical procedures or tests
- Documentation of changes or updates relating to certification, accreditation, established quality control, or other validation (where required) of all medical/lab/technical procedures or tests<sup>10</sup>

### **Recommended Standard Operating Procedures**

- Data Cleaning
- Laboratory Data Entry
- Laboratory Data Transfers

### **References**

1. Pelton L, McKusick VA. Dissecting Human Disease in the Postgenomic Era. *Science*. 2001, 291: 1224-1229.
2. Shargel L, Yu A. *Applied Biopharmaceutics and Pharmacokinetics*. Norwalk, CN: Appleton-Century-Crofts, 1985: 25 ff.
3. Scherer SW, Cheung J, MacDonald JR, Osborne LR, Nakabayashi K, Herbrick JA, et al. Human Chromosome 7: DNA Sequence and Biology. *Science* (April 10, 2003).
4. Young DS. Implementation of SI Units for Clinical Laboratory Data. *Annals of Internal Medicine*. 1987; 106:114-129.
5. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Section 8.2.11; 1996.
6. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Section 8.3.7; 1996.
7. Joshi B. 'Lab Data' in Clinical Trials; Evolving Perspectives. *Data Basics*: 14:3, pgs. 13–15.



8. Gligic L. Status of Development and Implementation of Medical Laboratories Accreditation in Serbia. *JMB*. 2008, 27: 144-147.
9. US Department of Health and Human Services. *Clinical Laboratory Improvement Amendments (CLIA) Overview*. Available at <http://www.cms.hhs.gov/CLIA>. Accessed January 8, 2009.
10. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, 1996.

### **Further Reading**

Eastman JR, Bixler J. Serum Alkaline Phosphatase: Normal Values by Sex and Age. *Clinical Chemistry*. 1977, 23(9): 1769-1770.

Prokscha S. *Practical Guide to Clinical Data Management*. New York, NY: CRC Press, 2007:75-88.

Snelling LF. Local Lab Ranges. *Data Basics*. 2008; 14:3, pg. 17.

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
October 2009	Initial publication.

**This page is intentionally blank.**

---

# External Data Transfers

May 2007

## *Abstract*

Data collected from external sources can be essential to the quality of a clinical trial. This chapter reviews some of the types of external data that may be utilized within a clinical trial and discusses the best practices for handling such data. Processing steps for the validation, editing, and verification of external data are examined, and the importance of key variables is emphasized. Discussions are included concerning file and record formats, transmission of data, procedures for database updates, and archiving of external data.

## *Introduction*

Often during the conduct of a clinical trial, much data external to the case report forms (CRFs) will be collected. If not included in the primary safety or efficacy parameters, these data can be used for subject screening, routine safety and quality-of-life monitoring, or trend analysis. To speed up this process and minimize the use of different analyzing methodologies and equipment, it is common for sponsors to refer to the use of centralized vendors. Such vendors provide electronic transfer of computerized data into the sponsor's database, thereby offering quick results, standardized testing, and reference and calibration values applied to data collected across study sites with the potential to eliminate transcription errors and key entry of data. This chapter focuses on the structure and handling of external data most often required in clinical trials.

## *Scope*

What follows is the data management perspective of the challenges involved in incorporating any external data into a clinical database while assuring that the quality, integrity, confidentiality, and plausibility of the clinical

information is maintained. Further, processing steps that affect the data quality are identified, and a solution framework proposed.

Since regulatory guidance exists and data interchange standards have already been proposed, this chapter will reference on a smaller scale (but not attempt to fully cover) the subjects of providing data for regulatory submissions, clinical data interchange standards (FDA,<sup>1</sup> CDISC<sup>2</sup>), and validation of computer programs (FDA,<sup>3</sup> ACDM/PSI<sup>4</sup>).

For information specific to the handling of laboratory data, see the chapter of *Good Clinical Data Management Practices* entitled “Laboratory Data Handling.”

### ***Minimum Standards***

- Establish the procedures for collecting, transferring, loading, validating, and editing external data through sponsor and vendor collaboration.
- Identify and involve vendors as early in the process as possible.
- Identify key individuals for communication and follow through.
- Provide written specifications for loading external data into the sponsor’s database. In advance of loading the data, identify and agree upon mandatory fields or critical variables.
- Maintain a documentation trail.
- Ensure that parties involved have written standard operating procedures and documentation to support that the SOPs have been followed.
- Establish written procedures for safeguarding the blind when primary efficacy data are collected externally.
- Apply quality control procedures to each stage of data handling to ensure that all data are reliable and have been processed correctly.

### ***Best Practices***

- Audit external data providers on a regular basis as part of your vendor audit practice (see also the Vendor Selection and Management chapter).

- Enforce a formal data clarification process for handling data discrepancies and data updates.
- Validate all programs and systems used for processing clinical trial data in a clinical research environment (see also the Database Validation, Programming, and Standards chapter).
- Provide vendor-specific training. A clear understanding of what is expected by both sides is critical for quality and efficient conduct of the clinical research.

### ***Types of External Data***

External data can originate from different sources, but it is a common practice for a centralized vendor to specialize and produce one or more major data types. Examples of data types include:

- Laboratory and PK/PD Data
- Device Data (ECG, Flowmetry, Vital Signs, Images, and other)
- Electronic Patient Diaries

It is significantly important to identify and describe the variables that must be included in any data transfer, regardless of where the data originate or the information contained within the data transfer. The purpose of these variables is to merge the external data to the sponsor's clinical database; safeguard the blind; and ensure that data belonging to a particular protocol, investigator, and subject cannot be loaded to a subject enrolled into a different protocol or to an incorrect visit. Working with the end goal in mind, one can observe that these data may constitute an integral part of the dataset domains proposed by FDA/CDISC for submission:<sup>1,2</sup>

<b>Dataset</b>	<b>Description</b>
DEMO	Demographics and subject characteristics
DISPOSIT	Disposition
EXPOSURE	Drug exposure
AE	Adverse events
CONMEDS	Concomitant medications
CHEM	Labs – chemistry
HEMAT	Labs – hematology
URINE	Labs – urinalysis
ECG	Electrocardiogram
VITAL	Vital signs
PE	Physical examination
MEDHIST	Past medical history

Refer to CDISC for additional information.

## ***External Data Processing Steps Affecting the Data Quality***

The following areas may adversely affect the integration of external data and should be accounted for during database setup:

- Definition of key variable and mandatory fields
- Data editing and verification procedures
- Record formatting and file formats (e.g. SAS<sup>®</sup>, ASCII)
- Data transmission
- Database updates
- Data storage and archiving

## ***Key Variables***

To ensure that sufficient information is available to identify and process data at the sponsor's site, it is imperative that key variables (those data that uniquely describe each sample record) be carefully selected. Without such variables, it proves difficult (if not impossible) to match patient, sample, and visit with the result records accurately.

While these variables are intended to uniquely identify and clarify subject visit records, incomplete data collection or presentation of errors in either primary or secondary key variables can result in inadequate information. Therefore, completeness in the choice of variables collected and transferred offers a way to increase the accuracy and overall quality of the process. Primary (protocol subject identifiers) and secondary (additional subject and unique vendor identifiers) key variables can include the following:

<b>Primary Key Variables (Protocol subject identifiers)</b>	<b>Secondary Key Variables (Additional subject and vendor identifiers)</b>
Sponsor Name / ID	Subject's Gender
Study / Protocol ID (any combination of project and protocol)	Subject's Date of Birth
Site / Investigator ID	Subject's Initials
Subject Identifier (Subject Number, Screening Number or number assigned by the CRF used)	Transmission Date / Time
Clinical Event ID (Visit Number)	Date associated with the Subject visit
Sample ID (vendor or device specific sample identifier or a subject visit)	Sequence Number (when more than one observation per record exists)

Data acquisition forms, whether conventional or electronic (i.e., CRF, e-CRF), should be designed to facilitate the full and accurate reporting of key information at the study site.

Parties involved in the process should identify in writing and agree in advance upon key variables or fields for loading external data into the sponsor's database. They should also avoid duplication of information. For example, if subject initials and date of birth are already in the database from the CRF and are not selected as primary keys, these variables should not be transferred on the external file. The key variables and value ranges should be specified in advance so that they can be incorporated in range-checking programs.



When any of the efficacy parameters are collected in the external data, particular attention should be paid to safeguard the blind. For example, bone density indicators in an osteoporosis trial may be collected with a study's lab data and could be blinded to the physicians and clinical personnel at the sponsor's site. In case of full double-blind or full triple-blind trial, these data must only be disclosed to parties not directly involved in the trial or data safety monitoring committee. A written procedure must exist describing how this data will be handled and to whom it can be disclosed before the clinical database lock. In a similar scenario, subjects may be excluded from the efficacy analysis for loss of baseline data if any of the pre-treatment blind results are incidentally revealed to personnel directly involved in handling the subject.

### ***Data Editing and Verification Procedures***

For quality and timely processing of data, errors must be eliminated at the source or as close to the source as possible. To facilitate this goal, sponsors and vendors must work together to develop editing and verification procedures. These procedures should include:

- Provisions for treatment of partial data
- Checking for duplicate demographic details and results (real or near real time where possible)
- Range of subject numbers allocated for the study, investigator, or both
- Range of treatment codes allocated per study, investigator, or both

The sponsor and vendor should identify key individuals for communication and follow-through. A representative from clinical data management should be included. It is recommended that the sponsor provide a range of subject and treatment codes for each protocol before external data are received for integration. The allocated ranges should be included in data validation routines and any discrepancies handled as part of a formal discrepancy management mechanism. Very often, a centralized vendor (ECG, laboratory organization) with quick results turnaround time will be able to identify and resolve data discrepancies before any other clinical information is entered into the database or even reviewed.

The vendor should perform duplicate record checks as subject visit data is received. Duplicates should be resolved following a formal data clarification process with the investigative site.

Whenever possible, the sponsor should provide the vendor with a complete listing of subjects' demographic details or IVRS demographic data for an independent reconciliation of the sponsor database and remote database during the study conduct or before database lock.

The vendor and sponsor should agree upon procedures for assuring that the sponsor receives complete data. If partial records are included in a data delivery, they should be indicated as such. The vendor should provide procedural verification and assurance that a hard copy of the results is identical to the electronically transferred results. Any changes to the system or the programs used to create either of the reports must be tested and documented accordingly. If data are transformed during processing, a comparison of the original data and observations to the processed data should always be possible.

If applicable, the vendor should provide a complete list of reference values and their effective dates at the onset of the study. Procedures to minimize the possibility of changes during the course of the study must be implemented.

Definition and details of the process for resolution of discrepancies between external and CRF data should be established as part of the study setup. The process should address the issues of both sponsor and vendor or third-party participant.

## ***Record Formatting and File Formats***

Quality and efficient integration of data demands up-front consensus between the sponsor and vendor with respect to record and file format. Areas for initial discussion include the size of data fields, clarification of numeric versus character fields, decimal granularity, use of characters such as ">" and "<", quotation marks, commas, and other special characters. Special consideration should be paid to handling of null or missing data.

Depending upon the characteristics of the database management systems and expertise at the sponsor and vendor sites, there may be a wide variety of

acceptable record, field, and file formats. Thus, both parties must negotiate in writing a mutually acceptable and detailed record format structure.

Areas to consider include the following:

- The sponsor should provide in writing a complete list of reportable variables in the order required. If data is requested in a SAS dataset, the output of the CONTENTS procedure should be provided as part of the specification. For ASCII files, the column positions or delimiter, record heading, and field justification should be specified.
- Character and numeric fields should be differentiated. Field formats should be specified, in advance, as numeric or character. Reporting of results that can be either character or numeric should be minimized.
- Sponsor requirements on date and time reporting should be negotiated and specified in writing; examples include DATE9., YYYYMMDD or TIME5., HH:MM (24 hr).
- Procedures should explicitly describe the handling of greater-than (>) or less-than (<) signs. Absolute values should be used where possible or to separate the numeric and character portion of the observation into two fields.
- If applicable, comments regarding the condition of the sample or its non-availability should be reported in a field that is separate from the results.
- The test data in the agreed upon format should be available in a file to be used during database set-up and validation at the receiving Sponsor or designee. Successful generation, transmittal, receipt, loading, and screening of the test data validate the data transmittal process.

Data management professionals should evaluate and leverage the experience of some of the existing and emerging vendor independent standards for data interchange between clinical systems, including HL7,<sup>5</sup> ACDM's Standards for Electronic Transfer of Laboratory Data,<sup>6</sup> and CDISC.<sup>2</sup>

## **Data Transmission**

Problems encountered with transmission of data from vendor to sponsor will result in data being lost or incorrectly loaded. To facilitate the transmission process in all cases, complete naming conventions and labeling information must be established. Any data transferred between the vendor and sponsor must contain sufficient information to be uniquely linked to the source of the data and corresponding project and protocol. Origin, date created, date sent, number of records, and a version-controlled file naming convention should be followed.

Public encryption mechanisms such as *PGP*<sup>®</sup> (*Pretty Good Privacy*<sup>®</sup>) are recommended for use when transferring data via the Internet. Thus, the data transfer process will ensure compliance with the regulatory guidelines and provide authenticity and confidentiality protection. Not all countries allow the use of strong encryption software. In such cases, consider the use of password-protected files such as ZIP archives or dial-up FTP transfer. Both processes will verify the integrity of the file being transferred and provide feedback in case of file corruption.

## **Procedures for Database Updates**

The processes by which updates to subjects' records are made are among the most vulnerable for generation of errors. Special consideration should be paid if the edit affects any of the primary key variables, and thus propagates multiple records (see also the Data Processing chapter).

Errors generated by the data-cleaning process in the sponsor's database should be communicated back to the vendor for follow up and resolution through a formal data-clarification process. To update a record when the original records are either incomplete or contain erroneous data, the vendor frequently will send a second transmission. Updates can be sent either as a full or partial transmission depending upon the capabilities of the systems in place. It is essential that the vendor and sponsor establish procedures that define how retransmissions are identified and handled throughout the study.

Strategies to consider include the following:

- During study set up, provide the vendor with a list of in-house data checks, supporting documentation, and sample subject-number allocations.
- Use correction flags. When possible, two separate types of flags should be used to distinguish an initial record from a correction or addition.
- Corrections to key variables should be identified and flagged. Updates to key variables should be sent as full records (i.e., including result variables) and should be flagged at a record level.
- Only current results should be reported.
- Maintain an audit trail. The source systems should be designed to permit data changes in such a way that data changes are documented and that there is no deletion of entered data.<sup>7</sup>

If applicable, vendors should provide the investigator site and sponsor with updated hard-copy information in addition to electronic updates.

## ***File Storage and Archiving***

Ultimate responsibility for the quality and integrity of the trial data always resides with the sponsor.<sup>8</sup> Thus, the sponsor should specify in the contract a definitive time period beyond the initial transmission of information during which the records will be maintained by the vendor for access by the sponsor and regulatory agencies. It is desirable that vendors maintain active copies of data files during the study stages that require unconstrained accessibility. After these stages, the vendor should maintain an archived version for the remainder of the retention period. When all reports have been finalized and the sponsor's database has been locked, a study should no longer require access to the records except for auditing purposes during the record-retention period.

For additional information, see the Data Storage chapter, the Database Closure chapter, and the FDA's *Guidance for Industry: Computer Systems Used in Clinical Trials*.<sup>3</sup>

## **Recommended Standard Operating Procedures**

SOPs should be established for, but not limited to, the following:

<b>Sponsor (CRO)</b>	<b>External Data Provider (Vendor)</b>
External Data Loading and Validation	Data Extraction and Validation
Query Generation and Vendor (remote) Database updates	Data Transfer and Discrepancy Handling
Vendor Auditing	Database Updates
Database lock procedures	Database Archiving and Security
Study-specific procedures (including the handling of extra/unscheduled data)	Study-specific procedures (including the handling of extra/unscheduled data)

## **References**

1. US Food and Drug Administration. *Guidance for Industry: Providing Regulatory Submissions in Electronic Format, General Considerations (IT2) and NDAs (IT3)*. Rockville, MD: US Department of Health and Human Services; 2003.
2. Christiansen D, Kubick W. *Submission Metadata Model: Version 2.0*. Austin, TX: Clinical Data Interchange Standards Consortium; 2001. Available at: <http://www.cdisc.org/standards/SubmissionMetadataModelV2.pdf>. Accessed February 20, 2007.
3. US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Rockville, MD: US Department of Health and Human Services; 1999. Chapter VI, sections C and D.

4. Association for Clinical Data Management and Statisticians in the Pharmaceutical Industry. *Computer System Validation in Clinical Research, a Practical Guide*. Macclesfield, UK: Author; 1997.
5. HL7 Working Group. *Health Level Seven Standard Version 2.5: An Application Protocol for Electronic Data Exchange in Healthcare Environments*. New York: American National Standards Institute; 2003.
6. Association for Clinical Data Management. *Standards for Electronic Transfer of Laboratory Data*. St. Albans, UK: Author: 1997.
7. *Code of Federal Regulations, Title 21, Volume 1, Part 11: Electronic Records; Electronic Signatures*. Washington, DC: US Government Printing Office; 1998.
8. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E6, Section 5.2*. Geneva, Switzerland: Author; 1996.

### **Further Reading**

Bloom JC. Drug Information Association (DIA). Clinical Trial Symposium, January 12–13, 1998.

Bornstein S. CDM task list. *Data Basics*. Fall 1999.

Goodman DB. Standard for ECGs in clinical trials. *Drug Information Journal*. 1996;30:237–242.

Peairs D, Ginsberg A, Gillibrand D, Malcolm R. New technologies enhance central laboratory data collection, transfer, and analysis. *Applied Clinical Trials*. 1996;5.

Pharmaceutical Manufacturers Association, Clinical Data Management Group Task Force. Electronic exchange of laboratory data: Recommendations for improving quality and reducing the hidden costs. *Drug Information Journal*. 1993;27:933–942.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
October 2009	Name of chapter changed from “Lab and Other External Data” to “External Data Transfers” and a reference to the “Laboratory Data Handling” chapter added to the Scope of this chapter.



---

# Patient-Reported Outcomes

July 2009

## *Abstract*

Clinical studies frequently rely on patient-reported outcomes to fully evaluate the efficacy of a drug, device or treatment. This chapter differentiates between traditional and electronic methods of capturing patient-reported outcomes and discusses features of each approach. The chapter also examines the impact of regulatory requirements on patient-reported data collection.

## *Introduction*

Certain situations necessitate information be reported by a subject rather than being objectively measured by study personnel. These types of data are known as patient-reported outcomes (PRO). PRO data give researchers the opportunity to quantify subjective experiences, which can be crucial in studies that measure, for example, symptoms, disability, emotional state, social functioning, or subjects' perceived response to symptoms, treatments or disability. PRO data also allows for data collection outside of scheduled visits and without any investigator interpretation of data.

The processes involved with the development of a PRO questionnaire are part of a field known as psychometrics. Psychometrics can be described as a scientific discipline concerned with the theories and techniques of questionnaire construction, quantification, testing and validation. Psychometrics relies heavily on statistics, and a well-designed questionnaire will reliably measure the attribute(s) it is designed to measure.

Many clinical studies use PRO data to some degree, and for some of these studies PRO data is a primary efficacy parameter. If an investigator cannot observe, quantify or measure a variable, it may instead be reported by the subject. The best way to learn about an individual's personal experience is to

ask the individual to describe it. In some cases, PRO data can also be used for registration or economic evaluation purposes.

Although PRO data have been used in clinical studies for many years, the advent of electronic tools may improve the quality of data and ease of data collection. Electronic patient-reported outcomes (ePRO) provide potential advantages over traditional PRO collection methods, but there are circumstances where ePRO may not be the best choice for a study. Each study should be individually evaluated to determine which collection method is most appropriate.

## **Scope**

This chapter discusses various methods of capturing both traditional (sometimes referred to as paper-based or non-electronic PRO) and electronic PRO data. The chapter also describes different types of PRO data, advantages of different data collection methods, and clinical data management (CDM) considerations when choosing a PRO or ePRO method for a study. For this chapter, the term “questionnaire” refers to any instrument or measure, including patient diaries, used to collect PRO data.

## **Minimum Standards**

- Provide detailed instructions to subjects for completion of any PRO questionnaire, whether electronic or paper-based.
- Employ strict version control on all PRO questionnaires. Changing one word of one question, changing the order of questions or changing the instruction to subjects can invalidate the comparability of results.
- If PRO data-collection tools are administered in multiple languages, ensure all translations are linguistically consistent.
- Ensure processes are in place to assure compliance with regulatory requirements regarding protection and ownership of ePRO data (electronic source data).
- Consult with information technology personnel to ensure networked ePRO tools have the appropriate level of network security and infrastructure.

- Ensure ePRO systems are properly validated in compliance with US 21 CFR Part 11 and any other regulations specific to the location of the study.<sup>1</sup>

### **Best Practices**

- Use PRO only for variables that cannot be directly measured. For example, do not ask a subject if their reflexes have slowed when a neurologic exam can be administered instead.
- Use standardized, validated questionnaires when possible to avoid needing to test the psychometric properties of a newly developed PRO questionnaire. For example, there should be no need to develop a questionnaire to screen for depression when many standardized questionnaires already exist.
- Document the development processes for newly developed PRO questionnaires.
- Document any study-specific modifications or revisions to PRO questionnaires. Refer to the FDA guidance to confirm whether changes require additional testing before implementation.<sup>2</sup>
- Conduct appropriate training and retraining (as necessary) with subjects to familiarize or refamiliarize them with the PRO questionnaire used. Giving subjects hands-on experience with ePRO tools may be more critical than with paper-based PROs, but adequate training should always be ensured regardless of collection method.
- Use a standard predetermined structure for collecting subject data (i.e., interview scripts, questionnaire layouts, electronic devices, telephone prompts).
- Ensure PRO questionnaires have been thoroughly psychometrically tested. Consult with a statistician for any questions about quality of psychometric testing.
- Avoid post data collection queries for missing or inconsistent data, as these data are a subjective account of the subject's experience at a particular point in time and no additional source is available to cross-

check reported data. Because queries are not generated for PRO data, the resultant database may contain inconsistencies that are not addressed, regardless of whether ePRO or traditional paper-based PRO capture methods are used.

### ***Data Suitable for PRO Collection Methods***

One approach to PRO data capture is intended to quantify the status of particular conditions or symptoms. These PRO data are often collected with questionnaires that ask questions about symptoms associated with a condition and use an established algorithm to quantify the results. This can include the state of discrete symptoms such as pain severity, or can be an assessment of the overall state of a condition such as depression or asthma. Numerous questionnaires exist employing this approach, many of which have been psychometrically tested. Some of these questionnaires may be protected by licenses, and appropriate authorizations may be required to use the questionnaires or modify them in any way, even if this modification only involves transferring the questionnaire to an electronic format or a different language.

Another type of PRO data involves a subject's self-assessment of feelings or opinions. This can include feelings or perceptions about a condition or treatment, or a self-assessment of current or recent emotional states such as depression or anxiety. Assessments of a subject's feelings about a treatment must oftentimes be designed for each study, as questionnaires may need to be tailored to the specific treatment.

PRO data may also involve subjects' self-assessments of their activities, social or physical functioning. These types of assessments may ask the subject how their social or physical functioning has been over a period of time, and these assessments frequently ask about specific activities or interactions. A number of widely accepted questionnaires exist to quantify some of these types of PRO data.

Finally, PRO data may consist of a subject reporting the frequency of certain events, such as bowel movements, headaches or taking pain medication. Because these data consist of counts, little or no psychometric testing may need to be applied. These data may be used as an indicator of treatment exposure, efficacy or safety or as an indicator of study compliance.

## **Traditional PRO Collection Methods**

The following methods are sometimes referred to as paper-based PRO, although in some cases the subject may report these outcomes verbally while a researcher records the information through an electronic medium.

- In-person or telephone interviews—In this form of PRO data collection, a researcher elicits verbal responses from a subject. The interview should be scripted and administered consistently following established guidelines. This approach reduces likelihood of a question being overlooked; however, subjects may be reluctant to share some personal information with an interviewer. Although this may be described as a paper-based PRO collection method, researchers may record subject responses on paper or electronically.
- Paper questionnaires—This is the most commonly used PRO collection method, and psychometrically tested paper questionnaires already exist for much PRO data. When subjects visit a study site for an assessment, treatment or follow-up visit, they complete paper forms designed to quantify various PRO data. Paper questionnaires may also be mailed to subjects for completion. A disadvantage of paper questionnaires is that subjects may not always answer all questions. When questionnaires are completed at a study site, a cursory review may confirm whether a question was intentionally left blank. Mailed questionnaires, however, are especially prone to missing responses.
- Paper-based diary—Diaries are meant to assess subjective information when subjects are going about their normal lives. Subjects may be asked to enter information at certain intervals (e.g., daily or hourly) or may be asked to record when certain events occur, such as an asthma attack or insomnia. Although diaries can capture a wide range of subjective information, they are susceptible to subjects filling in information both backward and forward in time. This is sometimes known as “parking lot compliance,” where a subject completes pages meant to cover a range of time, but instead completes all the pages at once, for example, in the parking lot of the doctor’s office before a visit.

## **Characteristics of Traditional PRO**

Traditional PRO collection methods sometimes hold advantages over ePRO. A side-by-side comparison of traditional PRO characteristics and ePRO characteristics can be found in Table 1. Some of the characteristics of traditional PRO capture include:

- Fewer startup resources (e.g., hardware, software, technical support) are typically needed for traditional PRO capture.
- Minimal setup time is usually required for traditional PRO capture.
- Site personnel do not need to train subjects in use of the capture instrument, because most people are familiar with paper questionnaires.
- Because some subject populations may be more comfortable with paper than electronics, there could be a potential for bias in subject selection if ePRO is used.
- Traditional PRO capture methods are not as susceptible to the impact of technology failures such as battery depletion, device malfunctions, busy telephones, Web server crashes, or ineffectual help desk support.
- Traditional PRO may be associated with compliance issues due to lack of subject surveillance.

## ***ePRO Collection Methods***

Technologic advances have enabled researchers to utilize electronic tools to capture PRO data. In most cases these tools provide greater overall efficiency or improved data quality, but they may not be the most appropriate solution in all cases. The study team should evaluate each study on a case-by-case basis to determine the best approach. The following tools are in common use today, but additional PRO data capture methods may emerge in the future.

- Handheld devices—Although hardware costs typically make study setup more expensive than other methods of capturing PRO data, personal digital assistants (PDAs) or other handheld electronic recording devices allow comprehensive capture of PRO data. These devices combine portability, ease of use, and the ability to capture a wide range of PRO data. The convenience and portability of these devices promotes reporting

information in real time rather than asking subjects to remember a prior period of time and then accurately report the requested information. Many devices can also be programmed to provide subjects with reminders of scheduled times to record information.

- **Web-enabled reporting**—This approach allows subjects to fill out questionnaires or diaries from a computer connected to the Internet, which allows comprehensive data capture, but lacks the portability of a handheld device. In addition to allowing a full range of PRO data collection, Web-enabled reporting can be relatively inexpensive to set up.
- **Interactive Voice Response (IVR) systems**—This method utilizes automated interactive telephone systems to capture PRO data. These systems typically are not as ideally suited for collecting as wide a range of data as handheld devices or Web-based questionnaires. For example, phone systems do not have an acceptable way to complete a Visual Analog Scale (VAS) or indicate the precise location of pain. Subjects also may not always have access to a phone when calls are meant to be made. Although this is classified as an ePRO capture method, subjects may perceive this as more similar to some traditional methods.

## **Characteristics of ePRO**

When used effectively, ePRO can provide some advantages over traditional PRO data collection methods. A side-by-side comparison of traditional PRO characteristics and ePRO characteristics can be found in Table 1. Some of the characteristics of ePRO include:

- **Greater data accuracy** may be associated with ePRO because improved surveillance may promote more timely data entry.<sup>3</sup> More timely data entry may translate into more accurate or more complete subject reporting. With paper-based diaries, there could be a tendency for subjects to fill in a week's worth of data in the parking lot of the site before a visit.<sup>4,5</sup> Using ePRO allows all data entry to be date and time stamped, helping to ensure the subject is entering information based on their recall at that specific point in time. This helps to avoid recall bias, which can be a confounding factor in many of the subjective measures that are captured by PRO instruments.

- Potential for improved subject compliance is provided through some ePRO systems' features such as automatic reminders, as well as the convenience of portable ePRO devices that allow more flexibility of when and where data are entered.
- Potential for fewer errors exists with ePRO because of the lack of ambiguous or unusable data due to illegible handwriting that may be associated with traditional PRO capture methods. The potential for fewer errors can also be facilitated by various front-end edit checks such as minimum/maximum values, time windows, response choice rules, etc.
- Reduced burden and increased convenience for subjects can be provided by ePRO capture methods, in part because ePRO offers question branching, which can allow for fewer, but more targeted questions.
- Quicker sponsor access to subject-reported data can be provided by ePRO capture methods, enabling proactive real-time study management. This could be helpful for studies with adaptive design. For example, if a study design includes study endpoints or decisions that are contingent upon PRO data, ePRO can provide advantages by making these data available more readily.
- Electronic integration with the clinical database is accommodated by many ePRO capture methods.
- More confidential collection of sensitive data may be achieved by ePRO capture methods.
- Some ePRO tools offer the opportunity to integrate interactive training for subjects.
- ePRO is not always readily available for some locations or populations.

### ***Choosing a Pro Method***

CDM personnel should carefully evaluate which PRO or ePRO capture methods will provide the best results for a study. The following factors should be considered when determining whether to use ePRO or traditional PRO.

- Complex or lengthy questionnaires may not be suitable for some capture methods, such as IVR systems.



- Studies of long duration may be subject to changes in technology used for ePRO data collection.
- The degree of psychometric testing applied to the mode of administration should be considered.<sup>2, 6</sup> For example, an established questionnaire will typically have more scientific validity than a structured interview.
- Clinical subject population and demographics may affect the suitability of some methods.
- An ePRO system's compliance with regulatory requirements should be thoroughly examined.
- Connectivity and data transmission abilities relative to the operating locale must be taken into account. For example, Web-enabled reporting may be a poor PRO capture method in a rural area with limited Internet access.
- The technological capabilities and quality standards of available vendors should be considered.
- If a study design involves study endpoints or decisions that are contingent upon PRO data, ePRO can provide advantages by making these data available more readily.

The following table presents some considerations of both traditional PRO and ePRO such that the two collection strategies can more easily be compared and contrasted.

**Table 1. Comparison of Paper-Based PRO and ePRO Methods**

<b>Consideration</b>	<b>Paper-Based PRO</b>	<b>ePRO</b>
Startup resources	Fewer startup resources are typically needed.	More startup resources are usually needed, but this may not be true for organizations already using ePRO.
Setup time	Less setup time is typically required.	More setup time may be required unless the organization has used similar ePRO instruments in the past.
Costs	Less startup costs are typically incurred.	Startup costs are usually higher, but overall study costs may be lower.
Data accuracy	Data may be more prone to errors because of legibility issues and data entry errors.	Data may be less prone to errors because legibility and secondary data entry become irrelevant.
Subject compliance	Subjects may be prone to “parking lot compliance,” where data cannot be definitively tied to a time of entry.	Subject compliance is better monitored with electronic date and time stamps.
Data accessibility	Data are not entered into the clinical database as quickly, because PRO data must be entered as CRF data.	Many ePRO capture methods allow electronic integration with clinical databases, allowing much faster access to PRO data.
Subject training	Subjects typically do not need to be trained in use of the capture instrument, because most people are familiar with paper questionnaires. However, subjects may need training in completion of the questionnaires used.	Subjects may need training in use of the capture instrument. However, instructions for completion of questionnaires may be integrated within the ePRO device used.

### **Costs Considerations**

One of the most common arguments against using ePRO is that it is in many cases more expensive to set up than a traditional PRO approach. Although ePRO costs may be more expensive during study setup, the possible savings ePRO may offer become more apparent when viewed in the context of an

entire study. A careful assessment of the following areas will help determine the most cost-effective manner to collect PRO data for a study.

- General considerations (fixed vs. recurring costs, amortization of fixed costs over multiple studies, need to acquire in-house expertise vs. outsourcing)
- Resources available to the organization at the beginning of a project
- Hardware and software expenses
- Personnel additions or reductions (less personnel may be needed for data entry and data cleaning depending on the collection method)
- Training requirements and costs
- Help desk or support needs for subjects and investigator sites
- Adequate infrastructure for hosting ePRO data
- Adequate disaster recovery plans to ensure continuous access to ePRO data
- Additional tools that may be used with ePRO (such as IVR systems, handheld devices or a Web page)
- Programming and integration costs must be considered for ePRO
- Validation requirements when going from paper to ePRO (e.g., when you already have a fully validated paper-based questionnaire)

### ***Regulatory Considerations***

Use of PRO or ePRO is subject to the same regulatory oversight as any other tool used in a clinical study. CDM personnel should take measures to ensure regulatory requirements are met for sponsors, vendors and investigators in regard to record keeping, maintenance and access. All ePRO tools should be fully compliant with 21 CFR Part 11, including a comprehensive audit trail preserved for every step of the data collection and handling processes.

*Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (DRAFT)* provides guidance for the selection or creation of a PRO questionnaire or test. It should be noted that this draft guidance also states, “If a patient diary or some other form of unsupervised data entry is used, the FDA plans to review the protocol to determine what measures are taken to ensure that patients make entries according to the study design and not, for example, just before a clinic visit when their reports will be collected.”<sup>2</sup>

For widely accepted questionnaires, ample information usually exists in the scientific literature to support the psychometric properties of the instrument. However, for modified or newly created instruments, “The FDA generally intends to review a PRO instrument for: reliability, validity, ability to detect change, and interpretability (e.g., minimum important differences).”<sup>2</sup>

### ***Recommended Standard Operating Procedures***

- CRF Design
- CRF Completion Guidelines
- Data Review
- ePRO System Validation
- Vendor Management

### ***References***

1. *Code of Federal Regulations, Title 21, Part 11, Electronic Records; Electronic Signatures*. Washington, DC: US Government Printing Office; 1997.
2. Food and Drug Administration. *Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (DRAFT)*. Rockville MD. US Department of Health and Human Services; 2006.

3. Hufford M. Placebo Effects, Memory, and the Value of Real Time Data in Drug Development. *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2002.
4. Hufford M, Shields A. Electronic Diaries: Applications and What Works in the Field? *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2002.
5. Stone A, Shiffman S, Shwartz, J, Broderick J, Hufford M. Patient Non-Compliance with Paper Diaries. *British Medical Journal*; 2002.
6. Coons SJ, Gwaltney CJ, Hays RD, et al. on behalf of the ISPOR ePRO Task Force. Recommendations on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Task Force Report. *Value In Health*. 2009;12;419–429.

### **Further Reading**

European Medicines Agency (EMA). *Reflection Paper on Expectations for Electronic Source Documents Used in Clinical Trials (DRAFT)*. London UK. EMA; 2007.

Hufford M, Stone A, Shiffman S, Schwartz J, Broderick J. Paper vs. Electronic Diaries: Compliance and Subject Evaluations. *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2002.

McKenzie S, Paty J, Grogan D, Rosano M, Curry L, Sciarappa, Hufford M. Proving the Diary Dividend. *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2004.

Shiffman S. Delivering on the eDiary Promise. *Applied Clinical Trials*. Walnut Creek, CA: The Biotech Communications Group; 2005.

US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Washington, DC: US Department of Health and Human Services; 2007.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
July 2009	Initial publication.

---

# **CDM Presentation at Investigator Meetings**

July 2008

## ***Abstract***

Clinical data management professionals serve an important role at investigator meetings, especially when the trial is large, complex or multisite. This chapter covers the procedures clinical data management professionals should follow when preparing a presentation for such meetings, including presenting examples of case report forms, discussing various types of error-checks, reviewing the role of the data manager, and emphasizing the proper use of data clarification forms.

## ***Introduction***

The investigator meeting provides an early opportunity for data managers to be in contact with site personnel for a clinical trial. It is often a joint effort among the project manager, the clinical team leader, the statistician, and the clinical data management (CDM) lead to describe procedures for preparing, conducting, and managing multicenter investigational trials. A CDM presence at this meeting should provide a well-rounded overview of the data collection strategies for a given study.

## ***Scope***

This task commences when the study team takes on the responsibility for preparing and conducting the meeting and ends when the required meeting documentation has been distributed following the meeting.

## ***Minimum Standards***

- The data manager should prepare their assigned presentation and materials for the meeting.

- Materials should include sample case report forms (CRFs), CRF completion guidelines, data queries and self-evident corrections.
- The data manager should prepare a visual presentation on the overall data collection process, including non-CRF data such as laboratory, ECG and imaging data.
- Role-based guidance should be provided for all project team members involved in the data cleaning process.
- Training on CRF and query completion should be documented.
- The data manager should present an overview of data collection processes in the study for adverse events (AEs) and serious adverse events (SAEs). In some organizations, the SAE collection process may be presented by a representative from a safety or pharmacovigilance group.
- Provide a presentation on the coding of AEs and concomitant medications.
- If self-evident corrections can be made for the study, the process should be addressed in the presentation, including site sign-off and examples of where this will be utilized.

### ***Best Practices***

- Avoid use of acronyms. If you must use an acronym, spell it out in the first instance and then use the acronym in subsequent references.
- Provide a copy of the presentation for all participants.
- Record the session (audio or video) for use in subsequent training as appropriate.
- Allow sufficient time at investigator meetings to answer CRF-related questions.
- Facilitate a breakout session where CRF completion exercises are performed and evaluated for common errors.
- If studies were completed in a previous indication with similar CRFs, provide targeted training based on discrepancy metrics per data panel or



field. Also provide training for edit procedures to address the most common failures.

- Provide, or support the preparation of, materials that are best suited for the type of meeting that will occur. For example, slides and flowcharts are appropriate for presentations or Web-cast meetings. Other presentation methods are more appropriate for self-study. It is best to consult with experts to determine the most appropriate method for presenting information.

## **Procedures**

The purpose of a CDM presentation is to familiarize site investigators and staff with case report forms, the electronic data capture (EDC) system and equipment if applicable, and clinical data management procedures such as CRF completion guidelines, data collection, and the query process. At minimum, the data manager should present the CRF completion guidelines and query workflow process at the investigator meeting.

In the past, investigator meetings were face-to-face meetings. Now, many companies are conducting investigator meetings via the Web, or preparing self-paced training modules to complete online or via CD. Some programs can track the amount of time a user was online, as well as the assessment score achieved on an e-learning module. Data management should ideally be included in investigator meetings because of the expertise in data collection and integration methodologies, as well as its ability to inform the investigative and clinical staff of the most effective and efficient measures to take to enhance data quality and timeliness.

CRF presentations should use completed CRFs as examples. It is valuable to present a CRF packet containing all unique CRFs completed as if they contained actual patient data. This would allow attendees to see proper data recording for various requirements of all CRF pages. Every effort should be made to have final, approved CRFs prior to the investigator's meeting. However, if a CRF is not finalized prior to the meeting, the study team should be reminded to plan sufficient time for CRF changes to be made before the start of the study. Another strategy would be to incorporate into the final draft any study coordinator or investigator data collection feedback provided at the

meeting. Valuable feedback on the reality of treatment may drastically reduce the discrepancy load and reduce queries to the sites.

The presenter should demonstrate consistency checks between pages, and should point out potential spots for errors. Some of the cross checks that can be discussed include, but are not limited to:

- Compare current medical history to any concomitant medications. For example, if a subject has hypertension and is taking medication, it should be appropriate to show that they are taking an antihypertensive.
- Compare medical history to physical examination. For example, if a subject has a history of bronchitis, the physical exam may show bronchitis.
- Compare termination page to AE page. For example, if a subject withdrew from the study due to an AE, an AE should be indicated on the AE page and the action taken should be discontinuation from study.
- Compare AE where a medication was administered to the concomitant medication page to ensure medication has been documented with the appropriate indication, which should be noted in the AE.
- Provide an example where efficacy and safety data show a logical progression. For example, compare baseline vital signs with subsequent vital signs.
- Make certain investigational product accountability corresponds with dosing regimens outlined in the protocol, as well as drug return logs in patient packets.
- Check that visit dates are in range with visit windows specified in the protocol.
- Compare subject history and physical examination to basic eligibility criteria.

The CDM presenter should use the opportunity to explain the data manager's role in the overall scheme of the trial, including but not limited to the following:

- Establish project standards for handling partial, missing or incomplete data, as well as illegible text entries.
- Ensure that *Good Clinical Data Management Practices*' guidelines are followed by providing examples that indicate the proper mechanism for making corrections to the CRF.
- Review the amount of time CDM needs to complete milestones and meet timelines.
- Review the process of providing performance feedback to sites, perhaps in the form of trend reports for the data query process.

The CDM presenter should use this opportunity to carefully review the data query process in a step-by-step manner, including but not limited to the following:

- Familiarize participants with various reports that organize data queries by data item or file names. Educate participants on problem areas of CRFs, or common mistakes made during the data query process.
- Demonstrate to site personnel how to address data issues before the monitoring visit in order to achieve best monitoring efficiency.
- Ensure site staff understand that the cleaner data are to start with, the quicker database lock can occur. The sites need to remain available to answer queries and questions at least until database lock.
- Explain any status reports site staff may receive that are based on data in the database. Some examples include outstanding CRF reports, outstanding data query reports, or outstanding lab samples.
- Explain the relevance of reports and any related workflow information. For example, if sites are paid for CRFs entered as of the 25th of each month, provide the site with the send-by or submit date that will assure data will be entered and included in the report run on the 25th.
- Describe the procedures for clarifying CRF questions or issues.

- For EDC studies, allow sites the opportunity to either participate in hands-on entry or to see a live demonstration of key points from the study and the EDC system.

### ***Recommended Standard Operating Procedures***

- Data collection and handling procedures
- Handling of standard clarifications or obvious data corrections

### ***References***

N/A

### ***Further Reading***

N/A

### ***Chapter Revision History***

<b>Publication Date</b>	<b>Comments</b>
January 2002	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
July 2008	Revised for content, style, grammar, and clarity.

---

# Training

May 2007

## *Abstract*

Clinical data management employees must receive the training necessary to complete their project-related responsibilities effectively and successfully. This chapter reviews the various factors to consider when adopting a training program for CDM employees. Approaches to the development of master training plans and training plans for individual employees are discussed. Topics which should be covered in data management training are reviewed. Effective strategies for facilitating the learning process are presented, including an overview of the principles of learning and different techniques for introducing course material to trainees. Online training is introduced as a solution to time and logistical constraints, and considerations for choosing and developing online training are reviewed. Trainer qualifications, the training environment, and evaluation and feedback from trainees are included as important factors to consider when adopting and maintaining a training program.

## *Introduction*

An effective training program plays a key role in ensuring regulatory compliance, performance effectiveness, and job satisfaction of clinical data management (CDM) employees. There are a number of compelling reasons for developing and implementing effective training programs. Good Clinical Practices (GCP) and other regulatory guidance documents state that all personnel involved in clinical trials must be qualified and properly trained to perform their respective tasks.<sup>1,2,3</sup> Changes in the technical, business, and regulatory environments are occurring more rapidly than ever. As a result, the demands placed on CDM personnel and the scrutinies under which they must perform are increasing. Trainers must be sensitive to these demands and ensure that they optimize the value of the training experience for participants. This chapter discusses the design, development, and delivery of effective training programs for clinical data managers. Core topics for inclusion in CDM training are also discussed.

## **Scope**

This chapter addresses issues relevant to CDM training. It includes a brief overview of classroom training, as well as computer-based and web-enabled training issues and techniques.

## **Minimum Standards**

- Document learning objectives for each component of the curriculum.
- Review and update curriculum and individual course offerings regularly, including applicable SOPs, to ensure that content remains current and relevant.
- Train all CDM staff members to perform the job functions that are currently required for their assigned roles.
- Ensure that training documentation is maintained and includes, at minimum for each course, the name of the course offering, the course objectives, the name of the course instructor, the date of the course offering, and the names of attendees. Ensure that this documentation also includes training that occurs outside the organization.

## **Best Practices**

- Document a role-specific training curriculum for each position within the CDM organization.
- Ensure that a master training plan, which is regularly reviewed and revised, documents and prioritizes training needs of the CDM function.
- Perform job-needs analyses and audience analyses to guide development of the training plan.
- Develop and document customized development plans for each employee according to the employee's career objectives and personal development needs.

- Evaluate each training curriculum to determine if the class is best suited for instructor-based training, online user training, or a combination of the two.
- Make training support available for all online user training.
- Evaluate the effectiveness of training.
- Use a variety of methods to enhance learning during training.
- Ensure that content is consistently represented across all training materials and consistently conveyed by instructors, coaches, mentors, peers and others who assist the learner to master targeted concepts and skills.
- Verify that instructors remain qualified by obtaining and maintaining current knowledge of the topics that they teach.
- Ensure that technical training occurs in a separate training environment that simulates the actual CDM work environment.
- Document the organizational responsibility for training in Standard Operating Procedures (SOPs).
- Ensure that managers actively allocate time for employee training and development. The amount of time that should be allocated depends on the needs of the employees and the organizations.

### ***Master Training Plan***

Training should be considered at both a macro level (i.e., overall training needs) and micro level (i.e., specific training needs). Appropriate training topics, such as computer systems usage, (SOPs) and working practices, should be included in a master training plan. The master training plan should be reviewed and approved by all key individuals involved in its development and use.

Training plan design should include an audience- and job-needs analysis. The Society of Clinical Data Management (SCDM) task list and capabilities matrix (both available from SCDM) provide a good starting point for this analysis. This analysis should be customized to the organization's assignment

of roles and responsibilities. When designing the proposed level of instruction, consider entry-behavior skills, knowledge, and abilities (SKAs). To determine the entry behavior, test a small sample of the target audience to establish if the starting points of the training program and threshold knowledge are accurately represented. The analysis should also consider various training-delivery mechanisms to accommodate differences between the learning styles, learning stages, sex, and ethnicity of the members of the audience.

Establish clear learning objectives. This step is critical as learning objectives form the basis for what is to be taught, the performance level, and the conditions under which the task is to be performed. To ensure that the stated objectives are valid and achievable, include a peer review or beta-test of training materials.

Once learning objectives have been established, the training method should be evaluated. Whether the course is most suitable as an instructor-based class or an online course should be determined. A cost analysis of the preferred training method should also be performed to the feasibility of online training. If an internal online training program is being considered, this cost analysis should include a review of the company's infrastructure and resources for maintenance of an internal training website. After implementation, the master training plan and materials should be assessed and updated based on feedback, changes to the audience, and job requirements.

## ***Data Management Training Topics***

This section covers topics that affect the daily work environment of data managers. This list is not intended to be exhaustive of training topics. Rather, it should serve as a reference guide for the development of a master training plan and individual development plans.

### **Standard Operating Procedures and Departmental Policies**

Data management departments are required to have SOPs that describe their processes and operations. All data management employees are required to understand and follow these SOPs. Frequently, required training on an SOP consists of having the employee sign a statement that he or she has read and understood the SOP. However, this practice, used in isolation, often falls short of meeting its intended purpose and should be avoided. An example of a more



effective approach is to have the trainer go over each required SOP with employees and explain how it affects their daily work flow. Such training sessions often encourage questions that reveal inconsistencies in the implementation of the SOP. Follow-up activities, such as study audits, may also reveal such inconsistencies. Issues identified during follow-up may be addressed by revising the SOP, if applicable, or by intervening to change the working practices of employees. As SOPs are revised, training must also be updated.

### **Computer Software and Technical Skills**

For data entry, cleaning, and analysis, data managers use various software applications, including clinical database management systems, case report form (CRF) imaging software, edit-specification development software, discrepancy management software, and others. To use these software packages, employees require training. Depending on time and budgetary restrictions, this training may be performed by the software vendor, a third party training vendor, or an in-house trainer.

### **Regulations and Industry Standards**

Data management is required to work within the constraints of Food and Drug Administration (FDA) codes and regulations. Additionally, industry standards give employees guidance in their common work practices. Information regarding standards such as GCP, ICH Guidelines, FDA regulations, FDA guidance documents, and the GCDMP can be found in various publications, educational seminars, or web sites. Trainers should make such references available to all employees.

### **Professional Growth**

Individual development plans should include topics that focus on the employee's growth outside of the technical skills required. Skills—such as leadership training, effective team skills, time management, conflict resolution, project management, presentation skills, listening skills, cultural diversity, and medical terminology—allow employees to cultivate their professional skills, helping them to be more productive in a group setting. Often, the human resources department can provide outside resources for

conducting such classes. Online courses also offer various training opportunities in this area.

### **Interdepartmental Processes**

To be fully effective, CDM employees must also understand the processes that occur before and after the data is handled in data management (such as site monitoring, safety monitoring, statistical analysis, and FDA submissions). One effective approach is to observe other departmental processes firsthand during cross training. Another effective approach is to invite personnel from other business units or departments to attend data management meetings as guest speakers.

### ***Training Techniques and Environment***

This section describes different training techniques that may be used to optimize participant learning satisfaction. The importance of controlling the environment to enhance the learning process is also discussed. Additional information regarding these methods may be obtained through the reference citations at the end of this chapter.

### ***Principles of Learning***

The principles and techniques described in this section are based on the Interaction Associates, Inc. workshop.<sup>4</sup> To establish an environment that is focused on the learner's needs, a trainer should balance the three principles of *service, respect, and authenticity*. These three principles facilitate the development of a sense of trust between the trainer and participants. The trainer demonstrates service to the participants by being prepared when questions arise, even during breaks. Service may also be exemplified by arriving prepared with innovative methods for teaching the topic. Mutual respect between the trainer and trainees must be established immediately. Creating a set of ground rules and expectations can facilitate an atmosphere of respect. Acknowledging and validating participant concerns and different learning styles also earn respect from the participants. Finally, being honest and genuine creates a presence of authenticity within the group.

## **Strategies in Learning**

Different strategies may be employed to guide decisions and steer the direction of a training session. Using the learning pathway enables trainees to learn new skills by passing through a logical sequence of steps.<sup>4</sup> The first of the five steps in the learning pathway is for the trainer to provide the *definition*, or meaning, of the skill or task. The second step is for the trainer to *validate* why the skill or task is important. The third step consists of *assimilation* or comprehension by the trainee of how the skill or task works. In the fourth step, the trainee must *integrate* how the skill or task is used in the daily working environment. Subsequently, the trainees reach the fifth step and *transition* or incorporate the task with relation to other skills or tasks that they perform in their job. A trainer can organize the teaching of any concept, skill, or task through the learning pathway.

A trainer also needs to balance the importance of *content, process, and relationship* when presenting a topic.<sup>4</sup> To ensure participant satisfaction, the trainer must provide enough *content* to keep trainees interested while covering the objectives and meeting the participants' expectations. However, if the training session is only made up of content, the learning process will be compromised. The trainer needs to think about the *process* or flow of the training session as well. Therefore, the trainer should include all participants in the session, monitor the pace of delivery, and consider the timeliness of each step. The trainer also needs to establish a trusting *relationship* with participants. Doing so promotes a comfort level for trainees and allows them to feel at ease to ask questions and participate in the class.

## **Presentation Delivery/Tools and Techniques**

Learning is best achieved by receiving information through a variety of methods or techniques. This section describes several methods used to present classroom-training materials. Methods often employed for on-the-job training are newsletters, fact sheets, or quick-tip reference guides. Special attention to mentor-based training should be given to ensure consistent delivery of information. The learner should be encouraged to seek clarification and validate information through multiple means rather than relying on a single source of information.

*Lecture* is the traditional method of transferring information from trainer to participant. However, research shows that using lecture alone for an extended period of time does not provide the optimum retention level of training materials. Lecture should be used in conjunction with other learning methods such as those described below. Lecture may be integrated with testing—thereby allowing time for self-assessment—or with a discussion of surveys or training materials within the group.

*Multi-sensory techniques* (e.g., visual, auditory, and kinesthetic) increase the acquisition of training material. Training that impacts as many of the human senses as possible accommodates different learning speeds, styles, and needs. Examples of visually stimulating training are the use of flip charts, colorful presentations, or other visualization techniques. Variation of voice tone during presentations or playing of music can stimulate the auditory senses during training. The kinesthetic sense of touch can be incorporated into training by using exercises with physical movement or objects.

*Group discussion* and interaction among participants is an effective way to present a common topic. Understanding and comprehension of the topic is enhanced when trainees discuss the topic with each other. Discussion of a topic enables a group to establish a personal connection with the content and provides a common basis for shared ideas. Triggers, such as short readings, role-playing, videos, or open-ended questions, help to stimulate discussions by connecting the participants with each other and the topic.

The “Open, Narrow, Close” technique of *questioning* is one approach that allows the trainer to maintain control of group discussions.<sup>4</sup> This technique is applied as follows. First, open up the content of a discussion with a broad question. Then, focus the discussion on a specific area or subtopic that was mentioned. Follow by closing and transitioning the discussion to the next topic. Questions posed by trainees should be recognized by the trainer as a learning opportunity or “teachable moment.” It is imperative for the trainer to understand the question being asked. This understanding can be achieved by paraphrasing the question, providing parallel comments to the question, or asking for clarification or expansion of certain aspects of the question or comment.

Assignments, simulations, games, or other activities are examples of *active learning* techniques. Research indicates that learning is enhanced by physically performing a related activity. Select the type of activity that best

supports learning objectives. Activities might include (but are not limited to) brainstorming, round-table discussions, role-playing, or practicing tasks in a test environment. Using a three-step learning cycle known as a construction spiral<sup>5</sup> is another method to engage trainees in the learning activity. Providing a post-training activity that facilitates the utilization of the new skills in the daily work environment can also be an effective technique.

*Storytelling* allows trainees to relate the topic of discussion to their own daily environment. Stories may relate a similar experience, draw an analogy to the topic being discussed, or give an example of how the topic relates to the participants. However, it is important not to generalize or make assumptions about participants when sharing stories. The trainer/trainee trust-level must be kept intact.

## **Online Training**

Due to time and logistical constraints, it is often necessary to provide online training materials for employees. Online training can consist of outside vendor courses performed via the Internet, as well as internally developed training. This type of training provides flexibility because the class may be taken at a convenient location, time, and pace. Online training is also beneficial since it avoids travel time and expenses involved in bringing employees to a central location for training.

Online training from *outside vendors* should be evaluated for the organization, accuracy, relevance, content, and cost of course materials. Products from different vendors should be compared to determine the most valuable and relevant course for the employees.

*Internal training* may be performed online via a department website on the company intranet. Training materials, such as presentations, examples, case studies, quizzes, glossaries, and tip sheets are easily referenced from the web. Links to other reference materials, such as forms, that are used for training may also be posted.

An internal training website should contain a main menu which lists the courses available. An introduction for each subtopic and a summary of the objectives for the course should also be provided. As with instructor-led training, it is important to measure the knowledge obtained from the course to

ensure that the objectives are understood. It is also important to use the different training techniques discussed earlier to keep the student interested. Visual graphics with screen shots are particularly helpful with online software applications training. With online training, it is imperative that an instructor or resource center be available for questions from the student. Online courses should be constructed to avoid boredom, which can lead to the student skipping sections or doing the minimum work necessary to advance to the next section. Worksheets, study guides, clear endpoints, and rewards for course completion can assist with these issues. Providing a practice environment for users is also beneficial.

Certain navigational features should be considered when an organization is assessing online training, internal or external. Forward, Back, and Exit buttons should be available at each menu to facilitate the student's movement from screen to screen. A Help button should be provided at each step to assist the student in navigation, as well as course guidance. Bookmark and sound-card options are also beneficial.

Accessibility of online training by students with language barriers or disabilities should also be evaluated. Internet services, such as Bobby Worldwide (<http://www.cast.org/bobby/>), can be used to test an internal website for issues related to language, vision, or hearing. Once the website is constructed, it can be sent for accessibility testing regarding obstacles such as sound cards for the hearing impaired, pages that are difficult to read or color-dependent for the visually impaired, and other issues.

## ***Trainer Qualifications***

A data management trainer should have experience in the topic of presentation, as well as experience in training techniques. The trainer must understand industry standards as well as departmental policies. Training techniques and methods may be found in various publications, some of which are listed in this chapter. Also, many companies offer training courses, which can be found on the Internet.

A trainer must always be prepared to handle strategic or "teachable moments." These situations may include an upset participant, an irrelevant or long-winded statement that guides the participants in an unplanned direction, or a compelling comment. When the need for transition from the current situation

to the next step in the training process is recognized, the trainer must choose the best route to reaching that next step. However, the principles of service, respect, and authenticity, as previously discussed, must be maintained during this process so the trainer/trainee trust stays intact.

## **Training Environment**

During training, regulating the physical and mental climate is important. Start by ensuring that pre-work assignments are distributed well in advance of the training event and that expectations are clearly understood. Temperature, room arrangement, lighting, and external noise should be kept at an optimal level during the session. Frequently, climate and tone are set at the beginning of a training session. Beginning with a fun activity, providing food and drinks, or playing music contributes to an optimistic training atmosphere. Closing the training session on a positive note is also important. Summarize the key contents covered during the class. Recognize the efforts of and the goals accomplished by each participant. Encourage participants to establish a plan to implement the topics discussed in their daily working environments.

## **Evaluation and Feedback Techniques**

Implement a 360° feedback process regarding all aspects of the training experience. Feedback should include comments about the trainer, the training materials, and the training environment. Testing may be appropriate at this stage. Explain to the trainees how the feedback will be managed. Encourage trainees to contact the trainer after the training session if necessary.

## **Recommended Standard Operating Procedures**

- Data Management Training Program and Documentation

## **References**

1. International Conference on Harmonization (ICH) Expert Working Group. *Guideline for Good Clinical Practice, E6*, Section 2.8. Geneva, Switzerland: Author; 1996.

2. *Code of Federal Regulations*, Title 21, Volume 1, Part 11.10 (i), *Electronic Signature Regulations*. Washington, DC: US Government Printing Office; 1997.
3. Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Washington, DC: U.S. Department of Health and Human Services; 1999.
4. Interaction Associates, Inc. *The Masterful Trainer<sup>sm</sup>: Principles, Strategies, and Tools* workshop. Boston, MA: Author; 1997.
5. Drummond T. A brief summary of the best practices in college teaching intended to challenge the professional development of all teachers Web page. Available at: <http://northonline.sccd.ctc.edu/eceprog/bstprac.htm>. Accessed February 26, 2007.
6. National Seminars Group. *Trainer's Creativity Day Camp<sup>sm</sup>* workshop. Kansas City, MO: Rockhurst University Continuing Education Center, Inc; 2007.

### **Further Reading**

Clark, D. Instructional system design: Design phase Web page. Available at: <http://www.nwlink.com/~donclark/hrd/sat3.html>. Accessed February 26, 2007.

Kolb, D. *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1984.

Lawson, K. *The Trainer's Handbook*. San Francisco, CA: Jossey-Bass/Pfeifer; 1998.

MacKeracher, D. *Making Sense of Adult Learning*. Toronto: Culture Concepts, Inc.; 1996.

Mager, RF. *Preparing Instructional Objectives* (2nd ed.). Belmont, CA: Fearon; 1975.

Menges, RJ. *Review of Higher Education*. Research on teaching and learning: The relevant and the redundant. 1975;11:259–268.



Trace R&D Center, University of Wisconsin-Madison. *Web Content Accessibility Guidelines 1.0* Web page. Available at: <http://www.w3.org/TR/WCAG10/>. Accessed February 26, 2007.

Wlodkowski, RJ. *Enhancing Adult Motivation to Learn: A Comprehensive Guide For Teaching All Adults*. San Francisco, CA: Jossey-Bass; 1999.

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
January 2002	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.

**This page is intentionally blank.**

---

# Metrics in Clinical Data Management

April 2011

## *Abstract*

A wide range of measurements, commonly referred to as “metrics,” are essential to evaluate the progress and outcomes of a clinical study. This chapter considers various metrics used in clinical data management, as well as the process of selecting metrics that are related to the goals and objectives of an organization. The chapter discusses the importance of standardizing metrics for a project and across projects, and gives suggestions to help ensure metrics are provided in a timely fashion with adequate contextual information to be understood and effectively used to measure, monitor performance and improve efficiencies.

## *Introduction*

The term “metric” simply refers to a measurement. In clinical data management, metrics can quantitatively and qualitatively assess whether or not a process or individual or group performance is efficient and effective, as well as indicate whether the factor being measured has or will have an expected level of quality. Metrics can be used at various intervals throughout a study to ascertain if processes are working as planned. When a process has been completed, well-designed metrics can help indicate if goals were achieved with the expected level of quality.

This chapter provides information on metrics that are particularly relevant to clinical data management (CDM) personnel. There are no regulatory mandates regarding specific metrics; however, metrics can assist in detecting potential regulatory issues, for example by measuring compliance with SOPs. The effective use of metrics also helps an organization evaluate and improve quality and productivity. This chapter is intended to provide helpful suggestions and considerations for CDM personnel involved with establishing a metrics program within a department or company.

## **Scope**

For the purposes of this chapter, the term “metrics” primarily refers to specific data management process-related measurements assessed during the course of a study, but may also refer to the data generated by these measurements. Roles and responsibilities vary between organizations, and some of the topics discussed in this chapter may be the responsibility of different departments in different organizations. Regardless of role assignment, CDM personnel should be aware of the processes discussed in this chapter and how they impact their roles as data managers.

## **Minimum Standards**

- Ensure CDM metrics are aligned with key performance indicators (KPI) (milestones, deliverables, timelines and other quantitative measurements) to meet the organizational needs and goals.
- Ensure that all metrics are clearly defined, quantifiable, documented and approved.
- Communicate approved metrics to relevant personnel and stakeholders within and across projects.
- Ensure adequate and appropriate resources (hardware, software, personnel, etc.) are made available to accurately and thoroughly measure and report metrics.
- Ensure the personnel responsible for defining, quantifying, documenting and communicating metrics have the proper training and relevant skills and competencies.
- Ensure all personnel and stakeholders are adequately trained regarding metrics definition and their relevance to process and project performance.
- Perform quality assurance on data used to determine the metrics, to ensure that the metrics are based on accurate and timely data.
- Establish and document corrective action to be taken if planned or actual metrics do not align with goals and objectives.

## **Best Practices**

- Include all stakeholders (e.g., project managers, clinical leads, data managers, management, etc.) in the development of metrics specifications.
- Ensure all stakeholders (e.g., project managers, contractors, clinicians, data managers, and management) understand and agree with the definition of the measurements and the parameters used to provide each metric before implementing use of the metric.
- Align metrics with project team/organizational goals as well as industry standards and contractual agreements, when and where appropriate.
- Standardize the definitions of metrics by using consistent terminology and parameters across projects and the organization.
- Agree upon well-defined metrics at the onset of a project, and use those metrics to evaluate performance during all stages of the project.
- Select a set of key metrics that apply to all projects. Use these metrics as the basis for comparison of process performance across all projects.
- Consider the aspects of cost, quantity, quality, timeliness and performance when deciding which metrics to implement.
- Identify metrics that will indicate progress to targets and also provide insight into historical performance.
- Ensure that the effort needed to collect and report a metric is appropriately offset by the benefit. Where possible, implement automated collection of data for metrics, and strive to use existing primary data (e.g., audit trails, tracking systems) to collect metrics.
- Ensure the tools used to collect and report metrics are thoroughly validated, and are 21 CFR Part 11 compliant where applicable.
- Establish benchmarks of expected performance based on pooling of similar data.
- Ensure metrics findings are visible to relevant stakeholders via a reporting plan (charts, dashboards, etc.) followed by a feedback loop and rigorous

action plan through root cause analysis (RCA) and corrective action/preventive action (CAPA).

- Document the process for collecting, reporting, and communicating metrics.
- Evaluate metrics collection and reporting processes frequently (for both internal and outsourced activities).
- Determine if metrics need revision, or if other metrics should be added or eliminated, based on changes in technology or process landscape.

## ***Identifying Metrics***

An organization's use of a set of key and relevant metrics will facilitate achievement of predetermined goals. Although agreement on certain metrics is obtained by the overall company or department, individual departments or project teams may need to maintain additional metrics to assess the progress toward the goals of their respective department or team.

Metrics should be based on goals and objectives set by an organization, and ideally, organizations and departments should strive to identify a set of metrics to use across all projects. Identifying the specific metrics that fit the needs of all involved parties is often difficult. Most goals and objectives set by groups or organizations revolve around the interdependent areas of quantity cost, time, quality, and performance, as shown later in this chapter in Table 1.

- **Quantity**—Quantity measurements are straightforward and objective, and are therefore among the easier metrics to quantify.
- **Time**—When measuring time, one of the most important considerations is defining the exact start and stop points and the unit of measure (e.g., business days, calendar days, or resource hours). Time measurements ensure that chronology of milestones is maintained. Organizations may follow a risk-based approach in adhering to the timelines over other metrics.
- **Cost**—Although costs are not typically a CDM responsibility, CDM may supply metrics that are used for cost analyses.

- **Quality**—Quality is the most important metric to be considered in CDM. Quality metrics may measure the quality of processes and deliverables and can be quantified in different ways. For more information about data quality, see the GCDMP chapters entitled “Measuring Data Quality” and “Assuring Data Quality.”
- **Performance**—Metrics intended to quantify performance are typically made up of some combination of measures of quantity, time, cost, and quality. Therefore, performance can also be assessed in terms of one or more of these measures in relation to another measure, such as performance over time, or performance compared to cost. Performance should typically be measured at multiple levels (for example, site, study, project etc.)

When considering a set of key metrics, an organization should design the metrics to allow for their application across projects, regardless of the project-specific process or technology used. This approach allows for an assessment of each project in comparison to similar projects. It also allows for an evaluation of processes that may be redesigned to take advantage of a new technology.

Two examples applicable for clinical studies using either paper-based data collection or electronic data capture (EDC) are: (1) measurement of the number of queries per data field for incoming data as opposed to the number of queries per page and (2) measurement of the time from subject visit to data entered in the database.

Clinical studies are often evaluated within the realm of strategic (i.e., organizational) and tactical (i.e., operational) objectives. Metrics assessments are generally based on the relationship between two or more (e.g., quantity over time, or quality of quantity) of the five core criteria of quantity, time, cost, quality, and performance.

One should be cautioned that focusing too much on one criterion may adversely affect another. For example, focusing too strongly on quality may impact study timelines, similarly focusing too strongly on study timelines may negatively impact quality. All of the above-mentioned criteria should be balanced to some degree in the metrics used by an organization.

Regardless of the measurement, or why a measure exists, a well-designed metric should be

- relevant—answers critical business questions
- enduring—is of lasting relevance
- robust—is not subject to manipulation or variation due to process changes
- valid—measures what it implies to measure accurately
- specific—is clear and consistent
- actionable—can drive decisions
- practical—is measured in a timely fashion without a significant drain on resources.<sup>1</sup>

The effort needed to collect and report a metric should be offset by the potential benefit. If a metric has no benefit, it should not be collected just because doing so is easy and inexpensive. Cost, quality, and performance metrics may be difficult to quantify, whereas metrics dealing with quantities and times are often much easier to collect. The metrics that are collected and reported should be able to answer questions that have been predefined to measure the success or failure of a project or process.

### **Linking Metrics with Organizational Goals**

A hierarchical relationship exists between the objectives of an organization, a department, and an individual project or clinical study. An organization may have strategic objectives that include achieving a certain level of quality in its product while achieving a particular profit margin at the end of the fiscal year. Each functional group within an organization, such as CDM, sets tactical goals and objectives to ensure quality while using resources efficiently. A particular project manager or project team may have budget and time constraints, yet be expected to deliver a quality end product.

Each functional group must develop its own objectives and metrics within the context of the organization's objectives. However, cross-functional input should be solicited to ensure consistent interpretation of the metrics. The



existence of these hierarchical objectives and concurrent timelines drives the need for consistency in the definition and utilization of metrics.

### **Linking Metrics with Project Goals and Deliverables**

Overall project goals and objectives must be considered when metrics are selected and evaluated. A set of metrics that only addresses some, but not all, of the five core criteria will provide only a partial assessment of overall project performance. If one metric is met, it does not imply that the others are achieved. For example, even if milestones are achieved on schedule, they may have required additional resources.

In addition to overall project goals, metrics should also be considered in relation to specific deliverables. For example, if the database lock is scheduled by a certain date, metrics that may indicate the possibility of delays should be carefully examined and communicated. This ensures that all stakeholders have realistic expectations of when the database lock will actually occur.

Even when the same set of metrics is used across projects, they may be prioritized differently for each project. For example, cost containment may be assigned a higher priority in an early phase exploratory study, while data quality may be prioritized in a phase III pivotal trial.

### **Identifying Users**

To optimize the effectiveness and efficiency of metrics, the users of each metric should be clearly identified. Each metric should be linked with documentation of who collects the metric, who reports the metric, and who is responsible for initiating any actions that may be taken based on the metric. If a metric is to be used for evaluating progress toward goals, all such stakeholders should be identified and documented.

Metrics should be shared with all stakeholders participating in a project when applicable, including CROs and vendors. Decisions should be made early in the project planning stages concerning which metrics will be collected, who will collect the metrics, how and when the metrics will be disseminated (e.g., with a common Web site or visualization tool, such as a dashboard, one month after the first patient signs the consent form, etc.).

Metrics results should be communicated to relevant stakeholders clearly and within prescribed timeframes, enabling needed corrective actions to be made in a timely manner.

## **Evaluating Metrics from Various Sources**

Obtaining metrics can be difficult when the parameters required for measurement are found in multiple databases. Even if all of a study's clinical data reside in a single database, data comprising project metrics may originate from a study database, a project tracking system, a CDMS (clinical data management system), or a system outside CDM altogether. This issue is further compounded when certain complementary metrics, such as the project budget and the status of various CDM processes, are not available for equivalent time frames. However, metrics can be synchronized with other relevant information if they are collected in a timely manner.

Automated data generation for metrics that can be shared electronically across various systems, will lower the chance of errors and the effort needed for re-entering the data. The use of technologies such as Web portals, clinical trial dashboards and visualization tools is a viable option for reviewing metrics data allowing proactive control of study progress. All such tools used in the clinical data management environment must be validated to ensure accuracy.

These tools may have the capability to aggregate real-time study data into intuitive views, eliminate the need to integrate databases or re-enter data, and allow for views of complementary data within the same time frame.

## ***Metrics in Different Types of Studies***

EDC systems offer the capability to have clinical data and queries available sooner (in real time) than in paper-based studies. Study or subject status indicators such as subject enrollment or visit completion may also be available within the EDC system. The quality and timeliness of metrics improves substantially when they are collected electronically.

In paper-based studies, CDM metrics can be generated electronically only after data are entered into the database or CDMS. Information regarding subject enrollment, visit completion, and other such status indicators can be difficult to obtain in a timely fashion. Teams often rely on each site to report

this information (e.g. using paper enrollment logs) and then subsequently re-enter the information into a project-management or project-tracking database.

### **Metrics Common to EDC and Paper-based Studies**

Many metrics common to EDC and paper-based studies relate to overall performance of the project, team, or organization. Because metrics measuring organizational or group performance are not contingent upon the data collection modality used, they are also usually independent of any CDMS or database software. Although there are some exceptions, most well-designed metrics are not dependent on a particular data collection strategy or software package.

### **Metrics Unique to Paper-based Studies**

Data entry is one area in which metrics for paper-based studies may be created. An example is the percentage of data entered relative to the number of completed CRFs received. Another example is performance metrics for data entry personnel (number of forms/patients entered per day, per employee). Paper-based studies will also have metrics related to data clarification forms used for query resolution, which are not needed in EDC studies due to the capability of generating queries electronically.

Some metrics used in paper-based studies may have a different meaning when used in EDC studies. For example, data entry percentage may also be measured in studies using EDC, although in that case it is an indication of site performance.

### **Metrics Unique to EDC Studies**

EDC-specific metrics are often directly associated with the EDC system. Examples include the percent of EDC system downtime or the average number and severity of EDC help desk calls. Another class of unique EDC metrics are those that would be prohibitively expensive to measure in a paper-based study, such as the number of modules pending PI review and signature. For more information about metrics in studies using EDC, see the GCDMP chapter entitled “Electronic Data Capture—Concepts and Study Start-up.”

## **Importance of Metrics Standardization**

Because metrics may be shared between various functional groups or stakeholders, metrics should be based on standard definitions. The need for standardized definitions is amplified if metrics are used for comparisons across studies, projects, or organizations (e.g., benchmarking projects). Communication between various groups using a metric is also enhanced by the use of standard definitions.

For example, “time to database lock” is one of the most frequently cited metrics used in clinical studies. However, this metric may be defined differently within different organizations. Depending on an organization’s definition of this metric, completion of database lock may be considered to occur:

- when data are “frozen” and a sponsor accepts data transferred from their CRO (e.g., the database or transferred datasets),
- after a QA audit is accepted and it is deemed permissible to break blinding of the study,
- multiple times, depending upon SOPs and whether or not a company allows for database “unlocking” to make changes to the database after it was originally locked.

Likewise, the starting point for this metric may be defined by different organizations as any one or more of the following criteria:

- the last subject completes the last visit (LPLV),
- the last data from the last subject visit are recorded on a paper CRF or entered into an EDC system,
- the last CRF is received by the group performing data entry,
- the data cleaning activity is deemed completed (i.e., generation of last query in database).
- the last query or discrepancy is resolved.

Due to various interpretations of the metric “time to database lock,” all parties could potentially be working in different directions based on their

presumption of when database lock occurs and what activities take place at that point. Without a standard definition of this metric, the goal may not be identified or achieved in an efficient and effective fashion. To ensure clarity and efficiency, all functions affected by a metric should be involved in the definition of the metric and made aware of the interpretation of the metric that is to be followed.

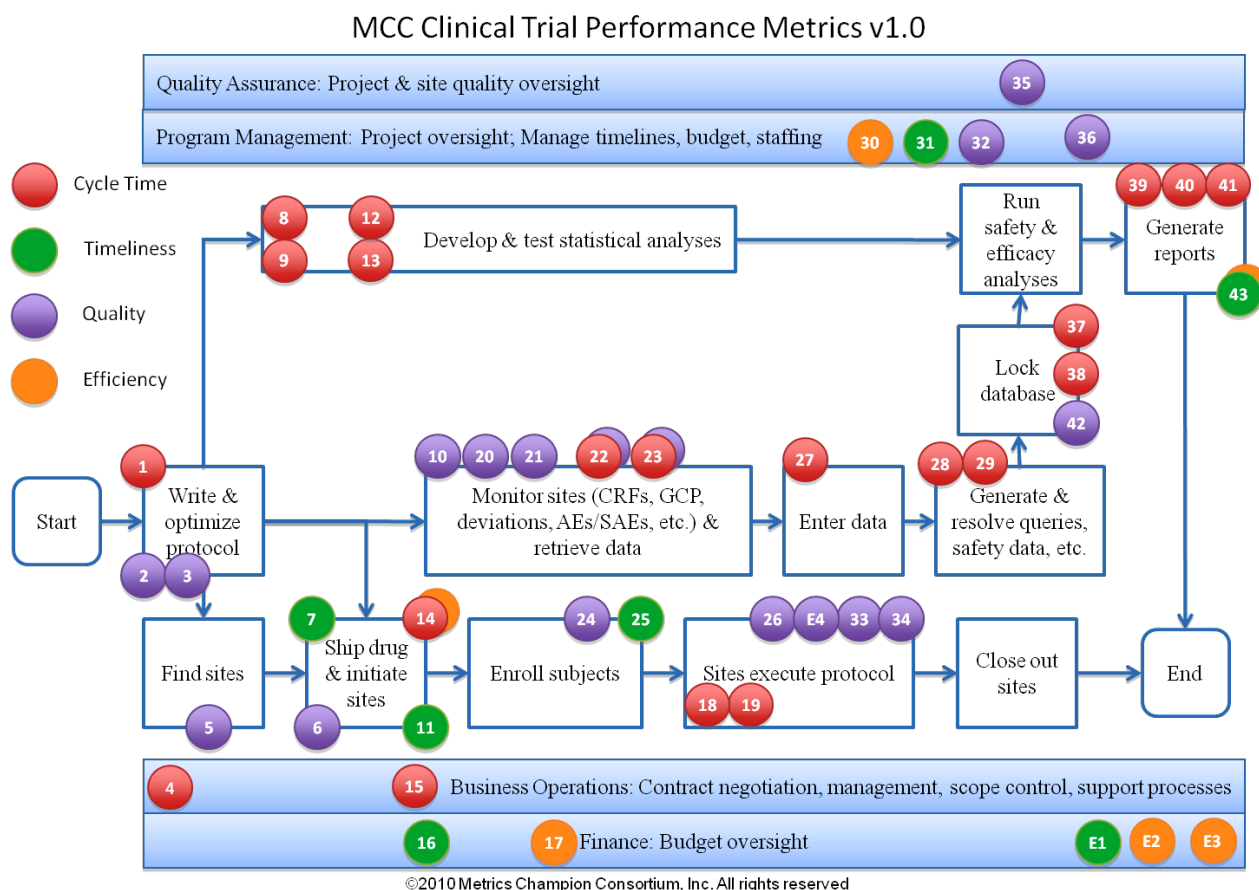
If the starting point for “time to database lock” is the date the last subject completes the last visit, the CRA or monitoring group should work with CDM to develop and agree upon definitions and the process used to achieve this milestone. As for the end point, if it is defined as the point that blinding of the study is broken, appropriate representatives (e.g. biostatistics, CDM and personnel responsible for randomization code storage) should work together to understand their respective roles in this process. The data management plan (or other applicable documentation) should be kept current to reflect any decisions that are made regarding metrics to be collected and their definitions.

Like other areas of clinical data management where standards are evolving, there is an initiative to develop industry-wide standards by the not-for-profit Metrics Champion Consortium (MCC).<sup>2</sup> Comprised of representatives from biotechnology, pharmaceutical, medical device and service provider organizations, the mission of MCC is to develop performance metrics within the biotechnology and pharmaceutical industry.

One of MCC’s initiatives focuses on clinical trial metrics where more than forty performance metrics have thus far (as of March 2011) been defined along with standard formulas and calculations used for reporting. Paired with standardized definitions and standard formulas for measuring each metric, all parties can stay informed of the criteria for measurement and the results being achieved not only within an individual study, but across studies that also use the identical metric definitions and formulas.

Figure one shows an example schematic of performance metrics within a clinical study and indicates when specific metrics may be used and the focus of that metric (for example, to evaluate quality or efficiency).

Figure 1. Example Schematic of Performance Metrics within a Clinical Study.



**Legend:**

MCC Metric Number	Definition
1	Number of calendar days from protocol synopsis to protocol approval
2	Number of versions prior to protocol approval
3	See Protocol Quality Score System
4	Contract execution timeliness (non functional outsourcing models)
5	See Site Selection Quality Score System
6	% country regulatory packets approved after first receipt
7	Timeliness of protocol approval to first site activated [country, region, study]
8 - EDC	Number of calendar days from final approved protocol to final approved eCRF
9 - paper	Number of calendar days from final approved protocol to final approved paper CRF
10	% monitoring plans completed prior to first site initiated
11	% planned sites activated
12 - EDC	Number of calendar days from eCRF sign-off to database "go live"
13 - paper	Number of calendar days from sign-off of final paper CRFs to database "go live"
14	Number of calendar days from Site Activation to FPFV (patient consented) [site, country, region, study level]
15	Number of calendar days from event threshold for change order (CO) generation to CO agreed and signed by both Sponsor and CRO.

16	% "On Time" payments of invoices
17	% actual contract value vs initial baseline contract value
18 – EDC	Calendar days from Patient Visit complete to eCRF page entered in EDC system
19 – paper	Calendar days from Patient Visit complete to CRF page entered in data management system
20	Monitoring Visit Frequency Compliance
21	Monitoring Visit Report Completion Compliance
22	Documented Monitoring Visit Report Review Compliance
23	Monitoring Follow-Up Letter Completion
24	% of sites meeting recruitment expectations (protocol specific) [Reported by tier level T0 – T4]
25	% subjects enrolled at point in time vs. target date
26	% enrolled subjects who remain in the study (did not voluntarily withdraw)
27 – paper	Calendar days from pages received and/or scanned to data entry complete.
28 – EDC	Calendar days from time query generated to query response on EDC system.
29 – paper	Calendar days from time query generated to query response updated on the DM system
30	% of drug not used versus planned amount (per patient per country)
31	% of drug kits available vs planned
32	Number of protocol amendments after protocol approved
33	Number enrolled subjects with protocol deviations per defined categories
34	% of active sites closed prior to study closeout
35	Number of site audit findings that are major and critical
36	% of critical issues escalated according to project plan
37 - EDC	Number of calendar days from last patient, last visit (LPLV) until database is locked by DM (EDC)
38 - paper	Number of calendar days from last patient, last visit (LPLV) until database is locked by DM (paper CRFs)
39	Number of calendar days from final database lock (DBL) to final TLGs/TLFs
40	Number of calendar days from final TLGs/TLFs to first draft clinical study report.
41	Number of calendar days from final DBL to first final approved clinical study report.
42 - paper	Final Database Error Rate
43	Number of calendar days from final TLGs delivered versus target date promised
<b>MCC Clinical Trial Performance Metrics version 1.0 – Exploratory Metrics</b>	
Exploratory Metric E1	Median number of calendar days from contractual milestone to invoice receipt
Exploratory Metric E2	Schedule Performance Index (SPI): <i>Original contract</i> planned amount of work completed versus work completed to determine if work is progressing as planned.
Exploratory Metric E3	Schedule Performance Index (SPI): <i>Adjusted contract</i> planned amount of work completed versus work completed to determine if work is progressing as planned.
Exploratory Metric E4	See Site Assessment Quality Score System



©2010 Metrics Champion Consortium, Inc. All rights reserved

## **Context and Attributes of Metrics**

The context in which a metric will be applied should be determined prior to reporting the metric. Each metric's data source(s), data extraction date, and

reporting window should be included with each report. Each metric should also be grouped according to its attributes, which can be described as characteristics of a metric that help stakeholders understand the underlying causes for performance variances.

Some attributes that may be used for grouping include:

- therapeutic area,
- indication,
- study phase,
- data collection mode (e.g., EDC, paper, imaging),
- study design,
- size or complexity factors (e.g., number of sites, number of subjects, number of procedures), or
- resourcing model (e.g., CRO, contractors, in-house staff, etc.).

Categorizing and summarizing metrics according to their attributes can result in more clear and concise metrics reporting, and minimize the potential for making invalid assessments and generalizations.

### ***Defining Time Points for Standardized Metrics Collection***

To provide maximum benefit, metrics reports should be available for review as soon as possible. Project and department managers frequently need to gather status information for an ongoing study, including information such as enrollment rates, the number of open queries, or the types of queries that occur most frequently on CRF data. The greatest opportunity to take corrective action occurs when information is timely. The earlier a problem is detected, the sooner it can be addressed. Although details may vary between organizations and studies, Table 1 presents some metrics commonly used during different periods of a study. Although the table groups these metrics by the five core criteria and by three study periods (startup, conduct and closeout), some of these metrics may be applicable to multiple criteria or time periods.



Table 1. Examples of Common Study Metrics

<b>Criterion</b>	<b>Study Startup</b>	<b>Study Conduct</b>	<b>Study Closeout</b>
Quantity	Number of expected subjects  Total number of data fields (may be quantified differently by different organizations)	Amount of data entered  Amount of data cleaned  Expected amount of entered data compared to data in database	Final number of subjects  Number of outstanding queries  Missing pages report
Cost	Total estimated resources (such as people, licenses, infrastructure, printing, etc.) needed for a study	Number of monitoring visits	Total study costs  Average cost per subject enrolled
Time	Projected overall study timeline  Time needed for protocol/CRF review and finalization  Final approved protocol to database activation	Time from subject visit to data available to CDM  Time from subject visit to data cleaned and locked	Time from first subject enrolled to last subject visit  Time from last subject visit to final database lock  Time from final database lock to clinical study report
Quality	Systems validation results	Number of queries and re-queries  Number of data transfer errors  Metrics generated from audit trail	Number of data errors per number of total data fields (error rate) (used in paper studies)  Number of protocol deviations
Performance	Number of programmed procedures that validate correctly	Comparison of data entry rates across sites  Time from subject visit to data entered  Average time for query resolution	Number of database unlocks to correct data errors  Number of protocol amendments

## **Action Plans: The Feedback Loop**

Ultimately, the desired outcome of using metrics is obtained through well-planned and executed processes that include interim assessments and feedback loops. An organization should carefully design the procedures that collect the metrics needed to assess whether a goal has been reached. However, the organization should also carefully design procedures describing the actions that may be taken based on the results of collected metrics.

Metrics reports are useful for both interim and final assessments of a project, therefore these reports should be run at agreed-upon times during and at the end of the project. Reports should summarize the metrics collected, and should include an assessment of results against goals or objectives. Metrics reports may also provide commentary about the results, which should include reasons for positive performance and plans for corrective action to improve performance.

Useful reports for the analysis of metrics include trend analyses, statistical techniques, summary tables, flagging of outliers, identifying unanticipated trends in the data, plots showing incoming data and query rates, and listings of values, such as changes from baseline values.<sup>3</sup> Ideally, metrics should be categorized according to their ability to assist in comparing a project's outcome to the outcomes of other projects inside or outside the organization.

## **Using Metrics to Improve Organizational Efficiency and Effectiveness**

Comparing metrics from different projects and studies can help improve the overall efficiency and effectiveness of an organization. If a particular process functioned more effectively and efficiently in a specific project, the organization can try to determine what factors made the process more efficient in that specific project and then try to apply those same factors to other projects. By using metrics to identify areas of strength or weakness within individual projects, an organization can apply lessons learned to projects in the future, thus improving the overall effectiveness and efficiency of the entire organization.

One of the means of ensuring visibility and transparency of metrics across all parties (sponsor, clinical research organization, and vendor) is by creating service level agreements (SLAs) and operational level agreements (OLAs) for

those metrics that form the key performance indicators. Routinely reviewing KPIs in governance meetings (strategic and operational) provides an indication of the health of the project and may identify areas needing corrective and preventive actions (CAPA).

### **Using Metrics to Improve Timeline Efficiencies**

Metrics can be used early in a study to identify areas where timeline efficiencies might be improved. For example, if a particular site is not entering data or resolving queries in a similar timeframe as other sites or within the expected timeframe, the root cause can be identified and, if warranted, corrective and preventive actions can be initiated, such as retraining relevant site staff. If particular milestones are not being reached as expected across an entire study, processes and data collection tools can be reevaluated to determine if adjustments could potentially improve timeline efficiencies.

### **Using Metrics to Improve Operational Efficiencies**

Frequently, operational efficiencies can also be improved by initiating corrective actions based on metrics reports. As with timeline efficiencies, identified operational inefficiencies at a particular site (e.g., delay with uploading data from ePRO) can often be improved by retraining relevant site staff. If metrics identify processes that are not working as efficiently as intended across an entire project or study, relevant processes and tools can be carefully examined to determine the most effective corrective actions needed to improve operational performance and efficiency.

## ***Metrics Documentation***

The data management plan (DMP) is a tool that can be used to document decisions about the use of metrics for a project (e.g., metrics definitions, the means of collecting metrics, the means of communicating metrics). However, some organizations may choose to document metrics separately from the DMP. Regardless of where they are documented, the metrics used for a project should be defined at the planning and initiation stages of the project.

All key metrics reports and other documents relevant across projects should be referenced in the project documentation, as well as all project assumptions and assertions for establishing particular metrics. If new terms are used or new

stakeholders or vendors are involved with a project, establishing and maintaining a project dictionary or glossary may be helpful.

### **Recommended Standard Operating Procedures**

- Definitions and Use of Performance Metrics
- Validation and Testing of Metrics Collection Tools
- Vendor Management

### **References**

1. Pitre A. *Definition of “Metric.”* PhRMA presentation; 2001. See also [www.amstat.org/sections/sbiop/br\\_spr99.pdf](http://www.amstat.org/sections/sbiop/br_spr99.pdf) (accessed March 9, 2011); page 7, 10. *Data Management Metrics*.
2. Metrics Champion Consortium. Clinical Trial Performance Metrics Initiative, <http://www.metricschampion.org/Pages/CRO.aspx> (accessed March 9, 2011).
3. Walton, M. *The Deming Management Method*. New York: Perigee Books; 1986.

### **Further Reading**

Terms used in this chapter may be found in the *Good Clinical Data Management Practices Glossary*.

E. Pena. Making Metrics Matter: The Changing Paradigm of R&D Metrics. *PharmaVoice*; 2005 (March):8–20.

Kush RD, Bleicher P, Kubick WR, Kush ST, Marks R, Raymond SA, Tardiff B. *eClinical Trials: Planning and Implementation*. Boston, MA: Thomson CenterWatch; 2003.

Meyer C. *Fast Cycle Time: How to Align Purpose, Strategy, and Structure for Speed*. New York: Free Press; 1993.

Tufte ER. *Envisioning Information*. Cheshire, CT: Graphics Press; 1995.

Tufte ER. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press; 1997.

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
October 2005	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
April 2011	Revised for content, style, grammar, and clarity. Name of chapter changed from “Metrics for Clinical Trials” to “Metrics in Clinical Data Management.”

**This page is intentionally blank.**

---

# Assuring Data Quality

October 2013

## *Abstract*

High quality clinical research data provide the basis for conclusions regarding the safety and efficacy of a medical treatment. This chapter discusses how the terminology and methodology for assuring quality, already well established in other industries, can be applied successfully to clinical research. General principles of quality systems and quality assurance in clinical data management are discussed. The key differences between quality assurance and quality control are presented and the roles of standardization, standard operating procedures, and auditing are reviewed.

## *Introduction*

Before discussing methods of assuring data quality, one must determine exactly what is meant by terms such as “quality,” “quality control” (QC) and “quality assurance” (QA). The American Society for Quality (ASQ) provides the following definitions for these terms.

- **Quality**—This is a subjective term for which each person or sector has its own definition. In technical usage, quality can have two meanings: 1. the characteristics of a product or service that bear on its ability to satisfy stated or implied needs; 2. a product or service free of deficiencies. According to some experts on quality, such as Joseph M. Juran, quality means “fitness for use,” and according to Philip B. Crosby, it means “conformance to requirements.”<sup>1-3</sup>
- **Quality control**—This term refers to the operational techniques and activities used to fulfill requirements for quality.<sup>3</sup>
- **Quality assurance**—This consists of all the planned and systematic activities implemented within the quality system that can be demonstrated

to provide confidence that a product or service will fulfill requirements for quality.<sup>3</sup>

In clinical data management, QA may be thought of as an overall management plan to ensure the integrity of data (the “system”), while QC may be thought of as a series of measurements used to assess the quality of the data (the “tools”). The terms QA and QC have been used in an imprecise manner in many industries including clinical research, and the ASQ and the American National Standards Institute (ANSI) both provide explanatory notes to that effect with their definitions of these terms.<sup>3,4</sup> The Institute of Medicine definition is often used for data quality within the context of clinical data management (CDM), and states quality data “are those that support the same conclusions as error free data.”<sup>5</sup>

A key aspect to remember about clinical research data quality is that it may be composed of numerous attributes.<sup>6</sup> For clinical research data, attributes of quality may include accuracy, consistency, timeliness, consumability, currency, completeness, relevance, granularity, unambiguity, precision and attribution.<sup>7</sup> Clinical research data quality may therefore refer to a dataset that accurately represents data points collected from subjects, has acceptable completeness, is defined sufficiently for use, is current, is attributable, and contains relevant data at the appropriate level of precision to answer the study’s primary hypotheses.

Quality assurance refers to all of the planned actions and systems implemented to impart confidence that a study will culminate with a quality dataset. Within this context, quality control refers to specific activities and techniques employed within the QA system to achieve the goal of finishing the study with a quality dataset. The most common approach to assuring quality is through a quality management system, which is the means by which an organization is controlled with respect to quality.<sup>8</sup>

Although the ultimate goal of CDM personnel is to complete a study with a quality dataset, proper principles and practices must be employed throughout the course of a study to ultimately ensure quality. If a study’s design, protocol or case report forms (CRFs) are of insufficient quality, the study is unlikely to accurately provide answers to its hypotheses. Lack of quality processes in any part of a clinical study can lead to results that are distorted, missing or inaccurate.



## Scope

This chapter emphasizes the infrastructures and practices that those managing clinical research data should use to ensure data quality. Although quality measurement methods are a necessary part of a plan to obtain quality data, a larger emphasis should be placed on error prevention, both in organizational infrastructure and early in the design stages of each protocol. For information about identifying and quantifying errors in clinical research data, see the *Good Clinical Data Management Practices* (GCDMP) chapter entitled “Measuring Data Quality.”

Many of the tasks described in this chapter may be joint responsibilities between different groups, just as many different groups may be involved in the implementation of various tasks. However, in all cases clinical data managers need to be conscious of whether or not these tasks have in fact been performed in a satisfactory manner.

## Minimum Standards

- Design and maintain data-handling processes according to the organization’s documented quality system.
- Attempt to collect only data that are essential for interpretation of study results and that are required by the protocol.
- Provide sufficient information in data-processing documentation to reproduce final analyses from source data.
- Assure data quality for all studies, whether submitted for regulatory review or not (e.g., marketing studies, observational studies or for publication-only studies).
- Ensure data quality is appropriate for study analyses according to parameters laid out in a statistical analysis plan, if one exists. Appropriate levels of data quality for analyses should always be determined by an experienced statistician.
- Use company-standardized data collection and handling processes.

## **Best Practices**

- Have an organizational quality policy that is strongly supported by upper management, understood by all staff, and supported by operational procedures.
- Create and maintain documentation of all roles and responsibilities involved in managing a clinical study.
- Use industry-standardized data collection and handling processes.
- Use well-documented processes for data collection and handling.
- Minimize the number of data-processing steps in order to minimize potential sources of error.
- Focus on error prevention with QA and focus on process monitoring with QC. The final product (database or software) of the clinical study should not be the focus of QA or QC.
- Ensure data quality audits assess compliance of procedures to regulations, compliance of practices to written documentation, conformance of data to source documentation, and conformance of data to written procedures.
- Apply data QC to each step of data management processes.
- Ensure all data management personnel are trained on and knowledgeable of the organization's quality policy.

## **Quality Systems**

A quality system encompasses the organizational structure, responsibilities, procedures, processes, and resources that are necessary to implement quality management.<sup>9</sup> This approach was standardized by the International Organization for Standardization (ISO) and is applicable across many industries, including clinical research. A quality system approach advocates an infrastructure that provides the flexibility to account for study differences in a controlled and consistent manner. Although not mandated for all clinical studies, a quality system approach has been adopted by the FDA in medical device regulations.<sup>10</sup>

Every study should establish an appropriate minimum level of quality, which should be determined through planned analyses specified in the protocol or statistical analysis plan. The assessment of data quality needs should address the study's purpose, characteristics and complexity.<sup>5</sup> A key concept of the quality system approach is that the structure, format, content, and method of presentation of documented procedures are contingent upon the needs of the organization.<sup>9</sup> Most organizations involved with clinical research already have some components of a quality system in place, for example, policies and procedures.

Within the context of CDM, a quality system should assure the following fundamentals:

- Written procedures and associated documentation should enable the clinical database to be reproduced from the site's source documentation.
- Written procedures must be followed.
- Data are consistently of sufficient quality to "support conclusions identical to those drawn from error free data."<sup>5</sup>

## **ISO Quality Systems**

The ISO provides the ISO 9000 series of standards to assist organizations with creating and maintaining quality systems.<sup>9</sup> The ISO quality management system describes a process-based approach in which organizations establish the infrastructure needed to control quality of their product sufficiently to meet customers' needs consistently.<sup>9</sup> To meet ISO quality management system infrastructure requirements, an "...approach to developing and implementing a quality management system consists of several steps including the following:

- a) determining the needs and expectations of customers and other interested parties;
- b) establishing the quality policy and quality objectives of the organization;
- c) determining the processes and responsibilities necessary to attain the quality objectives;

- d) determining and providing the resources necessary to attain the quality objectives;
- e) establishing methods to measure the effectiveness and efficiency of each process;
- f) applying these measures to determine the effectiveness and efficiency of each process;
- g) determining means of preventing nonconformities and eliminating their causes;
- h) establishing and applying a process for continual improvement of the quality management system.”<sup>9</sup>

Implementing a quality system starts with identifying processes that are required to produce a product. In CDM, these are processes for which most organizations already have standard operating procedures (SOPs). A quality system, however, goes beyond SOP documentation and includes confirmation that a methodology is effective, resources are available, and measurement and monitoring are sufficiently rigorous (i.e., a control cycle for those processes such as periodic process audits). The ISO standard provides specific documentation requirements as well as necessary roles and responsibilities.

### **Components of a CDM Quality System**

The components of a CDM quality system must take into consideration the practices and elements of a quality system infrastructure. The ISO quality system requirements can be translated into the following areas for CDM.

- *Defined processes* necessitate that all operations performed by CDM are identified and defined. The starting point is an inventory of processes for which the department or group is responsible. The quality system standard also requires specification of the sequence of processes, as well as interactions between processes. The quality system should be consistently applied to all departments of an organization, because CDM is but one component of the clinical research process—data are also collected at sites, verified by monitors and analyzed by statisticians. For departmental implementations of the standard, document the interface points of CDM

processes with processes from other departments. These interface points can be documented in SOPs for data management processes.

- *Position descriptions* list and describe the functions of specific jobs or titles. Position descriptions should accurately and thoroughly describe the requirements of a position, including responsibilities, tasks and education. Position descriptions serve as the basis for candidate selection, training, performance evaluations, and promotions. Each individual involved with a study should have a position description that accurately describes the work they regularly perform.
- *Training* is described in more depth in the “Training” chapter of the GCDMP. Both the ISO standard and FDA regulations require that individuals have documented training for their job tasks. For each work process in which an individual participates, training should be provided and documented. Organizations often create a training matrix listing each position and required training for each position. All job description tasks should be linked to SOPs and be adequately represented in the training matrix, although there is no regulatory requirement to provide these links.
- *Management oversight* is a good practice. Even if a quality system has documented work processes, job tasks, and training, factors such as comprehension, quality, judgment and consistency can vary. Many CDM tasks require review of an individual’s work, as well as an opportunity for the individual to receive constructive feedback. Although some review may be appropriately conducted by a peer, management oversight should also occur above the level of the individual. For example, departmental management should receive summary status reports of progress and QC activities. At an even higher level, management has oversight responsibility to assure the quality management system consistently produces acceptable quality.
- *Process control* refers to the capability of a process to consistently produce a particular result. Although process control can also be considered part of management oversight, it is important enough to be described separately. Management is responsible for designing and maintaining processes that produce consistent results. For CDM, consistent results may include acceptable database error rates, data timeliness, minimal errors in database programming, and meeting

milestone deadlines. Rather than establish separate measures and controls for each process, process control should be determined by global organizational goals. High-level assessment may be sufficient, but detailed measures on certain processes may be helpful to identify issues early.

A quality system approach is most powerful when employed by an entire organization, covering the entire clinical research process. Although a single department can achieve high performance in isolation, only local optimization will be achieved, which may not fully align with organizational goals.

## **Quality System Documentation**

### **Quality Policy**

An organization's quality policy is the highest level of a quality system. Specified by top management, the quality policy communicates and documents an organization's overall intentions and direction with respect to quality. The quality policy should detail various levels of the organization's quality system, such as management review procedures, the quality manual and the quality plan.<sup>8</sup>

An organization should have a written quality policy, and top-level management should demonstrate commitment to the quality policy by supporting the organization's infrastructure with adequate resources. Off-line QC activities, such as quality engineering, quality planning, and procedures applicable to each study, will be enhanced by this infrastructure and facilitate error prevention.

Although an organization-wide quality policy as the overarching directive is best practice, if a quality policy does not exist, data management should rely on department-specific documents such as the Data Management Plan, SOPs, study-specific procedures and a study's protocol to establish quality within the department. Because a quality product will never be achieved with only one department adhering to a quality system, it is important for data management to elevate the need for a corporate quality policy to upper management and to inform upper management of the working parameters that data management will apply in the absence of a corporate quality directive.

## **Quality Manual and Plans**

A quality manual is defined by ISO 9000 as a “document specifying the quality management system of an organization.”<sup>8</sup> An organization should have a written quality manual that defines the quality practices, resources, and activities relevant to the data-handling services of the organization. Most organizations already implement portions of a quality manual as SOPs, but a quality manual is much broader. A quality manual should describe not only processes, but also training, management oversight, positions, and process control.

Quality manuals and quality plans must be flexible enough to address differences in various studies. For highly standardized organizations, information that would otherwise be part of a study-specific or project-specific plan may be included in an organization’s quality system documentation (e.g., the quality manual, audit procedures, and SOPs). In these circumstances, the plan should reference these quality system documents and detail how the documents ensure data quality for each study. Quality plans may be designed to apply to one specific study or to all studies for which the organization takes full or partial responsibility. The organization’s quality plan or manual should also be subject to version and change control.

## **Role of SOPs**

As process definitions, SOPs are a large component of (and can be specified in) the quality manual. Organizations should have a documented process for creating, reviewing, and version control of SOPs. To easily identify time periods when an SOP should be used, effective dates should be assigned to each published version. Although they do not have to be archived with each study, SOPs should be archived according to documented organizational procedures and be available should a study be audited years after closing. Planned deviations from SOPs should receive the same level of review and approval as the SOPs from which they are deviating.

The level of standardization within an organization helps determine the level of detail that should be present in the organization’s SOPs. For example, an organization with standard CRF modules, database structure, and monitoring procedures may employ detailed SOPs and thus, require less study-specific documentation.

Each GCDMP chapter recommends a corresponding set of SOPs. For those needing to create SOPs, the Society for Clinical Data Management's (SCDM) European sister organization, the Association for Clinical Data Management (ACDM), has published *Guidelines for Writing Standard Operating Procedures*.<sup>11</sup>

## **Study-Specific Procedures**

A quality manual should account for the existence of study-specific documentation. Each study may have unique data-handling needs due to variations in sample size, visit schedule, type of data collected, amount of data collected, and method of data collection. Organizations should clearly document study-specific procedures to ensure the analysis database is reproducible from source documents. Study-specific procedures are also often known as data-handling plans, data management plans, data-handling protocols, and data quality management plans. Such documentation should provide supporting details to SOPs and may have a lower level of review and approval within the organization.

The ACDM has also published *ACDM Guidelines to Facilitate Production of a Data Handling Protocol* (DHP guidelines).<sup>12</sup> These guidelines provide an outline and list of items to be covered in an organization's study-specific procedures. Organizations may customize the content of the data-handling protocol, adjusting the level of detail to correspond to the level of detail present in their SOPs. The DHP guidelines are an excellent reference for defining and developing organizational study-specific procedures. Such references only provide a framework, however, and the content should be specific to the organization. For more information, see the GCDMP chapter entitled "Data Management Plan."

## **Creating a Quality System**

### **Structuring a CDM Quality System**

The structure of a quality system should be designed by organizational leadership to provide consistency between studies and departments. Careful consideration should be given to what processes should remain consistent across studies and departments. The organizational QA group (if one exists)



and organizational leadership will likely play an active role in establishing the appropriate level of consistency across studies and departments.

Once an organizational quality system has been designed, each department can then create and document department-specific components within the organizational structure. Although the organizational structure for a quality system is a top-down exercise and requires specialized knowledge, many of the departmental components (e.g., SOPs, training and process control) are best designed with participation from departmental staff. For example, CDM personnel will be able to suggest information that is consistent enough across studies to reside in SOPs, as opposed to information that is more suitable in study-specific documentation.

Although the level of departmental freedom to customize quality system components may vary, each department will likely have the five key components of a quality system—defined processes, position descriptions, training, management oversight, and process control.

### **Quality Assurance in the CDM Function**

Quality assurance is the set of activities that ensures procedures are in place and effective in producing a quality product. *ICH E6* defines quality assurance as, “All those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with GCP [Good Clinical Practice] and the applicable regulatory requirement(s).”<sup>13</sup> In clinical research, QA includes the administration and support of SOPs and documentation. In many cases, QA also assesses the compliance of policies, products, and work processes with regulatory standards.

An organization’s written procedures should describe the approach taken to assure data are reliable and processed correctly at each stage of data handling.<sup>13</sup> Specific tools and quantitative techniques are necessary to ensure study data meet required levels of quality at every point where data are manipulated. Process monitoring, auditing, sampling, and error rate calculation are essential processes for quantifying data quality and assessing the potential impact of data quality on a study’s conclusions. These tools and techniques should be included in the organization’s quality documentation.

## **Incorporating Risk-Based Assessment**

Because of the time and resources needed to obtain completely clean and error-free data, a risk-based approach to QA may be adopted. Most studies do not require error-free data, but rather, data of sufficient quality to support the same conclusions as error-free data. Random data-entry errors and data that fail established edit checks may have little or no effect on conclusions drawn from statistical analyses. A QA goal in a risk-based approach would be to identify and evaluate systemic patterns of errors. Systemic errors may be considered to be non-random, for example, errors introduced through a programming fault or site-specific errors resulting from misunderstanding the protocol or CRF completion instructions. If identified, these systemic errors are typically found late in a study's lifetime when corrective action is not as effective. Because of the number of data points in most studies, evaluating a systemic error on each data point may be an overwhelming task. A risk-based approach may be used to identify categories of data (e.g., adverse events, efficacy data, safety data) that have the highest risk levels for each study and then clean those data thoroughly.<sup>14</sup> Risk-based practices may also include identifying higher risk studies and more stringent procedures that apply to them.

## **Incorporating Standards**

The clinical research industry's interest in standardization has grown in recent years. Organizations such as ISO and the Clinical Data Interchange Standards Consortium (CDISC) have published standards to provide uniform terms and structures for data collection, data storage, data transfers, and regulatory submissions. Standardization has the potential to shorten timelines, reduce costs and increase data quality.

Clinical research processes associated with data collection and handling can be error-prone and complex, potentially involving many steps. An error rate is associated with every step where data are transcribed, transferred, or otherwise manipulated. Subsequent steps can increase or decrease that error rate. Standard data collection and handling processes can be designed to limit the number of manipulations and transfers, thus reducing the potential for errors.

Regardless of its level of complexity, a standard process will become more familiar to users. Sources of error become well known and are more easily

recognized and quantified, reducing unexpected errors and complications. Standardization also discourages the addition of unnecessary steps to a process. Using standard processes enables an organization to fully characterize the performance of processes and implement controlled and evaluated improvements. Successful standardization efforts can also allow the flexibility needed to address and document study-specific processes.

Opportunities for standardization may vary from organization to organization. For example, a large pharmaceutical company has more potential for standardization than does a contract research organization (CRO). For more information about standards used within clinical research, see the GCDMP chapter entitled “Data Management Standards in Clinical Research.”

## ***Maintaining a Quality System***

Once a quality system has been created, an organization’s leadership should encourage proactive maintenance of the quality system. Corporate policies often predefine the methods by which maintenance is performed. Whatever methodology is employed at a corporate level, it should not preclude employees from critiquing processes or proposing more effective and efficient practices.

### **CDM Quality Control**

*ICH E6* defines quality control as “the operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled.”<sup>13</sup> Data quality is the result of all of the planning, execution, and analysis of a clinical study. Each step in the clinical research process should be designed to ensure the necessary level of data quality is maintained throughout the study.

*ICH E6* section 5.1.3 states that every stage of data handling should have QC applied to ensure data are reliable and processed correctly.<sup>13</sup> A QC step is required for each process or step in which data are transcribed, transferred, updated or otherwise saved to a new medium. When data quality does not meet predefined acceptance criteria, appropriate corrective action should be initiated.

In clinical research, data quality is typically quantified through error rate calculations. To be useful for comparing the quality of different databases, error rates must use the same scale and precision (e.g., using errors per 10,000 fields consistently rather than some combination of errors per keystroke, errors per patient or errors per record). Error rates must also measure the same components of the process and use a standard method for counting errors and fields inspected. Ideally, all error rates would represent the same sources of error and count errors in the same manner. For more information about error rate calculation, see the GCDMP chapter entitled “Measuring Data Quality.”

Error prevention, detection and monitoring activities should be described in an organization’s written procedures and documented as evidence of ongoing QC. To maximize error prevention, QC activities should occur at the earliest feasible point in a process and should assess process control and provide quantitative measures of data quality.

Some examples of QC procedures include:

- Double data entry
- Programmatic data range and consistency checks
- Regular evaluation of error rates to assess process control
- Manager or peer review of CDM deliverables (listings review, queries issued, query closing, coding)

### **Ongoing Process Control in CDM**

Once a quality system is created and all processes are in place, personnel working within the quality system must adhere to the system for it to be effective. Management must provide oversight of process control for the quality system and ensure processes of the quality system are followed as intended, so that each function results in a quality product. For example, well-documented procedures do no good if over time, compliance decreases. Process control provided by CDM leadership helps ensure workflows and proper levels of quality are maintained.

Process control includes inspecting periodic samples of data, usually at regular intervals, and taking corrective action on the process when inspection results indicate a trend, an out-of-control process, or consistently poor quality.

Compared to the cascade effect of a design error, a process error only has an additive effect on the downstream data quality. However, each manipulation point that an incorrect data point passes through will have to be reworked to correct the error. A process that is operating in a state of control will not only meet the requirements of *ICH E6* section 5.1.3, but will also reduce reworking, data cleaning, and inspection costs.

## **Review and Revision**

Because organizations always experience change, a quality system must be able to accommodate changes. Once a quality system has been created, it should also be reviewed on a regular basis. The review may use a predetermined corporate methodology or be an ad hoc review of quality system components. Either way, if changes need to be made to the original quality system components, these changes must be reviewed and approved. Once changes have been made, all relevant personnel should be retrained on new quality system components to ensure proper implementation of the quality system.

## ***Auditing a Quality System***

The word “auditing” is described by the ASQ as a systematic and independent examination to determine whether quality activities and related results comply with planned arrangements, and whether these arrangements are implemented effectively and are suitable to achieve objectives.<sup>3</sup>

In a context more specific to clinical research, the word “audit” is defined by *ICH E6* as:

A systematic and independent examination of trial-related activities and documents to determine whether the evaluated trial-related activities were conducted, and the data were recorded, analyzed and accurately reported according to the protocol, sponsor’s standard operating procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s).<sup>13</sup>

To be qualified to audit CDM, one should be knowledgeable of auditing methodology, CDM functions, computer programming fundamentals, and

industry regulations. An auditor's training and experience should be sufficient to thoroughly and accurately assess compliance of CDM procedures with good clinical practice. Audits of CDM functions should be performed often enough to ensure CDM processes and QC procedures effectively produce reliable and reproducible data for analysis and regulatory review.

A comprehensive audit of CDM evaluates the entire CDM quality system. The following three levels should be examined in a CDM audit.

- Written CDM procedures should be compliant with regulatory requirements and should specify process steps and decision points required for handling and processing clinical data, including instructions for manual reviews, data-entry conventions, and data clarification procedures. Written procedures should be specific enough to enable the clinical database to be reproduced using source documentation. To determine the level of compliance with regulatory requirements, an auditor compares CDM procedures with current regulations.
- Documented compliance of the CDM organization or department to its written policy should exist, consisting of objective evidence that the written data-handling procedures were followed. This evidence can include a database audit trail, signed and dated checklists, signed data clarification forms from a site, or interviews with CDM personnel.
- Objective evidence should exist to indicate that CDM processes result in quantifiably high-quality, reliable clinical data for analysis and regulatory review. Several steps are required to obtain objective evidence that CDM processes produce reliable clinical data for analysis and regulatory review. The first step is quantifying the quality of clinical data, which is usually represented by an error rate. Additional objective evidence may include data demonstrating that an organization's data-handling process is operating in a state of control. Another important type of evidence is an assessment of the potential impact of the error rate on interpretations of data and conclusions that are ultimately derived from the data. This type of assessment may be carried out by departments outside of CDM, but the results provide CDM with information that may ultimately improve CDM processes.

## **Other Considerations for Quality Systems**

Different types of studies require different considerations in relation to QA.

### **Considerations for Electronic Data Capture (EDC)**

For studies using EDC systems, data are available very soon after initial data collection, including audit trails, electronic signatures and query information. Review of real-time “live” data allow errors to be identified earlier in the study, as well as enabling faster subsequent corrective actions. Studies using EDC also differ in regard to source document verification (SDV). Because in some studies the EDC system can be used to capture the original recording of data (the source), studies using EDC may have fewer source documents available for SDV than would be found with a paper-based study.

### **Considerations for Regulated vs. Nonregulated Studies**

Although regulated clinical studies undergo the additional scrutiny of regulatory authorities, data quality is critical in all clinical studies. The clinical protocol and analysis plans should drive the quality of any clinical study, whether regulated or not.

One of the primary differences between regulated and nonregulated studies is the level of risk associated with the study. Due to the differences in risk, the processes employed and the degree of QC may vary between the two. For example, nonregulated observational studies would not need as thorough and as frequent audits as a regulated study.

### **Considerations for External Data Sources**

Vendors supplying data to be included in clinical study databases and analyses should have quality systems in place. The recipient of the data must ascertain, usually through a vendor-qualification audit, if the vendor’s quality system is acceptable and will maintain the integrity of the clinical study databases.

A study protocol will determine what external data will be transferred into a clinical study database. This requires that CDM be aware that data is expected and communicate with the data provider to negotiate the details of data transfers. If laboratory data is being handled by a central lab, communication

will be on a one-to-one basis. If lab data is being handled by local labs, communication may be on a one-to-many basis and may be more complex.

The receipt of external data should be handled procedurally according to quality system components, SOPs, and study-specific requirements in the data management plan.

For more information concerning data quality from external data sources, please see the GCDMP chapters entitled “Laboratory Data Handling,” “External Data Transfers” and “Vendor Selection and Management.”

### ***Recommended Standard Operating Procedures***

- Development and Maintenance of Standard Operating Procedures
- Development of Planned Deviations from Standard Operating Procedures
- Development and Maintenance of Study-specific Procedures
- Quality Assurance Audits

### ***References***

4. Juran JM. *Juran's Quality Control Handbook*, 4th ed. New York: McGraw-Hill; 1988.
5. Crosby, PB. *Let's Talk Quality*, 1st ed. New York: McGraw-Hill; 1989.
6. American Society for Quality. *American Society for Quality Glossary*. American Society for Quality, Milwaukee; 2008.
7. American National Standards Institute. *Statistics—Vocabulary and Symbols—Statistical Quality Control*; 1993.
8. Davis JR, Nolan VP, Woodcock J, Estabrook RW, eds. *Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making: Workshop Report*. Washington, DC: National Academy Press; 1999.
9. Batini, C. and Scannapieco, M. *Data Quality Concepts, Methodologies and Techniques*. Springer-Verlag, Berlin; 2006.



10. Nahm M., *Data Quality in Clinical Research* [chapter] in *Clinical Research Informatics*. Richesson RL, Andrews JE, eds.; Springer. 2012.
11. American National Standards Institute. *ANSI/ISO/ASQ Q9000-2000: Quality Management Systems-Fundamentals and Vocabulary*. Englewood, CO: IHS, Inc.; 2005.
12. International Organization for Standardization (ISO). *ISO 9001:2008 Quality Management Systems – Requirements*; Geneva, Switzerland: International Organization for Standardization; 4th ed.; 15 Nov 2008.
13. *Code of Federal Regulation, Title 21, Parts 800-1299, and specifically 21 CFR 820 Quality System Regulations*. Washington, DC: US Government Printing Office.
14. Association for Clinical Data Management. *ACDM Standard Operating Procedures Working Party: Guidelines for Writing Standard Operating Procedures*, Version 2.0. Macclesfield, UK: Author; 1997.
15. Association for Clinical Data Management. *ACDM Data Handling Protocol Working Party: Data Handling Protocol Guidelines*, Version 2.0. Macclesfield, UK: Author; 1996.
16. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Geneva, Switzerland: Author; 1996.
17. Howard, K. *A risk based approach for assessing data quality*. Data Basics; Society for Clinical Data Management; Winter 2007; pp. 11-15.

### **Further Reading**

Terms used in this chapter may be found in the *Good Clinical Data Management Practices* Glossary.

Batini, C., Catarci, T. and Scannapieco, M. *A Survey of Data Quality Issues in Cooperative Information Systems*. In *Proceedings of the 23rd International Conference on Conceptual Modeling*, Shanghai, China; 2004.

Pipino, L., Lee, Y. and Wang, R. *Data Quality Assessment*. Communications of the ACM 45, 8; 2002.

Redman, T.C. *Data Quality for the Information Age*. Artech House, Boston, MA; 1996.

Tayi, G.K. and Ballou, D.P. *Examining Data Quality*. Communications of the ACM 41, 4; 1998.

Wand, Y. and Wang, R. *Anchoring Data Quality Dimensions in Ontological Foundations*. Communications of the ACM 39, 10; 1996.

Wang, R. and Strong, D. *Beyond Accuracy: What Data Quality Means to Data Consumers*. Journal of Management Information Systems 12, 30; 1996.

International Organization for Standardization (ISO). *Guidelines for Quality and/or Environmental Management Systems Auditing*; 2002.

US Food and Drug Administration. *Guidance for Industry: Medical Device Quality Systems Manual*. Rockville, MD: US Department of Health and Human Services; 1999.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2003	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
October 2013	Revised for content, style, grammar, and clarity. Added more explicit description of quality management system components important in clinical research data management.

---

# Measuring Data Quality

September 2008

## *Abstract*

Data collected during a clinical trial must have as few errors as possible to be able to support the findings or conclusions drawn from that trial. Moreover, proof of data quality is essential for meeting regulatory requirements. This chapter considers the challenges faced by clinical data management professionals in determining a dataset's level of quality, with an emphasis on the importance of calculating error rates. An algorithm for calculating error rates is presented in this chapter and is asserted to be the preferable method for determining the quality of data from a clinical trial.

## *Introduction*

This chapter concentrates on identifying, counting and interpreting errors in clinical trial data. Data quality measurement methods are very important and should be applied to clinical trial operations as part of an overall planned approach to achieving data quality. Although measuring data quality is important, it is equally if not more important to focus on preventing errors early in the protocol development and data handling process design stages. Error prevention will be addressed in the "Assuring Data Quality" chapter of the GCDMP.

Federal regulations and guidelines do not address minimum acceptable data quality levels for clinical trial data, therefore it is left up to each organization to set their own minimum acceptable quality level and methodology for determining that level. As a result, differences in methodology for determining data quality and estimated error rates are often not comparable between different trials, vendors, auditors or sponsors. It is important that data management professionals take a proactive role to set appropriate standards

for acceptable data quality levels, to utilize methods for quantifying data quality, and to implement practices to assure data quality.

## **Scope**

This chapter provides minimum standards, best practices, and methods for measuring data quality.

The Institute of Medicine (IOM) defines “quality data” as data that support conclusions and interpretations equivalent to those derived from error-free data<sup>1</sup>. To make the IOM definition of data quality operational, organizations must understand sources of errors, identify errors through inspections, use inspection results to measure data quality, and assess the impact of the data quality on conclusions drawn from the trial.

## **Minimum Standards**

- Use statistically appropriate inspection sample sizes for decision making.
- Document the method and frequency of data quality assessments in the study’s data management/quality plan.
- Perform at least one quality assessment of the study data prior to final lock.
- Document data quality findings and corrective actions, if needed.
- Determine acceptable error rates for primary and secondary safety and efficacy (also known as “critical”) variables.

## **Best Practices**

- Use quantitative methods to measure data quality.

NOTE: Quantitative methods for measuring data quality involve classifying the data, counting the data, and constructing statistical models to help explain database quality, database errors, and patterns of errors. Database errors, or “findings”, can be generalized to the entire data set, and direct comparisons can be made between the sample and the whole data population as long as valid sampling and significance techniques are used. Quantitative methods

help differentiate between data errors that might be pervasive in the data set and errors that are merely random occurrences.

- Compare trial data and processes in the beginning, middle, and end stages of the trial.
- Work with clinical operations to predefine criteria to trigger site comparisons based on monitoring reports.
- Perform quality control on 100% of key safety and efficacy (critical) variables.
- Monitor aggregate data by site to detect sites whose data differ significantly so that appropriate corrective actions can be taken.
- Perform quality control prior to release of data used for decision making.

### ***Other Best Practice Considerations***

- “When a long series of data processing steps occurs between the source document and the final summaries (as when the source document is transcribed to a subject’s chart, transcribed onto a case report form, entered into a database, and stored in data tables from which a narrative summary is produced)”<sup>2</sup> compare the final summaries directly against the source document, at least on a sample of cases.
- Streamline data collection and handling to limit the number of hand-offs and transfers.
- Perform a data quality impact analysis. Impact analysis in data quality is a methodical approach used to assess the impact of data errors or error patterns on the trial or project. Through impact analysis, potential risks or opportunities can be identified and analyzed. Impact analysis can provide key information to aid in decision making.
- Evaluate the results of the impact analysis and propose system and process changes.

- Perform the appropriate level of risk assessment to ensure data quality based on the type and purpose of the trial. For more on this, see the “Assuring Data Quality” chapter.

## **Data Errors**

A clinical research study is a complex project involving many processing steps. Each step where data are transcribed, transferred, or otherwise processed has an error potential associated with it.

A data error is defined as a data point that inaccurately represents a true value. There are many sources or causes of data errors, including but not limited to, incorrect transcription at a clinical site, incorrect data processing, unintended responses based on an ambiguous question, or collection of data outside a required time window.

Common errors in clinical trial data are compiled from several references and are shown in Table 1.<sup>2, 3, 4, 5</sup> Table 1 also suggests some detection methods data managers can employ to identify data errors.

### **Error Detection**

It is not practical, necessary, or efficient to design a quality check for every possible error, or to perform a 100% manual review of all data. There will always be errors that are not addressed by quality checks or reviews, and errors that slip through the quality check process undetected.

Programmatic checks (data validation and/or edit checks) should be applied consistently across trial data, and all errors that are identified in this manner may be corrected. At a minimum, these checks should target fields critical to the analysis where errors may have a greater impact on the outcome of the study. However, not all errors can be detected using these methods. For example, unreported adverse events may be difficult to identify using programmatic checks.

**Table 1. Common Sources of Error and Primary Detection Methods**

<i><b>SOURCES OF ERROR</b></i>	<i><b>DETECTION METHODS</b></i>				
	Programmatic Data Checks	Source Data Verification	Data Validation	Aggregate Statistics	CRF-to-Database Inspection
Subject completes questionnaire incorrectly or provides incorrect or incomplete answers to questions (lack of tool validation or bad form design)			<b>X</b>		
Subject does not follow trial conduct instructions		<b>X</b>			
Inadequate instructions given to the subject				<b>X</b>	
Site personnel trial conduct error (protocol violation)		<b>X</b>		<b>X</b>	
Data captured incorrectly on the source	<b>X</b>	<b>X</b>			
Site personnel transcription error	<b>X</b>	<b>X</b>	<b>X</b>		
Site equipment error				<b>X</b>	
Human error in reading equipment or print out or inter-rater-reliability		<b>X</b>			
Data entry error	<b>X</b>	<b>X</b>	<b>X</b>		<b>X</b>
Electronic data acquisition error (power glitch, back up that didn't run, lead not attached securely)			<b>X</b>		<b>X</b>
Data linked to the wrong subject		<b>X</b>	<b>X</b>		<b>X</b>
Database updated incorrectly from data clarification form or query					<b>X</b>
Missing data	<b>X</b>	<b>X</b>			
Outliers	<b>X</b>				
Data inconsistencies	<b>X</b>	<b>X</b>			
Programming error in user interface or database or data manipulations					<b>X</b>
Lost data		<b>X</b>	<b>X</b>		
Fraud		<b>X</b>		<b>X</b>	

Errors caused by fraud and protocol violations can be difficult to detect without the use of special programming and the use of aggregate statistics.<sup>3, 5, 6, 7, 8, 9</sup> Throughout a trial, aggregate statistics should be available to monitors to facilitate detection of misunderstandings, misconduct and fraud. Data management is the first point in many processes where the data are available

for viewing in aggregate across sites. It is at this earliest point that aggregate statistics should be provided to monitors and other study personnel to quickly identify sites that are behaving differently from the rest. Aggregate data reports may be designed to summarize the performance of individual centers in the areas of recruitment, extent of follow-up, compliance to treatment, completion of procedures, late visits, or data queries.<sup>2</sup>

Source data verification (SDV) may be used to identify errors that are difficult to catch with programmatic checks. For example, a clinical trial monitor at the investigator site performs SDV by comparing the medical record (a subject's chart) to the CRF. Any discrepancies between the two that are not explained by CRF completion instructions, the protocol, or other approved site conventions are counted as errors. In addition, if a study is using electronic data capture (EDC) methods, SDV may be the best way to check for data errors. The scope of SDV can be decided on a trial-by-trial basis and should be determined at the beginning of the trial.

### ***Inspection or Comparison of Data***

ICH E6 defines an inspection as “the act by a regulatory authority(ies) of conducting an official review of documents, facilities, records, and any other resources that are deemed by the authority(ies) to be related to the clinical trial, and that may be located at the site of the trial, at the sponsor's or CRO's facilities or both, or at other establishments deemed appropriate by the regulatory authority(ies).”<sup>10</sup>

The American Society for Quality (ASQ) defines inspection as “measuring, examining, testing, and gauging one or more characteristics of a product or service and comparing the results with specified requirements to determine whether conformity is achieved for each characteristic.”<sup>11</sup> Here the term inspection is used to indicate a scope narrower than a comparison, and is a process where measuring can be performed as a step in the work process with less independence than a comparison. For example, a CRF-to-database inspection may be performed by individuals in the same department or on the same project team as those who did the work, as long as they are not the individuals who performed the work being inspected. In contrast, a comparison is often performed by trained company or sponsor representatives. Many organizations require both ongoing inspections and/or more formal comparisons to assure high quality data for all their trials.



## Data Comparison

Errors can be detected by comparing two representations of data captured at different points in the data handling process. A CRF-to-database comparison is performed by comparing the CRF to the data stored in the database.

Depending on the needs of the study, the comparison may be performed on the clinical database immediately following data entry, or on the analysis-ready datasets at database lock. In either case, an error is defined as a discrepancy between the dataset and the CRF that is not explained by data handling conventions, site signed data clarification forms, or programming conventions defined in the trial analysis plan.

## Sample Size

The best practice for sample size selection is using a statistically appropriate sample size for each inspection. This assures that information obtained from the inspection is representative of the entire database and can be used in decision making. It is important that the data manager work with key study personnel to develop and document a sampling methodology. For studies having a large enough study population, one sample size algorithm commonly used by many organizations is the square root plus one ( $\sqrt{n} + 1$ ) of the total study population. Another approach used is having a sample size equal to ten percent (10%) of the total study population.

## Error Rates

Data quality can be quantified in two ways: (1) raw counts of numbers of errors, and (2) error rates. Calculating an error rate guards against misinterpretation of error counts, can facilitate comparison of data quality across database tables and trials, and therefore is the preferable method.

Caution should be used when interpreting raw error count data. These data can be misinterpreted if used to compare the quality of different database tables within the same database, or the data quality of two different trials.

The error rate is defined as the number of errors detected divided by the total number of fields inspected.

$$\text{Error Rate} = \frac{\text{Number of Errors Found}}{\text{Number of Fields Inspected}}$$

Error rates are sometimes expressed as the number of errors per 10,000 fields. Scaling the error counts in this way gives a distinct advantage over raw error counts. For example, say two database tables or datasets, DEMOG and VITALS were inspected for a sample of 20 subjects. There are 100 fields in the inspected sample of the DEMOG dataset and 400 fields in the VITALS dataset. There are 10 errors found in DEMOG and 20 errors found in VITALS. The error rate is 1000 errors per 10,000 fields in DEMOG and 500 errors per 10,000 fields in VITALS. The DEMOG panel error rate is twice the VITALS panel error rate even though half as many errors were detected in DEMOG as in VITALS. By presenting error counts as errors per 10,000 fields, the data quality can be compared across not only database panels or datasets, but also across trials. The error rate gives a common scale of measurement for data quality. This is why establishing error rate methodology is recommended as a minimum standard. Error rates should always be presented along with a description of how they were calculated. For the hypothetical DEMOG dataset used as an example, this may be presented as follows:

DEMOG error rate = 10,000(Number of Errors Found / Number of Fields Inspected)

DEMOG error rate = 10,000(10/100)

DEMOG error rate = 1000 errors per 10,000 fields

Include the mathematical calculation(s) and the final, calculated error rate(s) in a report that summarizes database quality.

This is just one example of how to express error rates, and error rates can also be expressed through other means, such as by a percentage or a p value.

### **Important Concepts About Error Rates**

- The error rate is only a part of data quality process evaluation. It is important to know if the errors are in critical or noncritical fields. If the high error rates are in noncritical fields, they may have little impact. In this case, an organization may determine that it is not worth the time and effort required to clean these data.
- Knowledge of the error rates can help you choose the process paths and technology that will yield the highest quality for your organization.

## **Acceptable Quality**

In the absence of industry-wide standards for acceptable error rates for CRF to database quality control, “quality data” means different things to different organizations. Popular definitions of an “acceptable quality level” include rates of 50 errors per 10,000 fields overall, and different standards for critical and noncritical variables. These standards range from 0 to 10 errors per 10,000 fields for critical variables, and 20 to 100 errors per 10,000 fields for noncritical variables.

There are many ways to quantify data quality and calculate an error rate. While the differences among the methods can be subtle, the differences among the results can be by a factor of two or more. For example, consider the hypothetical situation of two lab data vendors calculating error rates on the same database with three panels. The Protocol Number, Site Number, and Sponsor Number are default fields that do not require data entry, in all of three database panels. Vendor 1 includes each of these default fields in the field count as fields inspected, which results in a denominator of 100,000 fields inspected in the error rate calculation. Vendor 2 does not include them in the field count since they are default fields, for a denominator of 50,000 fields inspected. Both vendors do a data quality inspection and both vendors find 10 errors. When they calculate the error rates, Vendor 1 has an error rate half that of Vendor 2 only because they did not follow the same algorithm for field counts. This example illustrates how important it is for a common algorithm to be followed by all parties calculating error rates. It is imperative that the units in the numerator and denominator be the same. Some other examples of algorithm details that could skew results are:

- Should data errors involving derived fields be counted?
- Is an error in the month and year fields of a derived date one error or two?
- How should errors be counted in a header that are entered one time then electronically populated throughout the study pages?

Data managers, statisticians and other trial personnel must work together to define acceptable data quality levels for trials, to design data collection and processing so as to achieve the desired level of data quality, to measure data quality and monitor it throughout the trial, and to communicate data quality information to key stakeholders, including management and sponsors.

## **Documentation**

Documentation of data quality comparisons should include the number of errors found, the error rate, how the numerator and denominator were defined and the final error rate. Anyone reading the documentation should be able to recreate the sampling and error rate calculations and produce the exact same results. For information addressing how these processes may differ in studies using EDC, please refer to the chapters entitled "Electronic Data Capture—Study Conduct" and "Electronic Data Capture—Study Closeout."

## **Recommended Standard Operating Procedures**

- Measuring Data Quality
- Monitoring Data Quality
- Data Quality Acceptability Criterion

## **References**

1. Davis JR, Nolan VP, Woodcock J, Estabrook RW. eds. Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making. Workshop Report. *Roundtable on Research and Development of Drugs, Biologics, and Medical Devices, Division of Health Sciences Policy, Institute of Medicine*. Washington DC: National Academy Press; 1999.
2. Knatterud GL, Rockhold FW, George SL, et al. Guidelines for Quality Assurance in Multicenter Trials: A Position Paper. *Controlled Clinical Trials*. 1998;19:477-493.
3. Califf RM, Karnash SL, Woodlief LH. Developing Systems for Cost-Effective Auditing of Clinical Trials. *Controlled Clinical Trials*. 1997;18:651-660.
4. DeMets D. Distinctions between Fraud, Bias, Errors, Misunderstanding, and Incompetence. *Controlled Clinical Trials*. 1997;18:637-650.
5. Williams GW. Discussion: Errors, Misunderstandings, and Fraud. *Controlled Clinical Trials*. 1997;18:661-666.

6. Ventura D. Issues in Clinical Trial Management, Part II Corrupt Data: Sources and Solutions. *Research Nurse*. 1999; March/April:12-17.
7. Bauer P. Statistical Process Monitoring During Clinical Trials. In: *Abstracts from the DIA Thirty Fifth Annual Meeting; June 27 – July 1 1999, Baltimore, MD*: Drug Information Association; 1999
8. Knatterude GL. Quality Assurance in Multicenter Clinical Trials: An Overview. *Abstracts from the DIA Thirty Fifth Annual Meeting; June 27 – July 1 1999, Baltimore, MD*: Drug Information Association; 2000.
9. Nahm ML. Optimizing Data Quality Through Statistical Process Control. *Proceedings from the 15 Th. Annual US Clinical Data Management Symposium, March 2000, Philadelphia, PA*: Drug Information Association.
10. International Conference on Harmonisation. Guideline for Good Clinical Practice, E6: Good Clinical Practice; Consolidated Guideline. Geneva, Switzerland: Author; 1996.
11. American Society for Quality, “Glossary”, Available at: <http://www.asq.org/glossary/i.html>. Accessed on April 3, 2008.

### **Further Reading**

Calvert WS, Ma JM. *Concepts and Case studies in Data Management*. Cary, NC: SAS Institute Inc.; 1996.

Anderson P, Wortman B. *CQA Primer*. Indiana: Quality Council of Indiana; 1997.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
September 2008	Revised for content, style, grammar, and clarity.

---

# **Data Storage**

May 2007

## ***Abstract***

The storage of data that is collected during a clinical trial must be carefully planned. This chapter discusses issues that should be considered whether a study's data is stored electronically or on paper. Guidelines for securely storing data are provided, with an emphasis on preventing unauthorized access that could detract from the integrity of a study. Issues concerning passwords, access controls, electronic signatures (including 21 CFR 11), and audit trails are considered. Recommendations for the locking and archival of data at the conclusion of a study are detailed.

## ***Introduction***

The secure, efficient and accessible storage of clinical trial data is central to the success of clinical research. Whether collected using validated electronic tools or traditional paper forms, data are often transferred many times during the course of a clinical trial. These transfers occur between an organization's functional groups as well as between companies, contract research organizations (CROs), and regulatory agencies. Hence, the potential for data corruption and version control errors during data storage and transfer is significant and must be minimized to ensure consistency of results and data quality.

## ***Scope***

This chapter provides key considerations for the data storage and archival of clinical trial data.

## **Minimum Standards**

- During the conduct of a clinical trial, store all original data collected (e.g., case report forms and electronic laboratory data) in secured areas such as rooms or file cabinets with controlled access (e.g., locks). These original documents are to be considered part of the audit trail for tracing back to the source data and should be protected and controlled as rigorously as the electronic audit trail of database modifications or backup procedures.
- Document the procedures for granting access to database servers, establishing system controls, and assigning passwords. This process is especially important in a trial where the original data collection is done electronically and no paper backups will exist.

## **Best Practices**

- Store clinical data in such a way that backup copies can be easily and frequently made. For example, paper documents should be scanned and archived electronically.
- Use open formats for archival, storage, and transport of data (e.g., ASCII, SAS Transport, Portable Document Format (PDF), and the CDISC ODM Model) whenever possible. Adherence to this practice enables current and future access to the data by multiple systems or reviewers.

## **Physical Storage**

The physical security of original data sources (e.g., case report forms, electronic data files, and other original data documents) should be maintained carefully. Original paper and electronic documents should be warehoused in secure rooms or file cabinets with controlled access. Whenever possible, paper documents should be scanned soon after receipt and archived electronically so that they are included with the backup of other electronic files.

Database servers can be the primary warehouse of clinical data and should be physically secured, and appropriate standard operating procedures (SOPs) should exist to regulate access to them. Direct access to database servers should be restricted to individuals who are responsible for monitoring and backing up the system. All other access to database servers should be



controlled by logical security and should occur across a secure network protected by password access and appropriate system controls.

Special considerations must be given to the physical security of computers that are used for electronic data collection during a clinical trial. Whenever data are entered into a central database using a network connection, the physical security of the central server, most likely hosted by the sponsor or a vendor, is a primary consideration. If any data are stored locally at the study site before being sent to a central server (as is the case with a hybrid or “offline” system), the physical security of the system at the source of data entry is more critical. In either case, care must be taken to ensure the physical and logical security of computers that are used to store clinical data for any period of time.

Passwords and access-permission controls are vital to ensure that only authorized personnel may access study data. An administrator designated by company policy should assign permissions on an as-needed basis. Mechanisms should be implemented to capture and prevent unauthorized attempts to access the system. If such an attempt takes place, the administrator should be notified immediately. A procedure should be established describing best practices for the selection of passwords and the frequency that passwords should be changed. Passwords should never be shared among individuals or study teams. These operating procedures are designed to minimize the opportunity for data corruption via accidental or intentional manipulation of the electronic raw data.

To maintain compliance with Code of Federal Regulations Title 21 Part 11, trials that use electronic data collection and management will necessarily regard a user’s authentication (i.e., user name and password) as the user’s electronic signature. All data entry and modification should be captured and stored in an audit trail (user name, date and time stamps) that regards the electronic signature as evidence of the user’s authority to alter information in the clinical database.

## ***Electronic Storage and Warehousing***

In addition to access controls, database design and organization are important considerations for a thorough data storage system. Database validation and associated documentation is the cornerstone of a secure and reliable system.

All database validation and user acceptance documentation should be readily available to the study personnel to ensure that all database functions being performed on a study have been validated for quality and reliability.

Additionally, consideration should be given to ensure that project team access to clinical data is sufficient to expedite efficient and high-quality interim reporting, data-metrics evaluation, and safety-reporting requirements (see the Safety Data Management and Reporting chapter and the Measuring Data Quality chapter).

The need for thorough validation and trial database design is even more critical for trials utilizing electronic data collection. If a software malfunction or unintended loss of information occurs, data collected on paper CRFs can be re-entered. Because electronic data collection eliminates paper documents as an intermediary step between the study observations and the database, it is critical that database validation and reliability issues are resolved on the validated system prior to the entry of any study data. As electronic data entry moves closer to the point of patient care, electronic information more often will be the source data and, as such, require protection and secure warehousing.

## ***Data Archival***

Several archival procedures should be followed to ensure that the data are preserved in their raw format. Most importantly upon completion of a study, the database itself should be locked. That is, permissions to further modify the data should be removed from all except the most critical study personnel. A thorough study archive includes all of the following:

- Database design specifications: Documentation of the table definitions used to build the study database and file structure.
- Raw data: The final raw data files preserved within the study database format, and all original data transfers in their raw format.
- Audit trail: A complete electronic audit trail documenting all modifications to a database by date, time, and user identification.
- Final data: Preserved in a standard file format (e.g., ASCII, SAS transport) that can be easily accessed, reviewed, or migrated to another system.

- Original study documents: The original and/or scanned images of all original documents, which may be archived separately in a central records facility if necessary.
- Procedural variation documentation: Memos and relevant information about any variations from standard operating procedures or working practices that occurred during the trial.
- Database closure: Documentation of each database-lock and unlock, describing the time and conditions surrounding those procedures (for additional information, see the Database Closure chapter).
- Site copies of data: May be required for audit purposes; if needed, these copies should be locked, read-only datasets delivered on CD-ROM or a similar storage medium.

### ***Recommended Standard Operating Procedures***

In addition to SOPs, please also reference the chapters on Database Validation, Data Entry and Data Processing, and Database Closure. The following SOPs are recommended as a standard for controlling database storage and archival:

- Database validation.
- Database design.
- Database closure (including procedures for unlocking a locked database).
- Storage of original documents both during and after the trial.
- Forms management and e-data management—this procedure should cover shipping and handling of original and/or working copies of relevant study documents; If electronic documents and files are used, the SOP should specifically address file transfer specifications and storage for those files.
- Version/change control for revisions to software.
- Version/change control for revisions to hardware.
- Version/change control for revisions to data.

- Version/change control for revisions to documentation.
- Disaster recovery.
- System controls and security.

It is also advisable for investigational sites to maintain their own SOPs related to physical and logical computer security.

## References

1. US Food and Drug Administration. *Code of Federal Regulations*, Title 21, Volume 1, Part 11. Rockville, MD: US Department of Health and Human Services; 1998.
2. US Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Rockville, MD: US Department of Health and Human Services; April 1999.
3. US Food and Drug Administration. *Guidance for Industry: Electronic Submissions of Case Report Forms (CRFs), Case Report Tabulations (CRTs) and Data to the Center for Biologics Evaluation and Research*. Rockville, MD: US Department of Health and Human Services; 1998.

## Further Reading

Not applicable.

## Chapter Revision History

Publication Date	Comments
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.

---

# Data Entry Processes

October 2009

## *Abstract*

Established procedures for data receipt and entry are necessary for a study to successfully produce a clinical database of sufficient quality to support or refute study hypotheses. This chapter discusses considerations needed to reduce the likelihood of errors occurring during data entry processes and ensure consistency in a clinical database. These considerations cover topics including workflow components, data receipt and tracking, data entry, data review, data cleaning, and change control for case report forms, databases, and processes.

## *Introduction*

The purpose of data entry processes is to ensure data are reliable, complete, accurate, of high quality, and suitable for statistical analyses. Data entry processes encompass the efficient receipt, tracking, entering, cleaning, coding, reconciling and transferring of data. A number of factors should be considered when choosing a data entry process, such as the skill level and training of personnel, and the amount of time allocated for data entry. Clinical studies vary in study designs and operational plans, therefore the specific design and plan should address the unique requirements for a given study. Throughout a study, an effective plan will ensure each component or step of data entry processes provides an appropriate level of data quality. The International Conference on Harmonisation's *Guidance for Industry: E6 Good Clinical Practice* states, "Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly."<sup>1</sup>

With electronic data capture (EDC) systems, traditional data management roles may change. In most cases, site personnel conduct data entry and may have the capability to run edit checks and make data updates to resolve

discrepancies. When data managers are not able to make data edits, they may need to remotely guide site personnel through data cleaning processes. These processes may take the form of automated checks built into the computer system, or may operate through queries entered into the clinical data management system (CDMS). Whether a study is EDC- or paper-based, the functionality of the tools, the design of the study and the skill sets of staff should be carefully considered.

## **Scope**

This chapter focuses on data management functions of data entry processes, including data receipt, data tracking, data entry, change control, data review, data cleaning, and discrepancy identification, resolution, and reconciliation. The chapter is not intended to discuss audit or inspection processes in detail.

Although some of the specific topics addressed by this chapter may not be the direct responsibility of data management personnel, data managers must have an ongoing awareness of requirements and ensure these tasks have been completed in accordance with the principles and standards of their organization, regulatory bodies, and good clinical practice.

## **Minimum Standards**

- Utilize written procedures describing data flow, data entry, data processing, and required quality level. Ensure enough specificity to reproduce the analysis database from source documentation.
- Ensure employees are appropriately trained (including ICH-specified documentation of having been trained) on systems, procedures, guidelines, working practices, and appropriate references (e.g., materials such as medical dictionaries, medical abbreviations, etc.) and that these documents are current and available to employees throughout the course of the study.<sup>2</sup>
- Ensure all personnel involved with data entry or data management have the proper levels of access, grants and privileges.
- Maintain a list of individuals who are authorized to make data changes.<sup>3</sup>
- Apply quality control to each stage of data entry processes to ensure data are reliable and processed correctly.

## **Best Practices**

- Address the purpose, characteristics and complexity of each study in data entry training sessions, including, but not limited to a brief review of the protocol, scope of work, and identification of critical variables (usually privacy controlled subject identifiers, primary and secondary efficacy variables, and safety information).
- Verify in a test environment (before the data entry system is placed into active use) that entry fields function as planned (e.g., date fields only accept dates, drop-down lists contain appropriate values, skip patterns function properly). In some organizations, true test data pages may be entered for an entire case report form (CRF) packet, while other organizations may perform more focused testing. This is not to be considered a substitute for software validation or edit check testing.
- Provide comprehensive user training on CRF completion guidelines and data entry instructions.
- Provide sites, sponsors, vendors and study team members with timeline expectations for data receipt, data tracking, data entry, and turnaround times for data queries, file transfers and database deliverables.
- Establish thorough tracking mechanisms for the receipt of CRFs and other forms containing data to be entered. Tracking ensures control of the received records, identifies missing records and facilitates the archival of records at the end of the study.
- Establish database quality criteria, including a quality control plan that appropriately addresses primary efficacy and safety data.
- Monitor data entry functions while in active use to identify trends and ensure stable and desirable quality levels are consistent with study needs.
- Create and maintain comprehensive processes for change control.

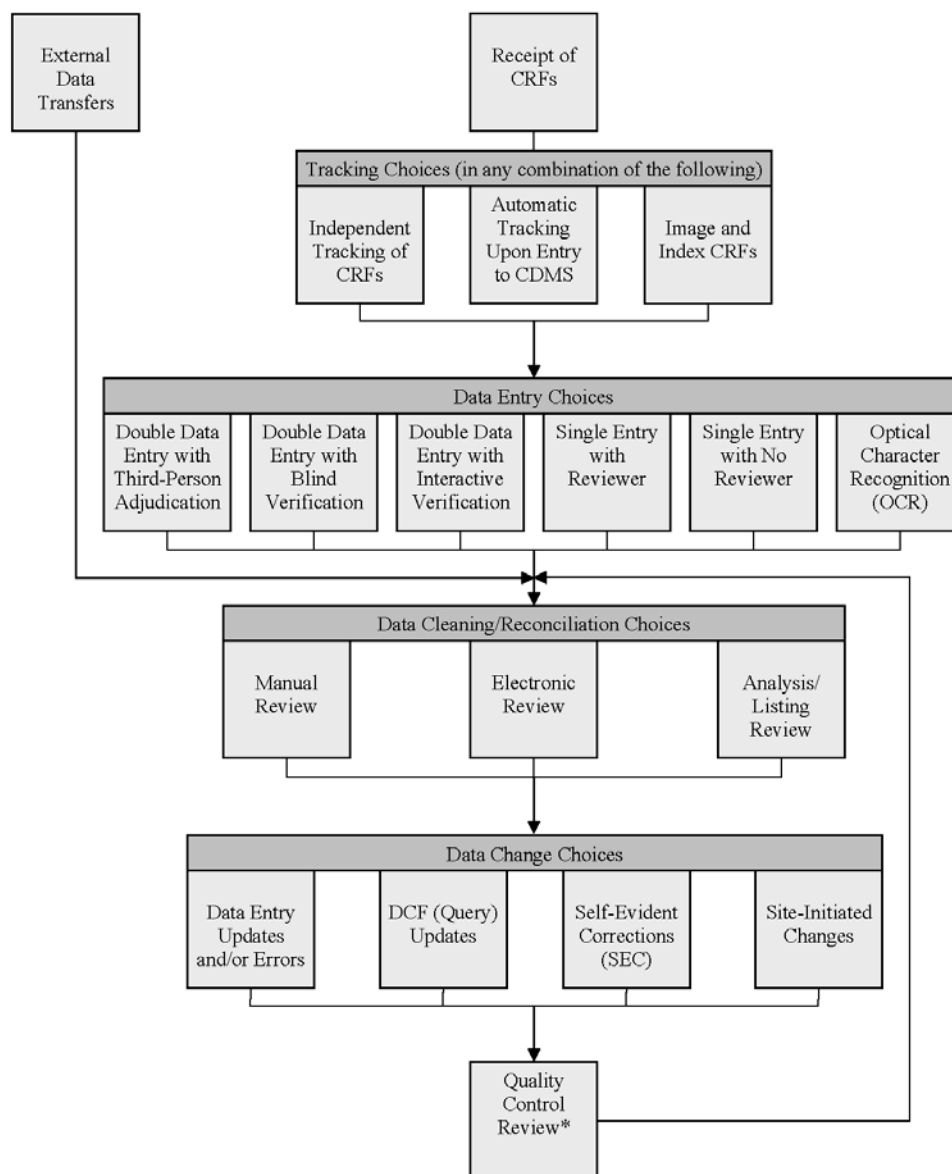
## **Workflow**

Although specific processes and steps may vary between studies and organizations, the flow of data should follow a logically prescribed path.

When data are received, it should first be tracked or logged, then entered, cleaned, and subjected to rigorous audit/inspection or quality control.

The general workflow of data entry processes for studies using paper CRFs is presented in Figure 1, as well as the choices available at each step. To determine which choices are made at each stage in the data workflow, every organization should have standard operating procedures (SOPs) and data processing conventions.

Figure 1. Paper CRF Data Processing Workflow



\* Quality control review may be performed at any stage, but should always be performed prior to database lock.



Workflow processes for EDC studies may vary according to the CDMS software used. For general principles of EDC workflow processes, see the GCDMP chapters entitled “Electronic Data Capture—Concepts and Study Start-up,” “Electronic Data Capture—Study Conduct,” and “Electronic Data Capture—Study Closeout.”

## ***Data Receipt***

Data receipt processes vary across the clinical research industry. Data may be received through fax transmissions, regular mail, express delivery companies with tracking ability, private couriers, hand delivery by monitors, Web entry, or transferred through other electronic means. Regardless of the data acquisition mechanism, the processes by which data are received, confirmed as received, and made available for data entry should be documented in the data management plan (DMP) in sufficient detail to ensure the origin of data is clear.

Standard operating procedures should be in place to ensure blinding of subject identifying information (e.g., name, address, or subject initials) submitted to the data center, unless collection of these data is authorized in the informed consent, protocol, and local regulations. Ensure a process is in place to quickly identify and report incidences of violations of data privacy conventions and laws. Missing CRF reports should be prepared for both paper-based and EDC studies to facilitate identifying forms that have not been received.

- Electronic data tracking—Computer-aided page checking can have higher integrity and efficiency than manual processes. Regardless of how data are received, procedures should facilitate timely, high-quality data processing. Expected visit date reports can be programmed into most reporting and tracking systems to follow a subject’s progression through a study and predict the last subject’s final visit dates.
- Paper CRF tracking—Tracking may occur on an individual CRF basis or per module. Ideally, all CRFs should be tracked, including mandatory, optional, and in some cases ancillary data. Data recorded on paper forms are recorded in one of the two following fashions, although details may vary between organizations. Some organizations may use a combination of independent or dependent logging with CRF imaging and indexing.

- Independent logging—This approach involves personnel manually registering that study data (not limited to CRFs) have been received. Data receipt may be recorded in the CDMS, although other tracking systems may be used as well.
- Dependent logging—This approach automatically records that a CRF has been received when data from the CRF are entered. This approach can eliminate an extensive and expensive manual process, replacing it with an electronic process in which tracking is a cost-free result of data entry. The trade-off is that any steps between receipt and entry may result in receipt dates that are not accurate. For reliable receipt dates, data should be entered when received, with little or no backlog of data to be entered.
- Tracking third-party data—Third-party data, such as laboratory data, may be received electronically or on paper forms. Documented procedures should be in place to track data from each external data provider within a study. For more information about processing third-party data, see the GCDMP chapter entitled “External Data Transfers.”
- Imaging and Indexing CRFs—To provide added security and flexibility for paper-based studies, CRFs may be imaged and stored electronically in addition to storing the paper forms. CRFs should be scanned using well-established formats, such as PDF (portable document format). The electronic files must be secured so they are only accessible to authorized and trained personnel. File-naming conventions should be strictly followed, and the repository of CRF image files should be indexed to allow specific files to be located quickly and accurately.

## **Data Entry**

Data entry processes should address data quality needs of the study. The following are some commonly used data entry strategies for studies using paper CRFs.

## **Methodologies**

- Double data entry (third-person adjudication)—Two people independently enter the same data and a third person independently resolves any discrepancies between first and second entry.
- Double data entry (blind verification)—Two people independently enter the same data, but remain unaware of what values the other entered. If the second entry operator enters a value that differs from the first value entered, the operator is warned that there is a discrepancy. After this warning, the second entry operator (who is responsible for verification) must carefully examine the form and determine the appropriate entry before saving. With this data entry strategy, the second entry will overwrite the prior value if it differs.
- Double data entry (interactive verification)—Two people independently enter the same data and the second entry operator resolves discrepancies between first and second entry while being aware of the values entered by the first entry operator.
- Single data entry with a review—One person enters the data and a second person reviews the data entered against the source data.
- Single data entry with no review—Although not recommended, situations may occur where one person enters data and the data are not subsequently reviewed.
- Optical character recognition (OCR)—Software packages are used to recognize characters from paper forms or faxed images and these data are placed directly into the database. Data obtained through OCR should always be reviewed for accuracy.

## **General Considerations**

Although specific data entry processes are not mandated by regulatory bodies or suggested by FDA and ICH guidance documents, a data handling document would most likely be a desired document in an audit or inspection. Having a set of standard data entry conventions for entry is encouraged to ensure consistency in the entry of data throughout the study. Data entry processes should be adapted according to the needed quality level for each data field.

Double data entry is typically used when frequent random keystroke errors may occur or if random errors would be likely to significantly impact analyses. However, a single-entry process with good manual review may be optimal in some circumstances, such as with free text fields.

Sites should have clear guidelines regarding timing expectations between a subject's visit and data being entered into an EDC system or recorded onto a paper CRF and forwarded to data management. The data management team is often responsible for producing reports that monitor compliance with established data entry timelines.

Although some clinical data management systems are capable of storing automatic default values, which are those written to the database with no action required by the entry operator (most frequently, but not limited to, subject identifiers, site numbers, and visit identifiers), this type of functionality should be used sparingly to reduce the likelihood of unexpected values being overlooked by data entry personnel.<sup>4</sup> In contrast, values that are derived, converted, calculated, or hard-coded based on the value of an entered field do not constitute automatic default values and are acceptable processes. Some organizations may perform these calculations outside the database, typically by those performing statistical analyses.

When applicable, system parameters should be set to allow an entry operator to exit the entry screen without saving the data that has been entered, as opposed to the system automatically saving entered data upon exiting. In this type of system, there should always be a prompt reminding the operator that data has not been saved. This approach enables data entry personnel to correct, upon discovery, situations where data may have been erroneously entered. Requiring a conscious decision to save data can also contribute to a higher level of data integrity. If the system does not allow for this data correction technique, a documented method to correct erroneously keyed information should exist.

Entry screens should be designed to minimize data entry errors. For paper studies, data entry screens should follow the pages of the CRFs, and may even be designed to appear identical to the paper CRFs. Some strategies for minimizing entry errors include displaying coded values and providing entry conventions (on entry screens or as a separate paper document), labeling entry fields clearly, and ensuring entry screens provide sufficient space to enter and view expected data.

## **Considerations for EDC**

For studies using EDC, sites should be contacted if they are falling behind in data entry. Although sites are typically entering and cleaning data, data management actions are still needed to help ensure data are entered and processed properly. These data management actions can include training site personnel on EDC system use, measuring site progress on data entry and cleaning, working through forms and data discrepancies with sites, data review, assessing aggregate data to identify subjects with outlying data, identifying data trends, verifying any and all coding, conducting data transfers and performing reconciliation.

Regardless of where data are entered, data entry personnel should be trained on the specific EDC system utilized in a study, as well as being taught the protocol and key data issues they might encounter. After data are entered, monitors verify data using source documents. In some systems, check boxes or particular fields on the entry screen are used by monitors to indicate which fields and visits were verified. In other systems, electronic forms may “graduate” through stages of, for example, data entry, monitored (or source document verified), and locked. In many systems, source document verification is negated if data are changed on the page. In such a case, source document verification must be repeated.

EDC systems may include user interface elements such as radio buttons and pick lists, and may allow fields to only accept specific variable types, such as only allowing numeric variables where appropriate. These systems may also be designed to allow numeric values to be checked against predetermined ranges upon entry. EDC systems can be designed to have dependencies for fields that should only have data when other criteria are met. An example of this design would be asking if a subject is of childbearing potential only if female gender had been selected.

The growing use of EDC systems has also had an impact on the training and desired skills for data entry personnel. In a traditional data entry method such as double data entry of paper CRFs, the skill emphasis is on the number of keystrokes made and the training emphasis is on the specific data entry system utilized. With EDC systems utilizing single entry, an overall understanding of the study becomes much more important in avoiding data entry errors. While

performing data entry in an EDC system, site personnel may need to check for online queries and recognize discrepancies as they enter data.

## **Data Entry Guidelines**

Whether using paper CRFs or an EDC system, detailed data entry guidelines should be provided to all data entry personnel. All data entry personnel should also provide written documentation that they have received and understood these guidelines. Data entry guidelines may be part of a broader user manual, particularly for studies using EDC systems. Both data entry guidelines and user manuals may take the form of paper documents or an online manual.

The following topics should be considered for inclusion in data entry guidelines or user manuals.

- Contact information of individuals available to troubleshoot computer problems and the hours such help is available
- Instructions or conventions describing how to enter data, delete data, and respond to queries
- Instructions or conventions describing how to enter data for single and multiple record panels if there is a difference in the system
- Reminders to users that a date/time stamp and a user name are recorded as part of the audit trail for every record. The audit trail may or may not be visible, depending on the computer system. Even if it is not visible during data entry, the audit trail must be readable by inspectors and auditors.
- Information on computer system security
- Instructions for proper computer shutdown procedures to prevent loss of data
- Instructions for data entry personnel explaining appropriate actions when edit checks trigger or reconciliation windows for double data entry systems appear

## **Data Review**

### **Data Cleaning**

Data cleaning refers to a collection of activities used to assure the completeness, validity and accuracy of data. Data cleaning activities may include manual reviews of data; computer checks that identify inaccurate or invalid data using ranges, missing data, protocol violations and consistency checks; or aggregate descriptive statistics that reveal unusual patterns in data. Early in a study, data should be reviewed from several subjects at each site to help detect problems with data entry screens not functioning as expected or a site's lack of compliance or understanding of the protocol.

The following list describes activities that may be included in data cleaning.<sup>4</sup>

- Verify raw data were accurately entered into a computer-readable file.
- Confirm code lists contain only valid values.
- Confirm numeric values are within predetermined ranges.
- Identify and eliminate duplicate data entries.
- Determine if there are missing values where complete data are required.
- Check the uniqueness of certain values, such as subject identification numbers or codes.
- Search for invalid date values and invalid date sequences.
- Verify that complex multifile (or cross-panel) rules have been followed. For example, if an adverse event of a particular type occurs, other data might be expected, such as concomitant medications or procedures.
- Check for any investigator comments entered on the CRF that could explain data anomalies.
- Reconcile all expected CRFs received with those that have been entered.
- Confirm inclusion of guidelines detailing reconciliation of adverse events, serious adverse events, lab data, or any additional third-party data.

- Check consistency of data across CRFs.
- Confirm that data are logical, even when outside expected parameters.

Range checks should be designed to identify statistical outliers, which are values that are physiologically impossible or outside normal variations of the population under study. Consistency checks should be designed to identify potential data errors (e.g., checking the sequential order of dates, corresponding events, and missing data noted to exist elsewhere). Checks designed to identify protocol violations should be closely monitored to allow timely action to be taken. A site should be monitored and investigated when aggregate statistics or other checks indicate substantial differences from other sites. Although manual review for data cleaning and validation is sufficient in some cases, programmatic validation provides high consistency and lower error rates.

Primary and other endpoints, key safety parameters and fields that uniquely identify subject data within the clinical database should be validated sufficiently to assure data are possible, complete, and reasonable. Data cleaning and validation procedures should not suggest bias or lead responses, because leading questions or forced responses can bias study results.

### **Data Cleaning Considerations for EDC**

Many of the data cleaning activities in the preceding list can be automated within a well-designed EDC system and may not require any post-entry effort. In an EDC environment, the site typically performs much of the data cleaning at the point of entry. The site is in control of the data and must either make the data edit or clarify the reason the data are acceptable. For a comparison of data cleaning processes between studies using paper CRFs or EDC systems, see Table 1.

The number of characters an EDC system will allow in a query is important for the data manager to know. Some data managers may be accustomed to paper queries with unlimited space, but most EDC systems require a scroll bar to view lengthy queries. This increases the importance of writing succinct queries to instruct the site to correct data or explain the reason for a discrepant or “abnormal” data value.



In an EDC system, built-in checks may initiate either at the time of data entry or when edit checks are run on batches of data. Additional edit checks may also be run and reviewed prior to issuing queries.

**Table 1. Data Cleaning Distinctions Between Paper and EDC**

<b>Data Cleaning Activity</b>	<b>Paper-based</b>	<b>EDC</b>
Discrepancies, Flags or Notes	<p>After entry and review are complete, flags or notes may be generated outside the database and submitted on individual data clarification forms (DCFs).</p> <p>Flags or notes may be compiled during entry and review, and subsequently addressed after data entry is completed.</p> <p>In some instances, items may also be flagged or noted during monitoring.</p>	<p>For data entry systems with no additional functions, flags or notes are identified by monitors and treated similarly to paper.</p> <p>Some systems show flags or notes on the screen in real time, allowing sites to address flags or notes sooner.</p> <p>Some systems close flags or notes automatically as values are updated, while others may require manual closing by monitors or data management personnel.</p>
Listings	<p>Cleaning listings differs from cleaning discrepancies, flags and notes in that cleaning may not occur as often due to a higher level of review by monitors, coders, statisticians, lab reconcilers, or safety managers.</p> <p>Listing reports may be sent to sites periodically to point out missing pages or overdue visits.</p>	<p>Some systems allow cross-page checks, but these may be limited in scope as many programmatic text checks must be manually reviewed.</p> <p>When posting responses or feedback, some systems may update or populate items right away, but others may be delayed due to system uploads occurring at predetermined intervals.</p>

To ease review and possible correction by site personnel, data managers should understand the EDC system and how data checks are attached to data fields. When checks are not issued against the correct panel, sites may be confused and not take appropriate actions. If a data check is to initiate automatically, it should check each data field only once. To prevent

duplication of effort, data management personnel should review previously issued data checks. Because sites must respond to data queries prior to any in-house review, it is critical that checks be properly tested prior to deployment. Deploying inadequately tested checks may result in unnecessary work for the sites and data management team.

Because sites may change data for various reasons, some users of EDC systems may not realize data that is clean today may not be clean tomorrow. These data changes may not be the result of data queries but rather a review of source data. Some systems are capable of locking data once it is clean, however a mechanism should allow the lock to be reversed for data changes if the site finds discrepancies that must be corrected.

### **Documenting Data Changes**

Data may be changed as a result of data cleaning procedures, in which case the site and data center or sponsor must retain a record of all such data changes. Data changes should be recorded and documented by a fax or original site signature acknowledging the new data. This documentation is usually accomplished using a query or data clarification form (DCF). In these cases, the site is expected to keep a record of the change within their study records.

In an EDC environment, site personnel usually make any necessary changes to the data. If nonsite personnel make data changes, a clearly defined SOP should document circumstances in which data can be changed, and a record of any data change should be provided to the site. All documentation of data changes is considered to be essential study documentation and is subject to audit or inspection. For comparison of differences in data-change documentation between paper-based studies and studies using EDC, see Table 2.

Data cleaning conventions may, under some circumstances, specify data that can be modified without a site's acknowledgement. These are known as self-evident corrections (SEC), and examples include appropriately qualified personnel correcting obvious spelling errors, converting values when units are provided, or providing missing identifiers when the true values are obvious. Because the site must have a record of all data changes, the site should receive and maintain a copy of each version of such data conventions.

Although strongly discouraged, situations do occasionally arise where telephone conversations with the site are utilized to authorize data changes. If this does occur, these changes should be clearly documented both by the site representative authorizing the change and by the data center representative talking with the site. In this way, a record of the conversation and authorization exists at both locations. In any case, any data change authorizations must be documented in writing and included in the study's documentation for audit or inspection purposes.

**Table 2. Data-change Documentation Distinctions Between Paper and EDC**

<b>Data Change Type</b>	<b>Paper-based</b>	<b>EDC</b>
Entry changes or errors	System or process changes should be reflected in data entry work instructions.  Database changes must be reflected in the audit trail.  When authorized changes are submitted by e-mail or phone, a hard copy should be created for patient folders both at the site and with data management.	When changing data, the EDC system should prompt the user to enter a reason for the data change. The reason provided will then be recorded in the database's electronic audit trail.  For non-site personnel, data entry work instructions or conventions will be used for documentation.
Data Clarification Form (DCF) updates	A hard copy of system-generated DCF submittals, once approved by authorized site personnel, should be kept with corresponding CRF pages at site(s) and with data management.	Rather than using paper DCFs, queries are generated and answered through the EDC system, which should include a comprehensive history of all queries recorded in the electronic audit trail.
Self-evident corrections	Self-evident corrections should be documented in study-specific conventions and data entry work instructions.	Self-evident corrections should be noted in the electronic audit trail.
Site-initiated changes	Site-initiated changes should be documented through manual DCF submitted with or without an updated CRF page.	Site-initiated changes should be noted in the electronic audit trail as new information provided by the site.

An audit trail is triggered by the initial data entry, and any changes to the entry are captured and should include the user name, the date and time of the change, the reason for the change, and the previous and current value. Recorded changes must not obscure previously recorded information.<sup>5</sup> To obtain consistent, accurate reasons for changes, some EDC systems offer a list of reasons for data changes as well as an option for free text. Since these reasons may vary, there should not be a default entry. Once a change has been committed and recorded in the audit trail, the reason cannot be edited.

A site's principal investigator should approve and sign off on data collected from that site prior to the data being finalized. This sign-off by the principal investigator must occur in both paper- and EDC-based studies. Any data changes that occur after the investigator signs must be re-signed by the investigator prior to study closeout.

## ***Change Control***

Protocol amendments are a fact of life in clinical studies. Changes to the protocol may be made when new information becomes available or is requested by the sponsor or regulatory agencies. While not all protocol amendments require CRF changes, procedures should be in place to handle these situations. IRB approval of protocol amendments must be received prior to deployment of new or changed CRFs. With paper-based studies, CRF changes may take a few weeks to be received by sites, by which time the sites may have received IRB approval for the protocol changes. However, with EDC systems CRF changes can be made remotely and implemented immediately upon IRB approval of the protocol changes.

## **Process Change Control**

All process changes initiated from a protocol amendment must be requested, reviewed, validated, approved and incorporated by following the organization's SOPs. If a process change involves modification of the clinical database, strict change control processes should be used to ensure preservation of clinical database integrity. Documentation of all process changes should always be stored and available for the entire project team.

At minimum, a process change should include identification and acknowledgement of the change, communication of the change to all

stakeholders, and a detailed request outlining any necessary modifications. The process change should be reviewed and approved by key stakeholders prior to implementation of the change.

Any process changes that involve investigative sites should come with clear communication and associated training (if training is deemed necessary). The clinical monitoring team should also be involved in process changes involving investigative sites. Process changes not involving site or monitoring staff should also include proper documentation, communication and training. Process changes should not be implemented until approval is received from all stakeholders and the change is thoroughly tested and validated. In some cases, changes may also require IRB approval.

### **Database or CRF Change Control**

If an approved change is made to an existing CRF or a new CRF is created as a result of a protocol amendment, data management is responsible for checking the consequences on the CRF completion rules and data entry guidelines, and if necessary, is responsible for modifying any existing database tables and creating any new database tables. Documentation of all changes and necessary validation testing is also the responsibility of appropriate data management personnel. As with process change controls, any changes should be communicated to the investigative sites in a timely manner. If CRF completion guidelines or data entry guidelines change as a result, ensure all changes are reflected and disseminated to appropriate personnel.

### **Change Control for External Data**

External data can originate from different sources, and are usually provided by previously selected vendors. External data include any data that are received as an electronic file rather than through paper- or EDC-based data entry. Any changes to external data should be corrected or updated at the source if possible. The vendor and data management should establish specifications and procedures at study start-up to describe how data changes will be communicated throughout the study.

Changes may be communicated between the site and vendor, site and data management, or vendor and data management. Changes communicated to sites are typically managed through DCFs or queries. Changes communicated

between the vendor and data management or the site and the vendor are usually communicated through a standardized process outlined in the DMP. Regardless of the methodology employed, any requested data changes must be tracked and documented.

### ***Recommended Standard Operating Procedures***

- Data Management Plan
- Data Validation Design and Testing
- Data Receipt
- Data Security and Storage
- Data Entry
- Data Review
- External Data Transfers
- Discrepancy Management
- Quality Control
- Database Lock Procedures
- CRF Archival

### ***References***

1. International Conference on Harmonisation. ICH Expert Working Group: *Guideline for Good Clinical Practice, E6, Section 5.1.3.* Geneva, Switzerland: Author; 1996.
2. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6.* Geneva, Switzerland: Author; 1996.
3. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6, Section 5.5.3.e.* Geneva, Switzerland: Author; 1996.

4. Cody, RP. *Cody's Data Cleaning Techniques using SAS Software*. Cary, NC: SAS Institute; 1999.
5. *Code of Federal Regulations, Title 21, Part 11, Electronic Records; Electronic Signatures*. Washington, DC: US Government Printing Office; 1997.

### **Further Reading**

N/A

### **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2000	Initial publication.
January 2002	Content revised.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
October 2009	Revised for content, style, grammar, and clarity.

**This page is intentionally blank.**



---

# Medical Coding Dictionary Management & Maintenance

May 2009

## *Abstract*

The use of medical coding dictionaries for medical terms data such as adverse events, medical history, medications, and treatments/procedures are valuable from the standpoint of minimizing variability in the way data are reported and analyzed. This chapter discusses the importance of medical coding dictionaries in streamlining and improving the quality of medical terms data obtained during collection and coding. Furthermore, reconciliation of medical terms data between a safety database and a clinical database is improved with the use of medical coding dictionaries during a clinical study. Issues that can affect conversion of reported terms to dictionary terms are considered, including autoencoders, the use of coded terms, and dictionary and software change control and versioning. Due to their widespread use, MedDRA and WHO Drug are discussed in more detail than other dictionaries.

## *Introduction*

Recording and storing data in a controlled, consistent and reproducible manner for data retrieval and analysis is a necessity for regulatory compliance and clinical study success. To provide control and consistency, a variety of medical coding dictionaries may be used to process, analyze and report collected data. These dictionaries range in size and complexity from simple code lists with a few entries to large and complex dictionary systems containing thousands of entries and related tables. Two examples of commonly used dictionaries are the *Medical Dictionary for Regulatory Activities* (MedDRA) and the *World Health Organization Drug Dictionary* (WHO Drug). Some dictionaries are well established and have been used for years, while others are more recent and may be revised or updated regularly. Establishing and maintaining medical coding dictionaries are important tasks that clinical data management (CDM) personnel or coding specialists must carefully manage.

Transitioning to a new or different coding dictionary presents multiple challenges to CDM. First and foremost is the consistency between the same or related data that are analyzed and reported with different dictionaries (even a different version of the same dictionary). Any lack of familiarity with the content, organization, use, or capabilities of the new dictionary must be addressed prior to its implementation. Processes must be established for managing the release of multiple versions of the same dictionary, handling different dictionaries or versions that have been used, and integrating data coded with different dictionaries or versions.

## **Scope**

This chapter focuses on the management of dictionaries used for coding adverse events, medications, medical history, and other types of clinical data. Although the chapter touches on custom medical dictionaries, the primary focus is on standardized, commercially available medical coding dictionaries and some available options for coding management tools. Any use of the words “dictionary” and “code” within this chapter refer specifically to medical coding dictionaries and medical coding, as opposed to programmatic coding dictionaries and coding.

This chapter does not cover the actual process of coding; for more information on coding please refer to the “Safety Data Management and Reporting” chapter of the GCDMP, as well as the ICH-endorsed guide for MedDRA users, *MedDRA® Term Selection: Points to Consider*.<sup>1</sup>

## **Minimum Standards**

- Select dictionaries that meet project and regulatory requirements.
- Follow established security procedures for dictionary installation and maintenance.
- Ensure user licenses are obtained and kept up to date for any dictionaries and applications used.
- Ensure all sponsor personnel and vendors who will use the dictionaries hold the appropriate licenses. If a vendor has access to the dictionary application, ensure the application license covers vendor access.

- Implement an audit trail for all changes/updates to the dictionaries or synonym listings and support tables associated with the dictionaries.
- Do not modify published commercially available coding dictionaries. If a commercially published dictionary has been modified, then do not refer to it by its commercially available product name.
- Specify the dictionary name and dictionary versions used during coding on all study reports and integrated summaries.
- Store all utilized versions of dictionaries for future reference.

### ***Best Practices***

- Select a coding tool to facilitate consistent dictionary use.
- Include the version(s) of utilized dictionaries in metadata.
- Ensure all levels and versions of dictionaries used for coding can be accessed by data management and other dictionary users.
- Establish a process for evaluating term or categorization changes in a dictionary and its effect on previously coded data when moving to a different version.
- Ensure the capability to recode to different versions of a dictionary. For example, this may be needed to allow integrated study analyses to be reported using the same version.
- Ensure individuals who code data have training and professional background appropriate to the dictionary and the version for which they are coding. Training must be completed and documented before coding with the dictionary or version.
- Educate individuals involved in recording, monitoring, reviewing, analyzing and reporting coded data on the functionality and capabilities of the coding dictionaries used.

- Submit requested dictionary changes to the organizations responsible for maintaining the dictionaries using the appropriate approved process of submitting changes.
- Ensure change control processes are in place for all dictionaries, whether commercially available or custom.

## ***Established Standardized Dictionaries in Common Use***

### **MedDRA**

Recognizing the increase of global studies and submission of marketing applications to multiple regulatory agencies, the International Conference on Harmonisation (ICH) undertook the development of a global dictionary, which resulted in MedDRA.<sup>2</sup> The US Food and Drug Administration (FDA) is currently using MedDRA in its Adverse Events Reporting Systems (AERS).<sup>3</sup> MedDRA was planned to eventually replace some of the coding dictionaries already in use, such as COSTART and WHO-ART, as well as proprietary variations of those dictionaries that had been developed by sponsors of clinical studies. In many organizations, MedDRA has already replaced other dictionaries that were used in the past.

When MedDRA was initially published, updates were released on a quarterly basis. Following the release of version 4.0 in June 2001, the frequency of updates was reduced to semiannually. The organization responsible for publishing and maintaining MedDRA is MSSO (Maintenance and Support Services Organization). MSSO recognizes the need to stabilize MedDRA terminology to address concerns that an overwhelming amount of resources are needed to maintain frequent version updates and subsequent recoding and reanalysis of adverse events.<sup>4</sup> Since the initial release of MedDRA, revisions have addressed topics in the following areas.

- Updated assignments to system organ class (SOC)
- Consistent use of terminology
- Retirement of terms from current status
- Addition of new terms identified during implementation of the dictionary in clinical studies

A review of the impact of each change and whether an improved coded term is available in a new dictionary version is facilitated by the ability to search within various versions for coded adverse events and dictionary entries at each level. It is possible that an update to a given version will not contain any new terms in a particular grouping or modifications of existing coded terms.

MedDRA is a multiaxial dictionary, meaning that a preferred term (PT) may be associated with multiple SOCs. Each PT, however, is associated with only one primary SOC, regardless of the number of secondary SOCs with which it is associated. Sponsors and contract research organizations (CROs) frequently made changes to dictionaries prior to MedDRA, leading some users to believe the same could be done with MedDRA. MedDRA, however, should not be modified in any way by users. This prohibition of user modifications includes both changes to terms and changes to the assignment of a primary SOC for a term. MSSO has established a detailed process for users to follow, which involves bringing the issue to the attention of MSSO if they find a term to be lacking or in error.

## **WHO Drug**

WHO Drug is one of the more commonly used dictionaries, and was designed by the World Health Organization (WHO) for coding medications in clinical studies.<sup>5</sup> Medications used by study participants prior to or concurrently with a study are commonly coded to facilitate reporting and analysis. A variety of dictionaries or medication references provide information about prescription, generic, and over-the-counter (OTC) medications, as well as herbal supplements. The medication references used for coding medications should be selected based on the scope of medications included, how recently the reference has been updated, the frequency of updates to include the release of new medications, and coding information available in the dictionary. Such coding information may include generic terms, active ingredients, indications, or Anatomical Therapeutic Chemical (ATC) classification. WHO Drug is widely considered the most comprehensive resource for medication coding, and is also associated with a quarterly journal, *WHO Drug Information*, that discusses the most recent news and trends relating to medications and medication development.

In recent years, WHO Drug has been distributed in several formats, known as format B-1, format B-2 and format C. In format B-1, drug names may be

repeated within the dictionary if the same name is used for different drugs, which may occur due to each drug being marketed in different languages or countries. Format B-2 is similar to format B-1, except each drug name appears only once within the dictionary. When a drug name appears more than once in the B-1 format, the first entry that was entered into B-1 is typically used as the B-2 entry.

Format C is the newest of the three formats, uses a different file structure than the B formats, and also includes each drug's available dosage formulations (e.g., caplet, liquid, intravenous, etc.) and dosage amounts (e.g., 10 mg, 20 ml, etc.). Format C is much more specific than the other two formats because it can contain many more entries for the same drug, with each entry representing a unique combination of that drug's formulation and strength. Format C was originally intended to replace the B formats, but many companies had difficulties implementing it. As a result, the Uppsala Monitoring Centre (UMC), which is responsible for maintaining and licensing WHO Drug, agreed to continue distributing format B-2 indefinitely. However, the UMC has indicated that starting in 2009 it will no longer distribute the B-1 format, although the files will be available upon request. A WHO Drug license entitles the subscriber to receive both available formats (B-2 and C) of the dictionary.

In 2005, the UMC introduced the WHO Drug Dictionary Enhanced (WHO-DDE). The WHO-DDE combines data from the original WHO Drug Dictionary (WHO-DD) with additional country-specific drug information collected through the UMC's collaboration with IMS Health (an international consulting and data services company). The WHO-DDE is therefore several times larger than the WHO-DD.

New versions of WHO Drug are released quarterly, but companies have the option to receive new versions on a quarterly, biannual or annual basis. The cost for a WHO Drug license is dependent on the frequency that a company chooses to receive updates, with higher costs for more frequent updates. New subscribers only have the option to subscribe to the WHO-DDE, whereas subscribers who are currently receiving the WHO-DD have the option to continue receiving the WHO-DD or upgrade to the WHO-DDE.

## **Other Dictionaries**

Although MedDRA and WHO Drug are the most commonly used dictionaries for clinical studies and postmarket surveillance, the following list briefly describes a few established but not as widely used dictionaries.

- WHO ART—The World Health Organization Adverse Reactions Terminology is a dictionary designed by WHO for coding adverse reactions.
- COSTART—The Coding Symbols for a Thesaurus of Adverse Reaction Terms was developed by the FDA for coding and reporting adverse reactions. It was originally used by the FDA for coding adverse events, although it has since been replaced by MedDRA.
- SNOMED CT—The Systemized Nomenclature of Medicine—Clinical Terms was developed by the College of American Pathologists as a coding system to capture information about medical history, treatments and outcomes.
- CTCAE—Common Terminology Criteria for Adverse Events was developed by the National Cancer Institute as a system for classifying the nature and severity of adverse events. Work is currently underway to integrate CTCAE with MedDRA.
- ICD-9—Published by the WHO in 1977, this dictionary consists of coding for diagnoses and procedures.
- ICD-9-CM—An update to ICD-9, this dictionary became the official system for assigning codes to diagnoses and procedures in hospitals within the United States. Medicare and Medicaid have required the use of ICD-9-CM codes since 1988. These codes are updated yearly.
- ICD-10—Completed by WHO in 1992, and while implemented in most of the world, the dictionary was not adopted in the United States. This dictionary was originally designed to report mortality; however, modified versions have since been created for diagnosis codes (ICD-10-CM) and procedure codes (ICD-10-PCS).

## **Custom Medical Coding Dictionaries**

Custom dictionaries are typically developed to meet company-specific processes. Most custom dictionaries display terminology in a hierarchical pathway ranging from broad terminology to very specific terms. These dictionaries can be used to code adverse event data, medical history data and more commonly, medication data. Some organizations may create a custom dictionary by adapting a commercially available dictionary to better meet the organization's specific needs. If this approach is used, the customized dictionary should not be referred to by the same name as the commercially available dictionary.

There are limitations to using a custom dictionary, such as the lack of a central governing body to maintain the dictionary hierarchy for terminology and classification. Custom dictionaries also may not be consistent with terminology as it evolves over time (e.g., drug formulations may change over time or cease to be marketed). Although versioning is important for all coding dictionaries, some companies may not have a rigorous versioning strategy for custom dictionaries. All relevant regulatory standards should be taken into consideration when developing custom medical coding dictionaries. Additional steps for data reconciliation between different sources might be required when using custom medical coding dictionaries.

## ***Dictionary Application Software Selection***

When choosing a coding dictionary, one must also consider the software that will be used to house and search the dictionary. Some dictionaries are already packaged with an accompanying software application, but there are cases where the software must be chosen separately. In all cases the applications need to be validated prior to being put into production. In addition, proper validation of changes and updates needs to be performed prior to any changes or updates being released into production.

## **Application Service Provider (ASP)**

An ASP system can come in many variants depending on the contract with the provider. An ASP may host and manage the implementation and validation of dictionaries, or may provide customized tools for managing and using dictionaries. All types of ASP systems, however, share a common model,



which is that the software is owned by the ASP, usually runs at the ASP's data center using the ASP's servers, and the customer pays a monthly or other contracted fee for service. Most ASPs allow for minimal customization and do not allow for most company-specific items. Support is usually supplied by the ASP, although depending on the contract, some support may be provided by the customer as well.

## **Commercially Available Applications**

Commercially available applications are software packages that are purchased by the user, and may also be referred to as “off-the-shelf” applications. One key difference between commercially available and ASP applications are that with commercially available applications, the customer usually hosts and manages the application on their own servers. Commercially available applications also allow for more configuration options to meet an organization's specific needs. Commercially available software packages are usually more amenable to the use of “add-ons” to allow interaction with other software packages. To make changes to the application software, the request will need to go through the company that owns the software. Application support is typically shared between the software producer and the customer's information technology (IT) support department.

## **User-Built Applications**

Some organizations may choose to build systems that are tailored to the specific needs of the organization (e.g., logistics and workflow). In these situations, the organization hosts the software on their servers and provides all support services. The organization is also responsible for ensuring applications needing validation have followed appropriate software development lifecycle processes to validate the application and the functionality of the application after installation.

The benefit of user-built applications is that they can be customized to meet the organization's specific needs. A risk of user-built applications is that they are dependent upon organization resources. Another risk is that poorly written business requirements may result in an application that does not adequately meet the organization's needs once the application development is completed.

## **Medical Coding Tools and Methods**

In addition to the actual dictionaries and software applications used to house them, CDM personnel and dictionary users should be familiar with the following tools and methods used in dictionary management.

### **Autoencoders**

Autoencoding is a programmatically assisted process for matching a reported term to a dictionary term. A basic autoencoder will take a list of reported terms and look for an exact match with dictionary terms. Various methods exist for autoencoding, such as character string matches with the dictionary, character string matches with synonym lists, and matches found using algorithms. Within the context of autoencoding, a synonym list is a repository of terms that have previously been coded. Advanced autoencoder designs allow the user to define algorithms to assist with finding suggested “best” matches. These coding algorithms should be evaluated for their ability to handle synonym listings, misspellings, capitalization, word variations, word order, exclusion of irrelevant verbatim text, and other issues that may impede accurate matching. An autoencoder is useful when a large number of entries must be coded, and can expedite consistent coding by eliminating the requirement of manual reevaluation of previously coded terms.

Consistency checks can be performed within some autoencoders. Some autoencoders may also allow for multiple dictionaries and versions of the dictionaries. Added features may also include the ability to access multiple coding jobs on demand; create and maintain synonym lists; configure algorithm lists to support autoencoding; and integrate to safety and clinical databases. These broad-spectrum coding systems decrease regulatory risks and increase efficiency, providing more consistent and high-quality coding output.

Some clinical data management systems include an autoencoder and provide the ability to load electronic versions of coding dictionaries. Other autoencoders may be available as separate applications. Both integrated and standalone autoencoding applications must be fully validated according to current regulatory standards. Other features to be considered when selecting an autoencoder include ease of use, security features and, ideally, the ability to load and store multiple dictionaries or versions. Despite the assistance

provided by autoencoders, a manual review of coded data should be performed to ensure consistency and accuracy.

## **Manual Coding**

Manual coding refers to a situation where a person selects an appropriate dictionary entry for each reported term, either in the patient database or in a module of the dictionary application that deals with discrepancies. This method may be used when an autoencoder is unable to code a term or an autoencoder is not being used. Some clinical data management systems have the ability to use manual coding, but standalone manual coding applications are also available. Both integrated and standalone manual coding applications must be fully validated according to current regulatory standards.

Some manual coding applications use the same types of algorithms as autoencoders to provide the user with a list of suggested dictionary terms for a given reported term. Coding applications with this capability should be user-configurable (i.e., allowing for the creation and maintenance of lists of synonyms appropriate to the dictionary) and allow for suitable testing of the configuration to ensure that the suggested terms are accurate and comprehensive.

Ideally, a manual coding application will allow the coder to view all components of a dictionary (e.g., the full hierarchy for MedDRA or the ingredient list and ATC codes for WHO Drug), and also have the ability to see how other reported terms have been coded to ensure consistent coding of similar terms. Additional features to consider for a manual coding application are the ability to review coded terms for accuracy and consistency, the ability to query a term when it cannot be coded, audit trails that record the user and date/time a term was coded, and extensive, easy-to-use search capabilities.

## **Hybrid Approaches to Coding**

A hybrid approach to coding uses an autoencoder to first automatically code those reported terms that match a dictionary term or that match a term that has previously been coded (i.e., a synonym list). The terms that are not autoencoded are then manually coded. Many clinical data management systems and standalone coding applications support this hybrid approach to coding.

A coding application being used in a hybrid approach should have the same features desired in an autoencoder or a manual coding application. In addition, a hybrid coding application should allow a coder to easily see the terms that did not autoencode, and which will therefore require manual coding. Some hybrid coding applications may provide the ability to distinguish between autoencoded and manually coded terms and a facility to manually override any autoencoded terms, if necessary.

## **Browsers**

A browser is a computerized tool used to aid in accessing terms in a specified dictionary. Browsers are designed to quickly find terms of interest and should be flexible, intuitive, and quick to use.

- Stand-alone browsers—These are applications that allow for the easy search and review of dictionaries. Some also possess a capability for limited linking to external applications (e.g., study databases), where one may not be able to affect a term or coding change from the browser, but would be able to call (or open) the browser from within the dictionary application.
  - WHO Drug—Several WHO Drug browsers with differing feature sets exist, including one produced by the Uppsala Monitoring Centre (which is an entity of WHO that works with international drug monitoring).
  - MedDRA—An application has been provided by the MSSO for searching the MedDRA dictionary, but other vendor-created browsers also exist, with differing feature sets.
- Browsers that are contained within dictionary management systems have enhanced capabilities, although the availability of these enhanced capabilities varies across available systems. Some of these systems can act as a browser, as well as a vehicle for importing and exporting individual reported terms or a batch of reported terms. Various coding approaches outlined above can be performed once the terms are imported into the system.

## **Dictionary System Validation**

Any dictionary application or system for housing dictionaries requires documented validation prior to being placed into production. This validation should include system validation, unit validation (if this level of detail is needed) and user acceptance testing. Full documentation should be maintained for the application, including business requirements, system requirements, design specifications and any other documents or support level agreements that are in place for the system.

The level of validation to be performed by an organization may vary depending on the origin of the system. ASP and commercially available applications may require less validation effort than a user-built application or system. Regardless of whether performed by an ASP, software vendor, or the organization conducting the research, all systems and applications require validation and supporting documentation to meet industry and regulatory standards, as well as to pass audit inquiries.

To prevent any untoward effect on subject data, changes to an application, whether a bug fix or a planned upgrade, may require validation and testing prior to being placed into production. The dictionary application or system that houses the dictionaries also requires documented change control and version control procedures. Change control procedures and version control schemas are usually set by the IT department of the organization to ensure clinical study software needs meet the standards of good clinical practice.

## **Change Control**

The practice of modifying published dictionaries is clearly discouraged by the ICH for the MedDRA dictionary.<sup>1</sup> The organizations that maintain dictionaries have an established process for submitting change requests if, for example, an adverse event term or medication is reported that is not included the dictionary. This process allows for a review of the requested change and dissemination of the request to others using the dictionary. An approved request will appear in a future release of the dictionary. A declined request will not.

Although in-house modifications are highly discouraged, any in-house modifications made to a published dictionary should be clearly stated in

reports, so as not to mislead reviewers who are familiar with the published dictionary. Any changes made to dictionary entries should also have documented authorization and be included in an audit trail.

Coding dictionaries may be available in electronic and/or printed format, and multiple versions may be released or published. The dictionary and version used for a given project, time period, or data set should be clearly documented. Where this information is documented may vary between organizations (e.g., in a data management plan or coding guidelines), but the dictionary and version should be referenced in clinical study reports or integrated summaries that report on the coded terms. For multiple ongoing studies, the study team should determine which dictionary and version will be used for coding each study. A systematic process and instructions should be in place to ensure the consistent use of the appropriate dictionary and version. Processes should be established for evaluation of the extent of changes between versions, the impact of changes on previously coded terms, and criteria for recoding and implementing the latest version.<sup>6</sup>

Using different dictionaries or versions over a period of time increases the importance of version control, documentation and standardized data reconciliation processes. For example, different versions may be used for coding postmarket safety data versus clinical data, between different studies for the same drug, or even within long-term studies. The impact of version changes can be greater for adverse events, because a term may be deactivated or reassigned to a more appropriate term, rendering the earlier term assignment outdated. Most of the changes to medication dictionaries simply introduce new medications.

Dictionary and version information may be maintained within the clinical database, within the autoencoder as the dictionary files are loaded, or within the metadata of data sets containing coded data. Because version information may be incorporated into electronic files by organizations maintaining published dictionaries, that information may be available without the need for additional in-house action.<sup>7</sup>

Process steps for installing and upgrading to new dictionary versions may vary between organizations and specific dictionaries. However, dictionary installations or upgrades should be subjected to a holistic approach to validation once installed, including processes such as remapping synonym tables and recoding subject data repositories.

## **Recommended Standard Operating Procedures**

- Maintenance of Dictionaries
- Security, Change and Version Control
- Validation and Testing Procedures

## **References**

1. International Conference on Harmonisation Secretariat. *MedDRA<sup>®</sup> Term Selection: Points to Consider*, Release 3.2. Geneva, Switzerland: IFPMA; 2003. Available at: [http://www.meddramsso.com/mssoweb/docs/PTC\\_3.2\\_final.pdf](http://www.meddramsso.com/mssoweb/docs/PTC_3.2_final.pdf). Accessed December 3, 2008.
2. MedDRA MSSO Web site. Available at: <http://www.meddramsso.com/>. Accessed December 3, 2008.
3. US Food and Drug Administration. *Reviewer Guidance: Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review*. Washington, DC: US Government Printing Office; 2005. Available at: <http://www.fda.gov/CDER/GUIDANCE/3580fml.pdf>. Accessed December 3, 2008.
4. MedDRA MSSO. *Messenger Quarterly Newsletter: May 2002*. Reston, VA: Author; 2002.
5. Uppsala Monitoring Centre Web site. Available at: <http://www.who-umc.org/>. Accessed December 3, 2008.
6. MedDRA MSSO. *Recommendations for MedDRA<sup>®</sup> Versioning for Summary Reporting*. Reston, VA: Author; 2000.
7. MedDRA MSSO. *Messenger Quarterly Newsletter: January 2001*. Reston, VA: Author; 2001.

## **Further Reading**

N/A

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2003	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
May 2009	Revised for content, style, grammar, and clarity.



---

# Safety Data Management and Reporting

May 2007

## *Abstract*

Collecting and reporting information about the safety of an experimental compound or product constitutes a significant challenge for clinical data management. This chapter reviews the wide range of factors that must be considered for the successful completion of a project's safety data management and reporting responsibilities. Industry guidelines and regulations for collecting and reporting reliable, high-quality safety data are discussed. The importance of degrees of precision and descriptions of severity when capturing data about adverse events is emphasized. The use of medical dictionaries, especially MedDRA, is reviewed with consideration for the process of encoding safety data to dictionary terms and various approaches to this task. Laboratory data and other forms of data, such as specialized tests, are discussed as potential sources of safety data. Special consideration is given for the capture of serious adverse events and their reporting to regulatory agencies. General issues to consider when reporting safety data to the FDA are also discussed.

## *Introduction*

Safety data often present the most challenging aspects of the management and reporting of clinical trial data. Consideration for return-on-investment frequently curtails the query process for cleaning safety data and limits reporting methods.

Estimating resource requirements and balancing business value against scientific theory are critical to the planning of effort. Scientific principles also motivate clinical trial scientists to use judgment in determining the standards to be set for a given study or program, the quality markers to be used, the levels of precision, and the depths of analysis and reporting. When information that has a soft basis is stored and cleaned as if it has a high degree of precision and reliability, reports can reflect an over-reliance on questionable data and lead to inferential errors. Soft information can still be

quite useful, but to avoid misrepresentation, a clear identification of the nature of the data is necessary.

The quality of data is really determined in the field. If the quality of the information that is recorded in source documents is poor, data managers or statisticians can do little to repair it. Instead, data managers should ensure that the database accurately conveys the limitations of the data's quality to users. Statisticians have an imperative to ensure that analyses and data displays acknowledge their limitations.

The processes of data capture, management, and reporting are highly integrated. Considerations of best practices for reporting guidelines would be deficient in absence of guidelines for the earlier processes.

## **Scope**

To the clinical trial scientist, the safety data in a clinical study are simultaneously a rich source of information and an enormous challenge. The data manager and statistician who are a part of the product team must work closely with each other and with other team members to ensure that safety data are captured in a sensible way to facilitate proper interpretation and meaningful analysis and summary. Ensuring quality requires that the team capture, process, and report the data in a way that facilitates the drawing of reliable conclusions. When determining the balance between business and science, data managers and statisticians must consider that resources may be expended on efforts that have no effect on conclusions.

Safety data may be displayed and reported in many ways. To ensure adequate reporting of results that pertain to product effects, judgment and scientific selection are needed to identify the trends and salient features of the data. Producing voluminous pages that are incomprehensible and clinically meaningless can dilute real effects. However, the discernment of these effects is the driving goal of the safety data processing and reporting.

This chapter discusses practices, procedures, and recommendations for data managers to operate within the project team and to work closely with statisticians, monitors, and clinical research so that data management practices support statistical and medical purposes. Data managers are better equipped to function as fully-integrated team members when they have a basic

understanding of the activities and needs of other team members, particularly statisticians.

### ***Minimum Standards***

When considering the capture, management, analysis, and reporting of safety data, the following minimum standards are recommended:

- Ensure compliance with regulations.
- Ensure that the standard of quality supports the utilization of the data.
- Ensure that conclusions about the safety profile of a compound can be reliably drawn from the database.
- Ensure that safety risks are identified and reported accurately.
- Ensure that normal ranges are properly linked to laboratory data. If normal ranges are unavailable, ensure that the reference ranges which are used are documented as such. This standard is especially crucial when normal ranges are updated frequently.

### ***Best Practices***

When considering the capture, management, analysis, and reporting of safety data, the following best practices are recommended:

- Develop CRFs with teams of individuals from the monitoring, data management, statistics, regulatory affairs, and medical departments, thereby ensuring adequate attention to the collection of safety data.
- Consider the level of precision that can be attained in the study and select the CRF format for collecting AEs appropriate for that level. Also, consider the level of precision in the analysis.
- Define severity, with an understanding of its uses and limitations.
- Examine laboratory data from the perspectives of categorical shifts, changes in magnitude for the group, individual significant values or

changes, and listings. Consider related parameters for compounds with potential toxicity in specific body systems.

- Consider laboratory normalization techniques when combining data across studies or centers where varying normal ranges are used.
- Include data managers and statisticians working together when considering computerization, management, reporting, and analysis of safety data. These tasks are highly integrated and require joint considerations of individual team constituents. Develop standard operating procedures (SOPs) for data capture, data validation, statistical analysis, and reporting of data. The SOPs should include guidelines for this team approach.
- Document the status and quality of safety data, and include this documentation with the database.
- Include clear links for comparators, such as normal ranges for laboratory data, with the database.
- Consider levels of precision in the capture and the reporting of safety data to reduce the likelihood of over-interpretation or misinterpretation.
- Understand that time-to-event analyses are only meaningful when the timing of the event is reliably known.
- Consider both categorical shifts (from a status of normal to abnormal) and magnitude changes for analysis and reporting of laboratory data. An examination of significant values may provide different information from an examination of significant changes.
- Apply standards commensurate with the utilization of the results residing in the databases when using databases for safety reporting (e.g., expedited reporting, ongoing review by monitoring boards, or routine reporting). If important decisions will be made based on the information in the database, know the data's appropriateness and level of quality.

### ***Available Guidelines***

One definition of “quality data” is “a collection of data from which reliable conclusions can be drawn.” The goal of reporting safety data is to convey

information that would facilitate the drawing of reliable conclusions. Generally, one of the key objectives in investigative clinical research trials, is to characterize, investigate, establish, or confirm the safety profile of an investigational product. The management and reporting of the safety data from the trial should support that objective.

The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has issued several guidelines to provide guidance to the industry for how to manage and report clinical trial safety data. These guidelines are as follows:

- *E1A* describes expectations for extent of population exposure for drugs intended for long-term treatment of non-life-threatening conditions. The guideline acknowledges that safety evaluation during clinical drug development is not expected to characterize rare adverse events (AEs), such as AEs that occur in less than 1 in 1000 subjects. Total short-term exposure is expected to be about 1500 subjects. Exposure for six months by 300 to 600 subjects should be adequate. Exposure for a minimum of one year by 100 subjects should be adequate. Exceptions are noted.
- *E2A*, *E2B*, and *E2C* are clinical safety data management guidelines. They provide guidance for definitions and standards for expedited reporting, for the data elements for transmission of individual case safety reports, and for periodic safety update reports for marketed drugs.
- *E3* is the guideline on “Structure and Content of Clinical Study Reports.” This guideline provides detailed recommendations and specific suggestions for data displays of safety data. It is noted that the guideline shows “demography” as a subsection of “efficacy evaluation” and “extent of exposure” as a subsection of “safety evaluation.” For studies for which doing so makes sense, and for integrated summaries, FDA regulations require that efficacy and safety data be analyzed with particular consideration in regard to age, sex, and race. ICH guidance encourages that the analysis of both efficacy and safety data consider extent of exposure, including compliance. It is imperative to understand that demography and dose exposure relate to efficacy and safety. Therefore, the analysis and reporting of safety data should consider the characteristics of the presenting population and the extent of exposure to the investigational compound.

- *E5, Ethnic Factors in the Acceptability of Foreign Clinical Data* advises that there are concerns “. . .that ethnic differences may affect the medication’s safety, efficacy, dosage, and dose regimen.” This guideline also delineates between extrinsic ethnic factors—those factors associated with environment and culture (e.g., diet, use of tobacco, use of alcohol)—and intrinsic ethnic factors—those factors that help define and identify a subpopulation (e.g., age, sex, weight, organ dysfunction).
- *E6* is the consolidated good clinical practice (GCP) guideline. This guideline contains principles of GCP that underscore the scientific basis of the clinical trial and specify qualifications for the personnel and systems involved in all aspects of the clinical trial. The guideline also asserts that adherence to good scientific principles is required and that the documentation of the adherence is needed.
- *E9* is a guideline geared toward the statistician, which includes substantial advice for the analysis of safety data.

Other guidance documents that give advice for capturing, managing, and reporting safety data are available from the ICH and from regulatory agencies. Sponsors should refer to IND regulations (21 CFR 312) and NDA regulations (21 CFR 314) to ensure compliance with FDA regulations for investigational and marketed products.

## **Safety Reporting**

Safety data are reported and examined at various stages of an investigation and by different assessors. IND regulations specify expedited reporting for serious or alarming adverse events. Many studies have safety data monitoring boards (SDMB) that review data as they accumulate in a study. The sponsor’s medical monitor reviews safety data, frequently masked to the treatment. Then, after market approval, there are NDA regulations that specify safety reporting. Data managers and statisticians need to ensure that the reports provided are supported by the quality appropriate for the purpose of the report.

The FDA requires sponsors to meet their obligations to Congress and to the public. Compliance with IND and NDA regulations is aided by an

understanding of the mission and motivation of these regulations. Before marketing, IND regulations apply.

One key purpose of the IND regulations is to facilitate the FDA's monitoring of the investigation, including protection of the safety and rights of individuals enrolled into trials, and the scientific quality of the investigation in terms of its ability to adequately demonstrate the efficacy of a compound. The FDA requires annual reports, which are brief updates concerning the progress of the investigation, including any newly identified safety trends or risks that may impact the investigation. FDA also requires expedited reports of "any adverse experience associated with the use of the drug that is both serious and unexpected" (21 CFR 312.32). Written notification of such events is required within 15 calendar days. For events that are fatal or life threatening, a telephone or facsimile transmission is required within seven calendar days. Additional details of IND safety reports, annual reports, and IND specifications are provided in 21 CFR 312.

After marketing, the FDA has a different perspective and a different goal. If a recall is necessary after the compound is in medicine cabinets, it becomes much more difficult (if not impossible) for the FDA to retrieve the compound. The regulations provided in 21 CFR 314 describe reporting requirements after approval as follows: For three years after approval, periodic reports are required quarterly. After the initial three years, reports are required annually. Moreover, under NDA regulations, each adverse experience that is both serious and unexpected, whether foreign or domestic, must be reported within 15 calendar days. Additional details of NDA safety reporting, periodic reports, annual reports, and NDA specifications are provided in 21 CFR 314.

In addition to the FDA's monitoring of investigations and review of safety data, the FDA requires sponsors to employ medical monitors who review safety data. Sponsors frequently have safety data monitoring boards, comprised of individuals separate from the conduct of the study, that conduct interim analyses and review accumulating data, blinded or unblinded. Data monitoring boards can make recommendations or decisions to halt an ongoing investigation due to (1) overwhelming efficacy, (2) unacceptable safety risk, or (3) futility. These boards may also make recommendations for changes in the ongoing study, such as a dose reduction or the elimination of an arm of the study with an unacceptable safety risk.

Any review of safety data that is based on reported information from a safety database (as opposed to CRFs) relies on that database. If the quality is poor, the decisions taken may be wrong. Review of accumulating data often implies a mixture of complete data with partial data and a mixture of clean data with dirty data. To provide the optimal information to the users of the dynamic database, the quality should be known and reported to the reviewers with the safety data. However, it is generally not helpful to report to data reviewers that some data are dirty without specifically identifying which data are dirty.

## ***Capture, Management, And Reporting Of Adverse Events***

Clinical adverse events frequently house the most important safety information in a clinical study. Ensuring that methods of collection, coding, analysis, and reporting facilitate the drawing of reliable conclusions requires an understanding of the characteristics and limitations of adverse event data.

### **Precision**

The precision with which AE data are captured relates directly to how the data can be analyzed and reported. There are three basic types of precision in a clinical trial:

- **High Precision**

Investigation in a Phase One sequestered environment (i.e., a phase one house) often incorporates medical monitoring that is continuous and high-precision. With a few subjects in a sequestered environment, a nurse or physician is by the bedside continuously. In such an environment, clock time may be recorded so that precise data can be collected for onset and offset of an AE. Hence, duration of the AE and elapsed time since initiation of treatment can be calculated in a meaningful way. Clock time is meaningful in such an environment for some events, although it may be difficult to assess the precise minute that sleepiness begins or a rash is cleared.

- **Moderate Precision**

Investigation in a hospital often incorporates medical monitoring that is daily, frequent (but not continuous), and moderate-precision. Hospitalization offers a controlled and sequestered environment such that



a nurse or physician can assess the subject daily. In such an environment, clock time may not make sense for all events, but date can be precisely recorded. Onset and offset of an AE can be recorded in terms of days but not hours. Duration (in days) and elapsed days since initiation of treatment of the AE can be calculated.

- **Low Precision**

Investigation in an outpatient study where subjects return to the facility after days, weeks, or months incorporates low precision. In such an environment, clock-time and date may not be meaningful. Use of subject diaries may assist with the determination of the duration of the AE or elapsed time since treatment. However, subject diaries are frequently inaccurate. In such studies, it is recommended to capture frequency (e.g., single episode, intermittent, continuous), maximal severity, most-harsh relationship, and other such information rather than to attempt to record each event with time of onset and offset.

When an investigation is of low precision but attempts have been made to record data as if it were moderate or high precision, the result is generally a database with dates (or times) that are rough guesses and that may be far from accurate.

The precision with which AE data were collected has an important impact on how the data can be analyzed in a meaningful way. In an outpatient study, dates cannot be interpreted with the same reliance as in a sequestered study. When dates are present in the database, it may be tempting for the statistician to employ survival-analysis techniques to analyze time-to-event. However, if these dates are inaccurate, the resulting analysis can lead to incorrect or unreliable conclusions.

## **Severity**

When considering the capture of severity of adverse events, it is tempting to make the assessment in terms of its impact on activities. This method of assessment may be meaningful for some events, such as “pain,” but not meaningful for others, such as “alopecia.” In some cases, severity is not assessable at all. For example, “mild suicide” is not meaningful. Some events are episodic rather than graduated by severity, such as “hair-line fracture.” For

example, an assessment of diarrhea as “severe” is often made because of duration or frequency of episodes (which are different parameters). However, diarrhea is episodic.

The concept of severity is only meaningful within a particular event. When one considers severity of AEs for an organ class (e.g., CNS), ranking among mild, moderate, and severe AEs is not meaningful. If one considers “mild stroke” and “severe flush” (both CNS events), these rankings are not sensible compared to rankings such as “mild headache” and “severe headache” for which a relative ranking does make sense.

A common data display that is encouraged by the ICH and the FDA is a breakdown by severity. In this context, it is easy to confuse severity with seriousness or to misinterpret severity altogether. A breakdown that ignores the particular events and that counts mild AEs separately from moderate AEs will give a distorted assessment when the same study includes reports of “mild stroke” or “mild MI” and also reports of “severe rash” or “severe sleepiness.” A more meaningful display breaks down severity within a particular event.

## ***Dictionaries***

AE dictionaries are needed to group data for meaningful analysis. MedDRA is the ICH-developed and recommended dictionary for all medical events captured in clinical trials, including, but not limited to, AEs.

Use of MedDRA requires an understanding of its levels of terms and an understanding of its multi-axial functionality. The levels of terms used in MedDRA are the follows:

- Lowest level term (LLT)
- Preferred term (PT)
- High level term (HLT)
- High level group term (HLGT)
- System Organ Class (SOC)

It is noted that the SOC level within MedDRA is really a dual level, because MedDRA permits a primary SOC and one or several secondary SOCs.

The multi-axiality of MedDRA permits a single AE to be simultaneously coded to many SOCs. For example, a migraine headache could be coded to the nervous system (because of the involvement in the brain), the vascular system (because it is a vascular disorder), the GI system (if there is associated nausea and vomiting), eye disorders (if there are visual disturbances), or other SOCs, as applicable.

MedDRA is not just another dictionary. It is a distinct approach to thinking about medical information. Managers of medical information have an imperative to understand the flexibility of MedDRA as well the implications that its storage and implementation can have on safety reporting.

### **Dictionary Version Control**

Updated versions of dictionaries frequently change pathways to body systems or organ classes. Such changes in a dictionary can have a substantial effect on conclusions regarding a product's effects on the body. Thus, the version of a dictionary used for classification of AEs into body systems can impact the labeling of the product. As there must be a clear trail leading from the data to the labeling, the data manager who will implement a dictionary for a study (or product) must ensure consistency, when possible, and the ability to replicate.

Most standard dictionaries that have been widely used have been reasonably stable (e.g., COSTART, WHO, ICD-series, and so on). MedDRA is updated periodically. Dictionary version management requires more resources when updates are more frequent.

For purposes of medical monitoring, interim analysis for safety review boards, or other important purposes, one suggested practice for ensuring consistency within a long-term study is to execute the dictionary against the AE data as the study progresses. Then, the dictionary should be re-executed using the most reasonably current version of the standard dictionary. This approach ensures that the entire study is executed against the same version of the dictionary. Because additional queries may result at the time of re-execution, this final execution of the dictionary should occur *prior to database-lock*.

To ensure reproducibility, the version of the dictionary used in any study should be stored with the database.

## **Encoding**

Auto-encoding is a highly recommended practice to facilitate the execution of a dictionary against AEs. Auto-encoding software is available to assist with the programming aspect of this task. To cultivate an understanding of the coding process, training of the monitors and site personnel should facilitate capture of AE data in a format that can be auto-encoded. Training should include guidelines such as the following:

- Avoid use of adjectives as initial words (e.g., “weeping wound” may be coded to “crying”; “faint rash” may be coded to “syncope”).
- Avoid the use of symbols and abbreviations in the AE text, as they may be interpreted differently.
- Avoid inclusion of severity in the AE text (e.g., “severe headache” in the AE text inhibits auto-encoding; severity should be recorded in the severity field, not the AE text).
- Ensure that AE text has a clinical meaning (e.g., “bouncing off the walls” and “feeling weird” are difficult to interpret).
- Ensure that AE text has a clear meaning (e.g., “cold feeling” may be interpreted as “chills” or “flu symptoms”).

Encoding within the database may add unnecessary complexity to the management of the database when final coding requires judgment. If the auto-encoding is done within the database itself and a medical judgment that is made after database lock indicates that the default pathway inaccurately captures the medical condition, the database would have to be unlocked. Performing auto-encoding in a separate file (e.g., an AE analysis file) offers the possibility of reflecting changes in medical judgment after database lock, if deemed essential. However, this practice imposes the need for an audit trail on analysis files.

## **Hard-coding**

Hard-coding, or coding outside the clinical database, is generally a dangerous practice. For coding AEs, hard-coding is sometimes used to introduce medical judgment that the standard dictionary does not offer. When events such as “strange feeling” are reported and no additional information from the site is

available, the medical monitor for the study may have insight that assists with the codification of the event, which can be inserted into the AE analysis file through hard-coding. It is possible to use “pass-through” text for the AE preferred term and hard-code the body system. Conventionally, many sponsors make use of quotation marks to indicate verbatim text that is passed through by a program to the preferred-term field. Any use of hard-coding requires careful documentation.

## **Lumping and Splitting**

Coders can be categorized into “lumpers” and “splitters.” No universally agreed-upon method exists for handling AE text with more than one event. “Tingling in hands and arms” is regarded by some coders as a single event and by other coders as two events. However, the decision to lump or split AE text has consequences.

When two events are reported in the same text field (e.g., “indigestion and diarrhea”) and splitting is done by the data management staff rather than the site, inconsistencies within the database may result. When the data manager splits the AE text into two or more events, the associated items are frequently duplicated (or replicated). For example, if a medication is given for treatment of the AE and the concomitant medications page of the CRF shows one event as the reason for use (e.g., “indigestion”), the splitting of the two events results in an AE with treatment given for which no treatment is recorded.

Medical judgment may also be inadvertently introduced into the database by the data manager. If the severity of the compound event is recorded as “severe,” the duplication of the attributes of the AE imputes “severe” to the other event(s). However, this outcome may not reflect the physician’s judgment for that particular component of the AE.

Coding of AEs has significant impact on the analysis and interpretation of the safety data for a product. The perspective that coding is a clerical function is naïve and risky. As the world moves toward the full implementation of MedDRA, the role of coding will have an even greater impact on the interpretation of safety data.

## **Capture, Management, and Reporting of Laboratory Data**

The characteristics of laboratory data differ importantly from most other types of data. Most clinical adverse events can be observed by either the subject or the physician. However, an elevation in bilirubin or cholesterol is not generally observable. For example, even in high-precision studies, it is impossible to know the time of an elevation of a clinical chemistry analyte. At the time of a blood draw, whether or not the value is elevated can be known, but when the value became elevated is unknown.

The peculiarities of laboratory data need to be respected in the management of the data. Attention is required to ensure that the storage of units of the data clearly reflects the values that were captured; In many databases, units are separated from the values. When data across studies are combined, it becomes particularly challenging and important to ensure proper linkage with the units. This linkage can protect against unreliable conclusions being drawn from the reported laboratory data.

One of the most challenging aspects of managing laboratory data is linking the data to the appropriate normal range. In the capture of data, if the data do not come to the data manager electronically, attention should be given to ensure the link between each value and the appropriate normal range.

When normal ranges are not available or not obtainable, reference ranges—ranges derived from normal ranges that are available in the study or from a reference book—may be used. However, documentation of the use of reference ranges in lieu of normal ranges must be clear for users of the database.

Normalization techniques for laboratory data are often employed for such purposes as conveniently combining data across studies. Normalization techniques generally include a transformation of the data into a unitless value between “0” and “1” when the value is normal, below “0” when the normal is below the lower limit of the normal range, and above “1” when the value is greater than the upper limit of the normal range.

If judgment and selection are not a part of the planning for data displays, reporting laboratory data can be prohibitively resource-intensive. The ICH and FDA have given specific guidance for how to report laboratory data.

## **Treatment-emergent Abnormal Values (TEAVs)**

For hematology, clinical chemistry, urinalysis, or other laboratory panel or group, comparative data summaries and supportive listings that provide a one-page summary by treatment group (for parallel studies) for analytes included in the study are strongly encouraged. Such a summary provides a valuable overview of movement from the normal state pre-dose to an abnormal state at any time post-treatment, in either direction, and for any analyte.

## **Clinically Significant Values or Changes**

Comparative data summaries and supportive listings are recommended. These documents provide summaries and details by treatment group of analytes with significant changes or values, such as an analyte for which the baseline value is doubled or tripled, an analyte for which the value is observed to be twice the upper limit of the normal range, or an analyte for which the change in value exceeds the width of the normal range.

## **Groups Means and Changes**

Displays of means and mean changes from baseline levels are useful within a group—to indicate a trend in an analyte—or among groups—to examine treatment group differences or trends that may be dose-related.

## **Shift Tables**

Shift tables frequently are 3x3 tables that show the status before treatment compared to the status after treatment (e.g., below normal, normal, above normal, in both cases). These displays ignore magnitude of change. The display depicts the movement, or lack thereof, from one category before treatment to another category after treatment.

## **Individual Data Displays**

Listings of individual data are needed for adequate reporting of most clinical trials. When the study is large, individual listings may be voluminous. Therefore, reporting needs to consider practical aspects of summarization.

## **Related Groups of Analytes**

Summaries by related groups of analytes are useful for some studies or integrated summaries. For example, products that may be prone to cause liver damage may need careful examination of analytes that relate to hepatic function. For the hepatic-function-related analytes, it may be useful to prepare a summary on a single page that includes proportions of subjects who double the baseline, triple the baseline, have a change of fixed magnitude, or exceed an alert or toxic threshold.

## **Other Data**

Safety data can have forms other than AEs and laboratory values. Capture of data from specialty tests (e.g., electrocardiograms, electroencephalographs) requires an understanding of the common data derived from the test and of the format, precision, and special attributes of the data.

Physical examinations are customary in clinical trials. In a broad sense, the physical exam is a screening method; if an unexpected, significant abnormality is detected during a physical exam, a specialty test is generally used to confirm the event. In this case, the data from the specialty test has greater reliability.

In considerations of data capture, free-text commentary boxes are generally discouraged. If they are used for medical monitoring purposes, they can be shaded so that the reviewing medical monitor can have the prose handy, but the text does not need to be computerized. Making effective use of the investigator's comment log can ensure that essential text (which is generally minimal) is computerized, if warranted.

The management of "other data" depends on the form of that information. For physical examinations or specialty tests for which free-text commentary is permitted, methods exist for managing the commentary without compromising the quality standards of the database.

Free-text commentary can be computerized using word-processing rather than a data entry system. Subsequently, the commentary can be proofread rather than double-keyed. Through this method, the free-text commentary can be computerized and linked to the database without being a part of the database



itself. As a result, quality standards can be maintained for the database proper, but reasonable standards may apply to free-text prose.

One method used by some sponsors that avoids computerization of verbose commentary is codification, in which a medically qualified individual reads the information and judges it to be relevant, not relevant, or perhaps critical. A code can be applied and keyed, where “0=no comment,” “1=comment, not relevant,” “3=comment, relevant,” and “4=comment, critical.”

### ***Serious Adverse Event Data***

Expedited reports are required by regulatory agencies for certain serious adverse events. In many companies, receiving reports of serious adverse events (SAEs), computerizing these reports, and managing these reports is the responsibility of a dedicated group of individuals. Often, this group is separate from the data management group that is responsible for computerizing and managing data reported from clinical trials.

The SAE database often includes safety data from various sources. Reports can be received from patients in clinical trials, from spouses who took trial medication (accidentally) and had AEs, or from patients who took marketed drugs and who are not participating in any trial. These reports can come from individuals who give reports over the telephone to a sponsor, from employees who report to the sponsor that they were told about adverse reactions to marketed products, from physicians, from the literature, and even from regulatory agencies. These reports are generally single-keyed, often by individuals other than professional data managers, and generally are not queried. The data within these SAE databases may be dirty, incomplete, duplicate, fragmentary, or have other issues. In contrast, the reports of SAEs from clinical trials that are reported on the AE page of the CRF are subjected to rigorous data management procedures, including scrubbing, querying, and verification to ensure accuracy.

These two types of databases generally have important differences in their sources, their quality levels, their uses, and their customers. Reconciliation of SAE data and the clinical trial database that houses the relevant SAE reports is not always straightforward. Different sponsors have vastly different methods of managing these two databases.

Good clinical data management practices include provisions for reconciling important disparities between serious adverse events that are captured both in the SAE database and in the clinical trial database. The business-balance perspective encourages users of these databases to recognize that clinical trial databases may be queried or updated while SAE databases are not and that, consequently, some discrepancies may exist because preliminary medical judgments were later changed in light of updated information.

## **General Safety Data**

The FDA draft document *Reviewer Guidance: Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review* (November 1996) provides specific guidance to industry that reflects thinking within the FDA about safety data.

In the above-referenced document, the FDA described the concept of clinical domains for a review of the following systems:

- Cardiovascular
- Gastrointestinal
- Hemic and Lymphatic
- Metabolic and endocrine
- Musculoskeletal
- Nervous
- Respiratory
- Dermatological
- Special Senses
- Genitourinary
- Miscellaneous

In the guidance document, the FDA specifies that an NDA should be reviewed against each clinical domain with two key questions as goals:

- Are the safety data adequate to assess the influence of the product on the clinical domain?
- What do the data indicate about the influence of the product on the clinical domain?

Statisticians who are involved with the reporting of safety data have an imperative to review safety data and ensure that the influence of the investigational product on each clinical domain is described clearly.

The design of the study must be considered in reporting clinical trial safety data. In a multi-center study, the ICH and the FDA urge an examination of the influence of center effects on the results to ensure that the results are not carried by a single center or dominated by a small proportion of the total study.

In a multi-center study, center effects are typical and are a nuisance. There are three sources of contributions to center effects:

- The investigator as an individual (e.g., the bedside manner, personal biases, and peculiar methods of assessment)
- The environment (e.g., equipment, SOPs, and staff)
- The subject population (e.g., those people who frequent the hospital or clinic, VA hospital, university hospital, or country clinic)

When the study employs one investigator who may be on the staff of several hospitals, or when a cluster of hospitals shares equipment and has common SOPs, or when a study makes heavy use of referrals, these attributes affect the interpretation of the center's effects. Reporting data in a multi-center study requires understanding the source of variability among centers and the reasonableness of displaying data by center or by clusters of centers.

### ***Recommended Standard Operating Procedures***

- Coding of Adverse Events

- Maintenance of Coding Dictionaries
- Reconciliation of Serious AEs in SAE Database with Clinical Trial Database
- Management of AE Analysis File
- Management of Laboratory Data and Normal Ranges
- Preparing Integrated Summaries of Safety Data

## **References**

1. *Code of Federal Regulations*, Title 21, Volume 1, Part 312, *IND Regulations*. Washington, DC: US Government Printing Office; 1997.
2. *Code of Federal Regulations*, Title 21, Volume 1, Part 314, *NDA Regulations*. Washington, DC: US Government Printing Office; 1997.
3. *Code of Federal Regulations*, Title 21, Volume 1, Part 11, *Electronic Signature Regulations*. Washington, DC: US Government Printing Office; 1997.
4. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E1A: The Extent of Population Exposure to Assess Clinical Safety*. Geneva, Switzerland: Author; 1996.
5. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E2A: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting*. Geneva, Switzerland: Author; 1996.
6. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E2B: Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports*. Geneva, Switzerland: Author; 1996.
7. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E2C: Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs*. Geneva, Switzerland: Author; 1996.

8. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E3: Structure and Content of Clinical Study Reports*. Geneva, Switzerland: Author; 1996.
9. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E4: Dose-Response Information to Support Drug Registration*. Geneva, Switzerland: Author; 1996.
10. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E5: Ethnic Factors in the Acceptability of Foreign Clinical Data*. Geneva, Switzerland: Author; 1996.
11. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E6: Good Clinical Practice; Consolidated Guideline*. Geneva, Switzerland: Author; 1996.
12. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E7: Studies in Support of Special Populations: Geriatrics*. Geneva, Switzerland: Author; 1996.
13. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E8: General Considerations for Clinical Trials*. Geneva, Switzerland: Author; 1996.
14. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E9: Statistical Principles for Clinical Trials*. Geneva, Switzerland: Author; 1996.
15. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E10: Choice of Control Group in Clinical Trials*. Geneva, Switzerland: Author; 1996.
16. International Conference on Harmonisation. *Guideline for Good Clinical Practice, M1: Standardization of Medical Terminology for Regulatory Purposes*. Geneva, Switzerland: Author; 1996.

## **Further Reading**

N/A

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.

---

# **Serious Adverse Event Data Reconciliation**

January 2008

## ***Abstract***

Because serious adverse event (SAE) data are typically stored in a safety database separate from the clinical trial data, a reconciliation of the two datasets must be carried out to ensure consistency. In covering the procedures for completing this task, this chapter discusses the importance of cooperating with safety representatives, and of creating proper documentation of discrepancies, missing data, reconciliation and other issues encountered during this process.

## ***Introduction***

Serious adverse event (SAE) data reconciliation involves the comparison of key safety data variables between two databases. Reconciliation is performed to ensure consistency between events residing in any SAE database and those residing in the clinical database. It is an iterative process that occurs several times during the study. When to reconcile is determined by the frequency of data receipt, scheduling of safety updates, and timing of interim and final reports.

## ***Scope***

This procedure applies to all projects where both a clinical database and a drug or device safety SAE database are maintained as two separate databases.

## ***Minimum Standards***

- Create entry and edit instructions, including deletion and change control procedures.

- Standardize the capture of SAE data elements in both the clinical database and the safety database.
- Conduct the reconciliation of event terms so they are at least similar if not exactly the same.

### ***Best Practices***

- Establish the time intervals in the project where reconciliation will be performed and in particular the mechanisms to cover interim analyses or safety data reporting. Often SAEs continue to be reported after a clinical trial has concluded. Some companies collect information in a single database and some companies collect information in two separate databases: a safety database and a clinical database. It is important to establish a cutoff point after which no SAEs will be added to the clinical database, even if the safety data or safety database is updated.
- Identify the data items to be reconciled. This may include, but not be limited to the following:
  - Protocol
  - Investigator
  - Subject identification
    - ◆ Randomization number
  - Initials
    - ◆ Date of Birth
    - ◆ Gender
    - ◆ Race
  - Event number
  - Diagnosis
  - Verbatim
  - Coded or preferred term



- Onset date
- Resolution date
- Date of death
- Outcome
- Severity
- Causality assessment
- Action taken with study drug

Sometimes data items are used from other modules for further reconciliation or clarification.

- From the demography module, items used may include but not be limited to the following:
  - Subject identification
  - Weight
  - Date of birth
  - Gender
  - Race
- From the discontinuation module, items used may include but not be limited to the following:
  - Subject identification
  - Primary reason for discontinuation being an event
  - Cause of hospitalization
  - Cause of death listed on the death certificate
  - Autopsy result

- From the concomitant medications module, items used may include but not be limited to the following:
  - Subject identification
  - Medication name
  - Start date
  - Stop date or ongoing
  - Indication
- When possible, customize database fields used in reconciliation to be programmatically compared without compromising the integrity of the software or databases. Even programmatic reconciliation of fewer than 100 events can be cost effective in both time and quality. The process can be validated once and run as frequently as data and time allow.
- When initiating the reconciliation process, clinical data management should confirm that all data to be included in the reconciliation have been entered and validated. Clinical data management should also confirm that any data clarifications have been returned and applied to the clinical database, and that the coding of AE verbatim terms against the common dictionary has been completed.
- Clinical data management, safety leads, and clinical operations should establish a mutually agreeable turnaround time for researching, retrieving, and correcting any discrepancies found during or since the last reconciliation period.
- Read–write access to either database (but not both) is granted to personnel trained in data entry for the purpose of and whose responsibilities include data entry, data modification, or data validation. Read–only access is granted to personnel related to reconciliation, but who are not directly responsible for those tasks related to data modification. System administration rights are limited to personnel responsible for database configuration.

## **Procedures**

- Some companies maintain two databases: a safety database and a clinical database. Conversely, some companies collect all information in a single database. When two databases are used, obtain the SAE information to be reconciled from both the safety and the clinical databases.
  - Listings are produced from either the safety database or the data management database, and the two databases are manually reconciled through direct comparison of these listings. However, in some instances the two databases can be compared programmatically and a listing of differences provided. Either way, the differences will require manual review by trained staff. Ancillary documents can also be used for clarification or corroboration, such as hospitalization discharge summaries, death certificates, or autopsy reports.
- Verify that all SAEs from the clinical database also reside in the drug safety database. Note that some SAEs from the safety database may not be in the clinical database until all CRFs are collected and entered.
- Document all SAEs included in the clinical database but not included in the safety database. These are potentially unreported events. Include copies of the appropriate CRFs to be forwarded to the safety contact person.
- Research all SAEs in the safety database that are not found in the clinical database.
  - If the visit has been monitored, collected, and entered by CDM, the site should be queried to request the original missing event page. Do not add SAEs to the clinical database without the data for that visit having been monitored against source documents according to the study's clinical monitoring guidelines. Only those updates signed and dated by site staff after the CRF page has been monitored and retrieved are acceptable for updating the clinical database.
- Research and resolve all differences between SAEs that are present in both databases.

- Depending on the nature of discrepancies, it may be necessary to seek input from the medical monitor or designee before deciding on a course of action.
- Some discrepancies are acceptable. For example, slight variations in terminology used in describing events may be of no consequence. Also, start dates may differ, as an event may start as nonserious before progressing to serious.
- Site-authorized updates to CRFs received by clinical data management are copied to drug safety for assessment and, if appropriate, for inclusion in the safety database. Clinical data management generates queries to clarify discrepancies, and forwards them to the sites for resolution. Resolved queries from the site are returned through data management, to be used to update either or both databases by their respective staff. Communication of these updates can be facilitated by use of a standard template, such as the Sample SAE Data Reconciliation Form provided in Appendix A of this chapter.
- Prior to data lock, verify that all queries have been correctly returned and integrated into the database. A quality control process should be in place to ensure this is done accurately and consistently. Ensure that all expected SAE information has been received and reconciliation has been performed on all events. Written notification should be made when reconciliation has been successfully completed. This helps avoid confusion should the safety database be held open for updates after the study ends.
- Any final inconsistencies that cannot be resolved should be documented in a CDM Data Handling Report or the equivalent.

### ***Recommended Standard Operating Procedures***

- Safety Database Setup, Management, and Validation
- Serious Adverse Event Reconciliation Work Instruction
- Coding of Clinical Data

**References**

N/A

**Further Reading**

N/A

**Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
January 2002	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
January 2008	Reviewed and content updated.

**Appendix A: Sample SAE Data Reconciliation Form**

<p><b>Sponsor Name</b>  <b>Protocol Number</b>  <b>Project Work Code</b>  <b>Reconciliation Session Date</b></p>
--

Investigator/ Patient Number	Field Name	Description of Inconsistency		Description of Resolution/Action Required	Resolved By (initial and date)
		Drug Safety Database	Clinical Database		

---

# Database Closure

October 2013

## *Abstract*

Study databases must have access removed and be properly closed to ensure data integrity for the generation of results, analyses and submissions. This chapter recommends processes, checklists, and essential documentation for locking and closing study databases. Reopening a locked database to evaluate and correct errors is discussed, with an examination of important considerations and decisions that should be made, procedures that should be followed, and documentation that must be produced.

## *Introduction*

At the culmination of a clinical study, database locking and closing procedures are imperative to prevent inadvertent or unauthorized data changes prior to analysis and reporting. For blinded studies, database closure procedures must also ensure blinding is not broken prematurely or by unauthorized personnel. Because data integrity and blinding are crucial to the success of a study, all clinical studies must have well-defined and documented processes for locking, unlocking, and closing study databases.

Terminology varies among different organizations, database systems, and clinical data management systems (CDMS), in particular regarding terms such as “interim lock,” “soft lock,” “final lock,” “database freeze,” and “database closure.” For the purposes of this chapter, the following definitions specify the meaning of some of the terms used in this chapter. These terms may be interpreted differently within different organizations.

- *Interim lock* refers to processes used to take a “snapshot” of a database at a particular point in time while the study is still in progress.

- *Soft lock* refers to processes during which access to the database is limited while CDM personnel confirm the suitability of the data for final analysis.
- *Final lock* (also known as a “freeze” or “hard lock”) refers to processes used to remove access to the database to ensure no further changes to data can be made. The final lock is a key process of database closure, but not the only process needed to close a database properly.
- *Database closure* refers to processes used to finalize the database as a final study deliverable.

These terms are not defined universally within the clinical research industry, so clinical data management (CDM) personnel should ensure they clearly understand how these terms are used or how they relate to terminologies used within their specific organizations.

## **Scope**

This chapter describes distinctions between interim lock, soft lock, final lock, and database closure procedures, as well as discussing considerations for unlocking and relocking a database. The importance of a thorough database closure checklist is discussed, and a sample database closure checklist is provided.

Although some of the specific topics addressed by this chapter may not be the direct responsibility of CDM personnel, data managers must have an ongoing awareness of requirements and ensure these tasks have been completed in accordance with the principles and standards of their organization, regulatory bodies, and good clinical practice.

## **Minimum Standards**

- Establish clearly documented procedures defining all steps of database lock, database closure, and unlocking and relocking the database after database closure.
- Clearly define all roles with respective responsibilities involved with database lock and closure procedures.



- Prior to database lock (interim, soft, or final) or database closure, ensure documentation of all defined tasks or criteria has been completed.
- At final database lock, ensure all team members are notified and access that allows database changes is removed and documented.

### ***Best Practices***

- Clearly define all terms relating to interim lock, soft lock, final lock, database unlock and final database closure.
- Develop and utilize a database closure checklist.
- Maintain documentation and approval or acknowledgement documents requiring signatures of all responsible parties involved in database lock, unlock and closure procedures.
- Where indicated, plan an interim or soft lock with a statistical analysis and data review prior to the final lock. This review may identify potential data errors, preventing the need to unlock the database after the final lock.

### ***Interim and Final Lock Distinctions***

Interim locks create a snapshot of a database at a particular point in time, and are typically performed to facilitate statistical analyses, listings, and reports prior to completion of the study. During the interim lock, access to the data to be analyzed may be restricted to certain personnel for a set period of time or until specific criteria are met. This restriction of access minimizes the possibility of inadvertent or unauthorized changes being made to the data. Because a study is often still in progress during an interim lock, the database will frequently have unresolved queries. Interim locks are often associated with some degree of data-cleaning activities, although data may not be cleaned as thoroughly as with a final lock. For example, data cleaning activities for an interim lock may be limited to critical data affecting safety and efficacy.

A final lock is only performed at the end of a study and is one of the initial steps toward complete study closeout. Once final lock has occurred, no changes to data can be made, and access to the database is typically restricted to only those personnel responsible for closing, delivering, and archiving the

database. Multiple interim locks may be performed in preparation for the final lock, allowing new or updated information that needs to be incorporated into the database prior to final lock. Before the final lock is initiated, all data queries are resolved, all medical coding is complete and approved, all external data reconciliation is completed and approved, and all data should be cleaned thoroughly. Some organizations may perform a soft lock prior to the final lock to ensure all activities required for the final lock have been adequately performed. Should it become necessary to unlock the database after the final lock has been initiated, database unlocking criteria and procedures typically become much more stringent than those for a soft lock or interim lock.

### ***Minimum Requirements for Database Lock***

Prior to database lock, the data manager should make certain the following tasks have been completed in accordance with established procedures and quality standards.

- For any database lock, identify and document the personnel responsible for performing various tasks for the database lock at the start of the project. Any subsequent changes in these roles or responsibilities must be documented. For more information on responsibility matrices see the GCDMP chapter entitled “Project Management for the Clinical Data Manager.”
- Notify relevant stakeholders and study team members in advance when an interim or final lock is to occur, then follow up with a notification when the lock occurs.
- For a final lock, ensure all applicable data have been received, processed and source verified (paper studies) or electronically signed by the principal investigator(s) (EDC studies). For interim locks, the degree of data processing and source verification may not be held to the same level as for a final lock.
- For a final lock, ensure all queries have been resolved. For interim locks, all queries may not need to be addressed, although all efforts should be made to resolve any critical queries potentially impacting interim analyses.

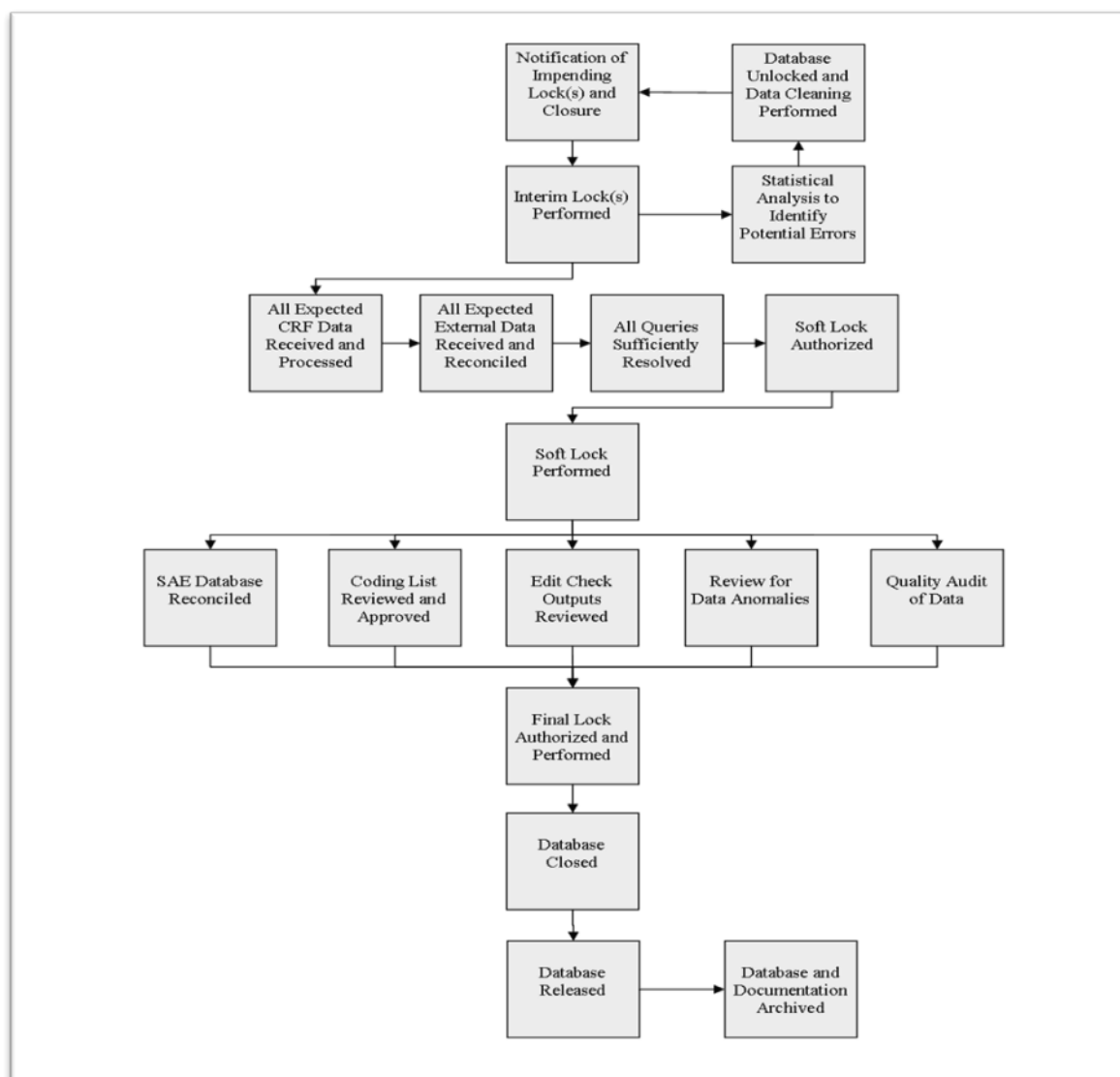
- For a final lock, ensure all expected external data (e.g., electronic laboratory data, IVRS, ePRO, ECG, etc.) are received, complete and reconciled with the study database. Interim locks may be performed with incomplete external data, but an effort should be made to reconcile external data that have been received.
- If a separate database exists for serious adverse events (SAEs), it should be reconciled with the main study database.
- The coding list (i.e., for adverse events, concomitant medications, etc.) should be reviewed for completeness and consistency, and should be approved by a medical monitor or safety surveillance team, if applicable.
- Ensure a final review of logic and consistency check output (edit checks) and data listings/patient profiles has been performed.
- Ensure a final review for obvious data anomalies has been performed.
- Perform a quality audit of the data and document the corresponding error rate. In many organizations, this audit may be performed by a separate quality assurance department. For more information on data quality, see the GCDMP chapter entitled “Assuring Data Quality.”
- Every database lock should follow a documented approval process with approval or acknowledgement by relevant study personnel (e.g., data manager, biostatisticians, monitoring representative, clinical/scientific representative). The ability to edit the database (“edit access”) should only be removed after the necessary approvals have been obtained, and the date of edit access removal should be documented.
- For any database lock, all documentation should be updated and stored according to standard operation procedures.

### ***Database Closure Processes***

Before a database can be closed, the final database lock must be completed and documented, including documentation that all edit access was removed at a defin-

itive point in time. Before the final lock is performed in preparation for database closure, clearly defined procedures must be followed to ensure all data have been received and processed, all data meet an acceptable quality level, and all relevant study personnel and stakeholders have been notified or have approved the final database lock. In addition to the tasks performed for a final database lock all documentation for database closure should be updated and stored according to standard operating procedures. Figure 1 shows a sample workflow of the main process steps of database closure.

**Figure 1: Sample Database Closure Steps**



## Database Closure Checklist

Every study should have a database closure checklist that should be closely followed to ensure all required tasks have been performed. Each task on the checklist should indicate who is responsible for the task and be accompanied by the date of completion and the signature or initials of the individual responsible for confirming the task was satisfactorily completed. For a sample database closure checklist, see Figure 2.

**Figure 2: Sample Database Closure Checklist**

<b>Task</b>	<b>Role Responsible</b>	<b>Date Completed</b>	<b>Signature</b>
Study team and stakeholders notified of database lock/closure.			
All applicable CRFs/data received, entered, and source verified.			
All external data received and reconciled.			
All queries sufficiently resolved.			
Final database lock approved and performed.			
SAE database reconciled.			
Coding list reviewed and approved.			
Final review of edit check outputs performed.			
Final review of data listings and patient profiles performed.			
Final statistical review for data anomalies performed.			
Quality audit of data completed.			
Database release authorized.			
Database release completed.			
Database archived.			
All required database lock and closure documents present.			

Although some elements of a database closure checklist should be relatively universal, other elements may be specific to the organization, study, CDMS, or database. In creating the checklist, the following documents and personnel should be consulted to determine the applicable items to include on the checklist.

- The organization's standard operating procedures
- Software platform requirements of the CDMS or database
- Protocol requirements
- Data management plan requirements
- Input from clinical, biostatistics, programming, and quality assurance personnel

### ***Database Lock and Closure Documentation***

All database locks and closures should be clearly documented, including all of the steps that go into the lock or closure. This documentation should include the reasons for any interim locks, the signed database closure checklist, and documentation of all tasks that comprise the lock or closure (e.g., final data receipt, SAE database reconciliation documentation, etc.). For most individual tasks that are part of the database lock or closure, documentation should include the date of completion and the signature or initials of the study team member(s) responsible for the task. Some organizations may archive the results of data quality audits with database closure documentation, but other organizations document data quality results separately.

### ***Final Database Release***

After a database has been closed with appropriately signed authorizations and documentation, the database is ready for release. Details of database release vary widely between organizations, particularly between CROs and sponsors. For example, a CRO may release the database directly to the sponsor, whereas a sponsor might release the database to a separate department within the organization. Regardless of the entity to which the database is to be released, the release should only be performed with signed authorization or acknowledgement.

If a study is blinded, release of the database is typically necessary prior to unblinding of the data. Although the data must be unblinded prior to statistical analyses, unblinding the data too soon can pose risks to the overall success of the study. Because of the magnitude of risk posed by premature unblinding, documented authorization for final database release is arguably even more crucial for blinded studies. Documented authorization should also precede unblinding of the data. In some situations, data is unblinded within the CDMS prior to final database lock and release. However, in these situations documented authorization for unblinding is still a crucial step.

### ***EDC and Paper-Based Distinctions***

Although there are differences in database lock and closure procedures between EDC and paper-based studies, these distinctions are often specific to the EDC systems that are used. For EDC studies, CDM should ensure that Principal Investigator(s)' electronic signatures are compliant with 21 CFR Part 11, which details the FDA's requirements for electronic signatures.<sup>1</sup> Source document verification procedures are also quite different between EDC and paper-based studies. In EDC systems, access to the database may be restricted to "read only" at the lock stages followed by complete removal of access occurring after database closure. For more information about distinctions between EDC and paper-based studies, see the GCDMP chapters entitled "Electronic Data Capture—Concepts and Study Start-up," "Electronic Data Capture—Study Conduct," and "Electronic Data Capture—Study Closeout."

### ***Database Unlocking/Relocking***

Ideally, a final lock should be truly final, and subsequent unlocking is discouraged. However, certain circumstances may necessitate temporarily unlocking the database after final lock has occurred. Before a request to unlock a database is authorized, appropriate management at the organization should carefully review the reasons provided to justify unlocking the database.<sup>2</sup>

Whenever there is a need to unlock a previously locked database, documented authorization should be received before unlocking commences. A thorough evaluation should be made before deciding to unlock a database. Well-defined

policies and procedures should be followed to ensure unlocking restores access to authorized personnel only. Although discouraged, any data changes made while the database is locked must be captured within the audit trail.

Despite being discouraged, unlocking a database after final lock occurs frequently enough that organizations should have policies in place to manage the details of unlocking. These policies should clearly define the reasons for unlocking a database after final lock, procedures to be followed when unlocking or relocking, and procedures to be followed for the period between unlocking and relocking. Procedures should include notification of the study team, a clear definition of the change(s) being made, the date changes are implemented, the specific individuals who will regain access to the database, and steps to ensure that only the authorized changes and no others have been made. Relocking the database should follow the same or similar processes for notification and approval as the initial lock.

Note that not all errors found after final lock must be corrected in the database itself. Most large clinical databases are not completely free of errors. The database should only be unlocked due to data errors relating to safety or efficacy parameters that may impact the statistical analysis and conclusions drawn from the study. Some errors may be documented in the statistical or clinical report. Some organizations may maintain a database error log that lists each error, the action taken, the reason why the database was not unlocked to correct the error, and who made the decision to not unlock the database. Regardless of the specific strategy employed by an organization, organizations should implement predefined processes to determine how such errors will be handled and documented.

### ***Recommended Standard Operating Procedures***

- Interim Database Lock
- Final Database Lock
- Database Unlock/Relock
- Database Closure
- Change Control/Errors



## References

17. *Code of Federal Regulation, Title 21, Volume 1, Part 11: Electronic Records, Electronic Signatures*. Washington, DC: Government Printing Office; 1998.
18. Prokscha S. *Practical Guide to Clinical Data Management*. New York, NY: CRC Press, 2007:75-88.

## Further Reading

Terms used in this chapter may be found in the *Good Clinical Data Management Practices* Glossary.

## Chapter Revision History

Publication Date	Comments
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
October 2013	Revised for content, style, grammar, and clarity with database closure sample flowchart and sample checklist added.

**This page is intentionally blank.**

---

# Clinical Data Archiving

June 2008

## *Abstract*

In order to meet the requirements of industry guidelines and regulations, clinical data managers must ensure that data captured during a clinical trial are retained correctly. This chapter provides an overview of the regulations that must be followed and discusses approaches to satisfying the requirements. Consideration is given to proper handling of electronic data that are collected in a clinical trial. The components that constitute a clinical data archive are reviewed, and technical requirements for the correct use of open electronic data formats, such as XML (Extensible Markup Language) and SAS<sup>®</sup>, are discussed with an emphasis on ensuring long-term accessibility.

## *Introduction*

Clinical data archiving includes planning, implementing and maintaining a repository of documents and/or electronic records containing clinical information, supporting documentation, and any interpretations from a clinical trial.

## *Scope*

This section provides an outline to help clinical data managers develop an archiving strategy, working in conjunction with the project team and/or other appropriate department(s). Included are details of the regulatory requirements surrounding clinical data archives, a description of the components of an archive and information about data formats that can be used to support an archive. This document focuses on the components of the study archive that are the responsibility of data management.

## **Minimum Standards**

- The clinical data archive should include a centralized table of contents.
- Accessibility of the clinical data archive electronic records should be tested following every major upgrade of the active clinical data management system.
- For paper case report form (CRF) studies, the original signed, dated, and completed CRF and original documentation of CRF corrections should be kept in the sponsor's files or offsite archive facility.
- The clinical data archive should be retrievable within a reasonable timeframe.
- For each study, the documentation should identify the hardware and software used, as well as specific version of the software or hardware.

## **Best Practices**

- All clinical data, metadata, administrative data, and reference data should be maintained in an industry standard, open system format, such as CDISC's Operational Data Model (ODM).
- An electronic repository should link all study components, including the clinical data, CRF images in Portable Document Format (PDF) format, program files, validation records, and regulatory documentation.
- The audit trail should be stored in open format files in a secure system location.
- Copies of all user and system documentation for any applications used to collect or manage clinical data should be retained in the corporate library or archive facility.
- Reports describing the metadata and validation of study metadata, including data structures, edit check descriptions, and electronic data-loading specifications should be stored in the clinical data archive.
- System security reports, including user listings, access rights and the dates of authorization, should be printed and filed or scanned.

- The archive should include all program code for edit checks, functions, and sub-procedures, together with a copy of the version control information and validation documentation.
- Paper CRFs should be scanned and indexed. If an electronic data capture (EDC) system is used, all entry screens should be archived in an easily accessible format, such as a PDF file.
- Address archive responsibility for external data management providers. The sponsor should ensure that any signed contract with a vendor includes a section on archiving.

## **Background**

The International Conference on Harmonisation Good Clinical Practice<sup>1</sup> (ICH GCP) requirements stipulate that data collected in a clinical trial must be maintained for a period of two years, following either the last regulatory submission or a decision to discontinue development of a compound, biologic, or medical device. To meet this requirement, as well as to ensure that the sponsor is able to answer questions related to clinical trial data that may emerge many years after the trial is conducted, it is important to archive clinical data, as well as the accompanying trial processing documentation.

Historically, the most common mechanism for long-term clinical data storage has been to extract the final data from the clinical data management system into SAS<sup>®</sup> datasets. The extracted SAS<sup>®</sup> datasets are still an important component of the clinical data archive; however, with the increasing importance of electronic regulatory submissions in recent years, requirements for clinical data archives are changing. As a result, clinical records that are part of an electronic submission must now comply with the 21 Code of Federal Regulations<sup>2</sup> (CFR) Part 11 ruling, which was originally published in 1997. Part 11 defines specific requirements with respect to authentication and auditing of electronic records. In addition, the Food and Drug Administration's (FDA) *Guidance for Industry: Computerized Systems Used in Clinical Investigations*<sup>3,4</sup> defines requirements for data archiving. This guidance was published in 1999 and updated in 2007. To fully meet the requirements of these regulations and guidelines, a comprehensive archiving strategy is needed.

## **Regulations and Guidance**

The tenets of 21 CFR Part 11 include no specific requirements for data retention or data archiving capabilities. However, the FDA has made it clear that the intent of the guidance is to supplement the predicate rules and ICH GCP requirements for those cases where electronic records are either directly or indirectly part of an electronic submission.

Guidance documents with specific mention of archive and record retention requirements include:

- *Guidance for Industry: Computerized Systems Used in Clinical Investigations*<sup>3,4</sup> (CSUCI) published by the FDA in 1999 and updated in May 2007. This document describes requirements surrounding the need to preserve the systems environment in which electronic records are captured and managed.
- *ICH Good Clinical Practice*<sup>1</sup> (Section 5 Sponsor requirements) provides information about record retention requirements.

Regulatory guidance is being actively developed in the area of electronic records handling. Before finalizing your clinical data archive design, it is necessary to consult with the regulatory affairs specialists within your organization to ensure your design approach is consistent with the organization's regulatory policies.

## **Archive Contents**

To successfully reconstruct a clinical trial, an auditor must be able to view not only the clinical data, but also the manner in which the data are obtained and managed. A summary of the types of information that should be included in a clinical data archive is provided in Table 1.

**Table 1**

<b>Archive Component</b>	<b>Requirement</b>
Clinical data	All data collected in the trial. This includes both CRF data and data that is collected externally (e.g., electronically submitted laboratory or patient diary data).
External data*	For data that are collected externally and loaded into a Clinical Data Management System (CDMS), the archive should include all loaded files, loading documentation, and quality control documentation.
Database Metadata	Information about the structure of clinical data, such as an annotated CRF. The annotated CRF will document the tables, variable item names, forms, visits and any other objects. It also includes codelists. This should also contain images of the entry screens (provided in PDF).
Coding Dictionaries	If data have been auto-encoded using a company dictionary or synonym table, a copy of the appropriate dictionary version should be included.
Laboratory Reference Ranges	If more than one version exists for reference ranges used during the course of the trial, each version should be retained in the archive. Documentation of the handling and processing of the laboratory ranges should also be present.
Audit trail	It is essential that the entire study's audit trail be included in the archive in a tamper-proof format.
Listings of edit checks, derived data, change controls	Edit check definitions and derived data change controls may be provided either as program listing files or as a report from the study definition application.
Discrepancy management logs, data handling guidelines	Listings of records that failed edit checks together with information on how the discrepancies were managed during the course of the study should be maintained.
Queries	Retain copies of all queries, query correspondence and query resolutions. Paper queries may be scanned and indexed.
Program code	Program code from data quality checking programs, data derivations and statistical analyses performed with clinical data and program documentation should be stored. Ideally, these documents should be stored online and indexed or hyperlinked.
CRF images in PDF format	For paper-based trials, CRF images are typically obtained by scanning the forms and converting them to PDF format. For trials using EDC, PDF images of the electronic forms may be created by the EDC application.
Data Management Plan (DMP)	PDF or paper version of MS Word and Power Point documents containing the study data management plan. The DMP may include sections or documents listed above.
Study Validation	Contents are described in the GCDMP chapter on systems

<b>Archive Component</b>	<b>Requirement</b>
Documentation	validation. This document may be in paper or electronic form.
Clinical Documents/Memos	Maintain copies of quality control documentation, database lock/freeze, SAE reconciliation, Personnel listing documents, etc.

\*For data managed externally and then loaded into an in-house system for reconciliation, reviews, or other purposes, it is generally sufficient to limit the archive to actual data and any information pertaining to how the data are managed internally. When using an external vendor, the vendor is responsible for archiving any records that reflect how the data are managed in the vendor's system. The trial sponsor is ultimately responsible for ensuring that any vendor who provides trial data works in accordance with regulatory requirements. Therefore, the sponsor should ensure that any signed contract with a vendor includes a section on archiving. The information in this section should comply with both sponsor and regulatory requirements.

## ***Technical Requirements***

Designing a clinical data archive for long-term accessibility presents a challenge in the face of proprietary applications, tools, and platforms. This design should include input from all team members to ensure that the archive will meet department, corporate and regulatory requirements. A well-designed clinical study archive can facilitate compliance with the long-term data access requirements of the regulations for both paper based and electronic clinical trials. For this reason, the ideal clinical data archive should be based on standards and open systems.

The open formats that are typically used for clinical study archives are described in Table 2. No single format is ideal in all circumstances. Due to the fact that a study archive will usually include many different types of information, it will most likely include multiple formats. The format chosen for each type of information should be based on the likely future use of the information.



**Table 2**

<b>Format</b>	<b>Description</b>	<b>Pro</b>	<b>Cons</b>
Comma Separated Values (CSV)	Plain ASCII text with commas used as field delimiters. CSV files can be edited with text editors, word processors, and spreadsheet programs such as Microsoft <sup>®</sup> Excel.	Conceptually straightforward readily imported into almost any database.	Requires separate handling of metadata, administrative data and audit trails.
XML	Extensible Markup Language. Vendor independent, ASCII based technology for transfer of structured information between dissimilar systems. Used as the basis for the CDISC Operational Data Model.	Open standard ideally suited for clinical trial data. XML can include structural metadata, administrative data, and clinical data within a single file.	Still unfamiliar to many data managers and IT staff.
SAS <sup>®</sup> Transport files	Open source format provided by SAS <sup>®</sup> Institute Inc. Commonly used for submitting clinical data to the FDA. Can be read by the SAS Viewer that is distributed free of charge on the SAS Web site.	Familiar to clinical data managers and regulators. Works well with SAS data analysis tools.	Proprietary format. Variable naming restrictions. Requires separate handling of metadata, audit trails, and administrative data.
Adobe <sup>®</sup> PDF	Product provided by Adobe Systems Incorporated. Widely used standard for transmission of text documents. Default format for transmission of information to the FDA. Can be read by the Acrobat Reader, which is available free of charge from the Adobe <sup>®</sup> Web site.	Many applications output PDF files as an optional output format. Reader is available free of charge.	Predefined PDF output from EDC applications may not comply with or have the flexibility to produce standardized Sponsor formats.

Long-term data access requirements suggest that the choice of data format is limited to ASCII based formats, or formats based on an open standard, such as SAS<sup>®</sup> Transport files. The choice may be further influenced by the format used in the original data management or data collection system.

## **Archives for Clinical Sites**

The CFR predicate rules and the ICH Good Clinical Practice (GCP) guidelines specify that a copy of clinical data must be retained at the investigator site throughout the records retention period. For paper based studies, this can be achieved by keeping a copy of the paper records at the site. For EDC studies it is important to have a strategy in place for ensuring that these guidelines are met appropriately. Many EDC vendors will provide PDF files for all of the electronic Case Report Forms (eCRFs) collected from the site during the trial. The Clinical Data Manager (CDM) may provide assistance and/or coordination with this procedure. If your company builds EDC studies in-house, the data manager will be responsible for ensuring the quality of the PDF outputs prior to sending the files back to the clinical sites.

## **Recommended Standard Operating Procedures**

- Study Archiving Procedures
- Document Archiving Procedures

## **References**

1. International Conference on Harmonisation. Good Clinical Practice. *Federal Register*: 1997.
2. *Code of Federal Regulations*, Title 21, Volume 1, Part 11. Food and Drug Administration. US Government Printing Office, 1998.
3. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; April 1999.
4. *Guidance for Industry: Computerized Systems Used in Clinical Investigations*. Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; May 2007.

## **Further Reading**

*The Data Archiving Challenge*. Clinical Data Interchange Standards Consortium; May 2000.

Paul Bleicher. Diamonds May Be Forever, But Data? *Applied Clinical Trials*; February 2002

CDISC (Clinical Data Interchange Standards Consortium) Web site, Available at: <http://www.cdisc.org/>. Accessed on November 28, 2007.

## **Chapter Revision History**

<b>Publication Date</b>	<b>Comments</b>
September 2003	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
June 2008	Revised for content, style, grammar, and clarity.

**This page is intentionally blank.**

---

## Glossary

October 2013

The *Good Clinical Data Management Practices* adopt the ICH definitions for terms defined within the ICH guidelines. Unless otherwise noted, these definitions were taken from *ICH E6*.<sup>1</sup>

(ASQ) in a definition indicates the American Society for Quality as a source.

### A

#### **access control**

Policy and procedure that defines accessibility to a physical space or electronic source of information. The policy usually includes the concept of audit trails, either paper (e.g., signature log) or electronic.

#### **adverse drug reaction (ADR)**

In the pre-approval clinical experience with a new medicinal product or with its new usage (particularly as the therapeutic dose[s] may not be established), all noxious and unintended responses to a medicinal product related to any dose should be considered adverse drug reactions. The phrase “responses to a medicinal product” means that a causal relationship between a medicinal product and an adverse event is at least a reasonable possibility (i.e., the relationship cannot be ruled out). Regarding marketed medicinal products, an ADR is a response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of diseases or for modification of physiological function (see *ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting*<sup>2</sup>).

**adverse event (AE)**

In a subject or clinical-investigation subject administered a pharmaceutical product, any untoward medical occurrence which does not necessarily have a causal relationship with the treatment. An adverse event (AE) can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product (see the *ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting*<sup>2</sup>).

**amendment (to the protocol)**

See **protocol amendment**.

**analysis dataset**

The final data set, including derived items and excluding redundant data points, which is used to perform the analyses required for safety assessment, efficacy assessment, submission to regulatory authorities, or other review. Can be comprised of one or more data files.

**analysis file**

Same as analysis dataset in the context of the GCDMP.

**annotated crf**

A document that maps the names of the collected items to their corresponding database tables, variable item names, forms, visits and any other objects needed for a person to correctly analyze data collected in a clinical trial. Annotated collection documents are required so that any person can understand where variables for analysis originate.

**applicable regulatory requirement(s)**

Any law(s) and regulation(s) addressing the conduct of clinical trials of investigational products.

**Application Service provider (ASP)**

An application service provider is a vendor who provides, manages and distributes software-based services to customers over a network.

**approval (in relation to institutional review boards)**

The affirmative decision of the institutional review board (IRB) that the clinical trial has been reviewed and may be conducted at the institution site within the constraints set forth by the IRB, the institution, Good Clinical Practice (GCP), and the applicable regulatory requirements.

**audit**

A systematic and independent examination of trial-related activities and documents to determine whether the trial-related activities being evaluated were conducted and the data were recorded, analyzed and accurately reported according to the protocol, the sponsor's standard operating procedures (SOPs), GCP, and the applicable regulatory requirement(s).

**audit certificate**

A declaration of confirmation by the auditor that an audit has taken place.

**audit report**

A written evaluation by the sponsor's auditor of the results of the audit.

**audit trail**

Documentation that allows reconstruction of the course of events.

## **B**

### **batch job**

A series of processes run in an electronic system that perform specific tasks, such as data validation, query generation, external data upload, or lab reference range normalization.

### **biologics**

A biological product (as a vaccine or blood serum) used in medicine

### **blinding/masking**

A procedure in which one or more parties to the trial is kept unaware of the treatment assignment(s). Single-blinding usually refers to the subject(s) being unaware, and double-blinding usually refers to the subject(s), investigator(s), monitor, and, in some cases, data analyst(s) being unaware of the treatment assignment(s).

## **C**

### **case report form (CRF)**

A printed, optical, or electronic document designed to record all of the protocol-required information to be reported to the sponsor on each trial subject.

### **CDISC**

Acronym for the **Clinical Data Interchange Standards Consortium**.

### **central lab**

A vendor contracted for a clinical trial that processes samples collected from subjects and provides the results of laboratory tests or other medical analyses



(e.g., ECG results, pathology results) to the sponsor. Refer to the Laboratory Data Handling chapter.

### **change control**

A procedure that defines how planned changes to any part of a computer system are handled in a manner as to maintain compliance with required functionality of that system. The procedure ensures that changes applied to the system do not unexpectedly impact the functionality of the system in question, or any other computer systems. The procedure should also define how unexpected changes to a system are prevented and managed.

### **checklist**

(ASQ) A tool used to ensure that all important steps or actions in an operation have been taken. Checklists contain items that are important or relevant to an issue or situation. Checklists are often confused with check sheets and data sheets.

### **CLIA**

See Clinical Laboratory Improvement Amendments.

### **Clinical Laboratory Improvement Amendments (CLIA)**

Congress passed the Clinical Laboratory Improvement Amendments (CLIA) in 1988 establishing quality standards for all laboratory testing to ensure the accuracy, reliability and timeliness of patient test results regardless of where the test was performed. See [www.fda.gov/medicaldevices/deviceregulationandguidance/](http://www.fda.gov/medicaldevices/deviceregulationandguidance/) for more information.

### **clinical trial/study**

Any investigation using human subjects that is intended to discover or verify the clinical, pharmacological, and/or other pharmacodynamic effects of an investigational product(s); and/or to identify any adverse reactions to an investigational product(s); and/or to study absorption, distribution,

metabolism, and excretion of an investigational product(s) for the purpose of ascertaining its safety and/or efficacy. The terms “clinical trial” and “clinical study” are synonymous.

**clinical trial/study report**

A written description of a trial/study of any therapeutic, prophylactic, or diagnostic agent conducted in human subjects, in which the clinical and statistical description, presentations, and analyses are fully integrated into a single report (see the *ICH Guideline for Structure and Content of Clinical Study Reports*<sup>3</sup>).

**code libraries**

A repository of validated programming logic that can be used during the programming of edit checks or other programs used in the collection, review, or analysis of clinical trial data.

**common causes**

(ASQ) Causes of variation that are inherent in a process over time. They affect every outcome of the process and everyone working in the process. See also **special causes**.

**comparator (product)**

An investigational or marketed product (i.e., active control) or placebo used as a reference in a clinical trial.

**compliance (in relation to trials)**

Adherence to all the trial-related requirements, GCP requirements, and the applicable regulatory requirements.

**composite endpoint**

Overall outcome that the protocol is designed to evaluate based on more than one common endpoint such as myocardial infarction plus repeat intervention.

**compound**

A chemical molecule with potential pharmacological activity.

**confidentiality**

Prevention of disclosure of a sponsor's proprietary information or of a subject's identity to unauthorized individuals.

**conformance**

(ASQ) An affirmative indication or judgment that a product or service has met the requirements of a relevant specification, contract, or regulation.

**contract**

A written, dated, and signed agreement that sets out any arrangements on delegation and distribution of tasks and obligations and, if appropriate, on financial matters between two or more involved parties. The protocol may serve as the basis of a contract.

**coordinating committee**

A committee that a sponsor may organize to coordinate the conduct of a multi-center trial.

**coordinating investigator**

An investigator assigned responsibility for the coordination of investigators at different centers that are participating in a multi-center trial.

**contract research organization (CRO)**

A person or an organization (e.g., commercial, academic, or otherwise) contracted by the sponsor to perform one or more of a sponsor's trial-related duties and functions.

**control chart**

(ASQ) A chart with upper and lower control limits on which values of some statistical measure for a series of samples or subgroups are plotted. The chart frequently shows a central line to help detect a trend of plotted values toward either control limit.

**corrective action (CA)**

(ASQ) The implementation of solutions that lead to the reduction or elimination of an identified problem.

**CS**

Clinically Significant.

**D**

**data cleaning**

The process of collecting, reviewing, and confirming modifications to clinical data in such a way that data provided for statistical analysis is complete, accurate, and consistent with other data points.

**data module**

A category of a type of data, such as CRF.

**database backup**

A duplicate copy of all electronic data and metadata that can be retrieved in the event of system failure or data corruption.

**database lock**

The closing of a database after all clinical trial data has been reviewed, queries resolved and issues addressed, such that the database cannot be altered in any way.

**development/test environment**

Computer system instances that are used for study build and test, prior to release to the production instance. Defined quality procedures and documentation allow transition of programming code from one instance to another.

**device**

I. A means of data collection such as a paper CRF, Personal Digital Assistant, or medical instrumentation. II. An instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is: recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them, intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or intended to affect the structure or any function of the body of man or other animals, and which does not achieve any of its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes.

**direct access**

Permission to examine, analyze, verify, and reproduce any records and reports that are important to evaluation of a clinical trial. Any party (e.g., domestic and foreign regulatory authorities, sponsor's monitors and auditors) with

direct access should take all reasonable precautions within the constraints of the applicable regulatory requirement(s) to maintain the confidentiality of subjects' identities and sponsor's proprietary information.

### **disaster recovery plan**

A disaster recovery plan is a comprehensive statement of consistent actions to be taken before, during and after a disaster. The plan should be documented and tested to ensure the continuity of operations and availability of critical resources in the event of a disaster. (www.drj.com)

### **discrepancy**

Inconsistency in two or more data points collected in a clinical trial that must be addressed prior to database lock.

### **documentation**

All records, in any form (including, but not limited to, written, electronic, magnetic, and optical records and scans, x-rays, and electrocardiograms) that describe or record the methods, conduct, or results of a trial; the factors affecting a trial; and the actions taken.

### **double data entry**

The process of purposely entering clinical trial data twice for studies with paper collection media. The two entries are done independently. The goal is to ensure entry into the electronic system is completed without transcription errors.

## **E**

### **e-CRF**

Acronym for **electronic case report form**. An auditable electronic record designed to record information to be reported to the sponsor on each trial

subject, as required by the clinical trial protocol. See also **case report form**.

### **edits - hard and soft edit**

Programmed or manual verifications performed on a clinical database for the purpose of ensuring a quality final analysis set for analysis. Hard edits refer to verifications that require a data change or entry in order to resolve it while Soft edits also accept a confirmation of the existing data.

### **EHR**

Electronic Health Record.

### **electronic record**

Electronic record means any combination of text, graphics, data, audio, pictorial, or other information representation in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer system.

### **electronic signature**

Electronic signature means a computer data compilation of any symbol or series of symbols executed, adopted, or authorized by an individual to be the legally binding equivalent of the individual's handwritten signature.

### **electronic submission**

The set of required documents for a submission, rendered in an acceptable electronic format that is transmitted to a regulatory agency in lieu of paper documents for review and approval.

### **EMR**

Electronic Medical Record.

**endpoint**

Overall outcome that the protocol is designed to evaluate. Common endpoints are severe toxicity, disease progression, or death.

**essential documents**

Documents which individually and collectively permit evaluation of the conduct of a study and the quality of the data produced (see *ICH E6*, Section 8. “Essential Documents for the Conduct of a Clinical Trial”<sup>1</sup>).

**EU**

European Union.

**exposure**

The condition of being subject to some effect or influence; in context of a clinical trial this generally refers to exposure to the test article/drug.

**external data**

Data that are collected externally and merged in the CDMS or analyzed together with data collected on the e/CRF.

**F**

**false negative**

A test result that is erroneously classified in a negative category (as of diagnosis) because of imperfect testing methods or procedures. In statistics a Type II error.



**false positive**

A test result that shows evidence of a result or condition although it is not actually present. In statistics, a Type I error.

**field**

A particular area (as of a record in a database) in which the same type of information is regularly recorded.

**flag**

A tag placed on a data point that defines a status (e.g., discrepant, closed, or other status) that indicates an action is required.

**flow diagram, flow chart**

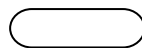
A graphic means for depicting the steps or activities that constitute a process. The flow diagram (flow chart) is constructed from standard symbols (the delay and database symbols have been added to Juran's list<sup>4</sup>).



The *activity symbol* is a rectangle that designates an activity. Within the rectangle is a brief description of that activity.



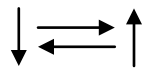
The *decision symbol* is a diamond that designates a decision point from which the process branches into two or more paths. The path taken depends on the answer to the question that appears within the diamond. Each path is labeled to correspond to an answer to the question.



The *terminal symbol* is a rounded rectangle that unambiguously identifies the beginning or end of a process. "Start" or "begin" is used to designate the starting point of a process flow. "Stop" or "end" is used to designate the end of process flow.



The *document symbol* is a document pertinent to the process.



The flow line represents a process path that connects process elements. The arrowhead indicates the direction of the flow.



The *connector* is a circle that is used to indicate a continuation of the flow diagram.



The *delay symbol* is a rectangle rounded on one side that identifies a waiting point or delay in the process flow.



The *database symbol* is a cylinder that represents a database application and the contained data.

### **frozen**

A temporary locked state for data that allows the generation of queries but does not allow a change to data points.

## **G**

### **global library**

In a Clinical Data Management System, the superset of all standard objects (e.g., CRF modules, edit checks, fields, etc.).

### **Good Clinical Practice (GCP)**

A standard for the design, conduct, performance, monitoring, auditing, recording, analyses, and reporting of clinical trials that provides assurance that the data and reported results are credible and accurate, and that the rights, integrity, and confidentiality of trial subjects are protected.

## **H**

### **hard coding**

Computer programs utilize logic and hardware to allow dynamic responses based on user input. For example, Web site can be programmed to tabulate the total bill when books are selected for purchase on-line or the average weight of the patients in the active treatment arm each time a program is run on a dataset. “Hard coding” is the limiting of the dynamic response by actually typing the data *in* the computer program itself rather than letting the data come from a dataset or the user. This approach can be dangerous because it is not visible in the analysis tables and listings or to the regulatory authorities and because it is easily forgotten once typed into the computer program.

### **hard lock**

The final state of the database where no changes are permitted and all user access is removed.

## **I**

### **impartial witness**

A person who is independent of the trial, who cannot be unfairly influenced by people involved with the trial, who attends the informed consent process if the subject or the subject’s legally acceptable representative cannot read, and who reads the informed consent form and any other written information supplied to the subject.

### **in-control process**

(ASQ) A process in which the statistical measure being evaluated is in a state of statistical control (i.e., the variations among the observed sampling results can be attributed to a constant system of chance causes). See also **out-of-control process**.

**independent data-monitoring committee (IDMC) (data and safety monitoring board, monitoring committee, data monitoring committee)**

An independent data-monitoring committee that may be established by the sponsor to assess at intervals the progress of a clinical trial, the safety data, and the critical efficacy endpoints. Such a committee may also recommend to the sponsor whether to continue, modify, or stop a trial.

**independent ethics committee (IEC)**

An independent body—i.e., a review board or a committee, whether institutional, regional, national, or supranational, constituted of medical professionals and non-medical members—that is responsible for ensuring the protection of the rights, safety, and well-being of human subjects involved in a trial and to provide public assurance of that protection. These responsibilities are accomplished by, among other things, reviewing and approving/providing favorable opinion on the trial protocol, the suitability of the investigator(s), facilities, and the methods and material to be used in obtaining and documenting informed consent of the trial subjects. The legal status, composition, function, operations, and regulatory requirements pertaining to IECs may differ among countries but should allow the IEC to act in agreement with GCP, as described in this guideline.

**informed consent**

A process by which a subject voluntarily confirms his or her willingness to participate in a particular trial after having been informed of all aspects of the trial that are relevant to the subject's decision to participate. Informed consent is documented by means of a written, signed, and dated informed-consent form.

**inspection**

1. (ICH) The act by a regulatory authority (or authorities) of conducting an official review of documents, facilities, records, and any other resources that are deemed by the authority to be related to the clinical trial and that may be located at the site of the trial, at the sponsor's and/or contract research

organization's (CRO's) facilities, or at other establishments deemed appropriate by the regulatory authority. 2. (ASQ) Measuring, examining, testing, and gauging one or more characteristics of a product or service and comparing the results with specified requirements to determine whether conformity is achieved for each characteristic.

**institution (medical)**

Any public or private entity or agency or medical or dental facility where clinical trials are conducted.

**institutional review board (IRB)**

An independent body—constituted of medical, scientific, and non-scientific members—that is responsible for ensuring the protection of the rights, safety, and well-being of human subjects involved in a trial by, among other things, reviewing, approving, and providing continuing review of trial protocol and amendments and of the methods and material to be used in obtaining and documenting informed consent of the trial subjects.

**instrument**

A device for capturing or measuring the present value of a quantity under observation.

**interim clinical trial/study report**

A report of intermediate results and their evaluation based on analyses performed during the course of a trial.

**intervention**

A method of interfering with the outcome or course, especially of a condition or process.

### **Investigational New Drug application (IND)**

An IND application is submitted to the FDA when a sponsor or investigator wishes to initiate trials with human subjects. The IND regulations can be found at the following link: [https:// www.fda.gov/cber/ind/ind.htm](https://www.fda.gov/cber/ind/ind.htm). “ND” is synonymous with “Notice of Claimed Investigational Exemption for a New Drug.”

### **investigational product**

A pharmaceutical form of an active ingredient or placebo that is being tested or used as a reference in a clinical trial, including a product with a marketing authorization when used or assembled (formulated or packaged) in a way different from the approved form, for an unapproved indication, or to gain further information about an approved use.

### **investigator**

A person responsible for the conduct of the clinical trial at a trial site. If a trial is conducted by a team of individuals at a trial site, the investigator is the responsible leader of the team and may be called the principal investigator. See also **subinvestigator**.

### **investigator/institution**

An expression meaning “the investigator and/or institution, where required by the applicable regulatory requirements.”

### **investigator meeting**

The kickoff meeting for an upcoming trial where the participating investigators review and provide feedback on the protocol or procedures in a protocol. Training of the principal investigator or other site staff on protocol procedures and/or EDC system entry is conducted at the investigator meeting as well.

**investigator's brochure**

A compilation of the clinical and non-clinical data on the investigational product(s) that is relevant to the study of the investigational product(s) in human subjects (see *ICH E6*, Section 7. "Investigator's Brochure"<sup>1</sup>).

**IOM**

Institute of Medicine.

**ISE**

Integrated Summary of Efficacy.

**ISO**

(ASQ) English acronym for **International Organization for Standardization**.

**ISO 9000 series standards**

(ASQ) A set of five individual, but related, international standards on quality management and quality assurance developed to help companies effectively document the elements that should be implemented to maintain an efficient quality system. Initially published in 1987, the standards are not specific to any particular industry, product, or service. The standards were developed by the International Organization for Standardization (ISO), a specialized international agency for standardization that is composed of the national standards bodies of 91 countries.

**ISS**

Integrated Summary of Safety.

## **L**

### **legacy system**

An electronic system previously in production, but no longer actively used, that may contain data needed for current analysis or other use and therefore must be maintained by the sponsor organization.

### **legally acceptable representative**

An individual, juridical, or other type of body that is authorized under applicable law to consent, on behalf of a prospective subject, to the subject's participation in the clinical trial.

### **local lab**

Local labs are labs in close proximity to individual clinical study sites or patients and are most often used when timely results are needed.

## **M**

### **MedDRA**

Medical Dictionary for Regulatory Activities is a medical terminology used to classify adverse event information associated with the use of biopharmaceuticals and other medical products. See [www.meddra.org](http://www.meddra.org) for additional information.

### **medical monitor**

An individual, other than the principle investigator, who evaluates clinical trial data from a safety perspective.

### **medical monitoring**

The act of evaluating the clinical trial data from a safety perspective.



**monitoring**

The act of overseeing the progress of a clinical trial and of ensuring that it is conducted, recorded, and reported in accordance with the protocol, standard operating procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s).

**monitoring report**

A written report to the sponsor that is produced by the monitor after each site visit and/or other trial-related communication, as specified by the sponsor's SOPs.

**multi-center trial**

A clinical trial that is conducted according to a single protocol but at more than one site and therefore is carried out by more than one investigator.

**N****NCS**

Non Clinically Significant.

**new drug application (NDA)**

The documentation submitted to the U.S. Food and Drug Administration. As described by the FDA:

The goals of the NDA are to provide enough information to permit FDA reviewer to reach the following key decisions: Whether the drug is safe and effective in its proposed use(s), and whether the benefits of the drug outweigh the risks. Whether the drug's proposed labeling (package insert) is appropriate, and what it should contain. Whether the methods used in manufacturing the drug and the controls used to maintain the drug's quality are adequate to preserve the drug's identity, strength, quality, and purity. . .

. . . The documentation required in an NDA is supposed to tell the drug's whole story, including what happened during the clinical tests, what the ingredients of the drug are, the results of the animal studies, how the drug behaves in the body, and how it is manufactured, processed and packaged.<sup>5</sup>

The NDA regulations are 21 CFR 314.

### **non-clinical study**

Biomedical studies that are not performed on human subjects.

## **O**

### **OCR**

Optical Character Recognition.

### **open access**

See National Cancer Institute's cancer Biomedical Informatics Grid (caBIG®) for additional details.

### **open development**

See National Cancer Institute's cancer Biomedical Informatics Grid (caBIG®) for additional details.

### **open source**

See National Cancer Institute's cancer Biomedical Informatics Grid (caBIG®) for additional details.

### **opinion (in relation to an independent ethics committee)**

The judgment and/or the advice provided by an independent ethics committee (IEC). See also **independent ethics committee**.

**out-of-control process**

(ASQ) A process in which the statistical measure being evaluated is not in a state of statistical control (i.e., the variations among the observed sampling results can be attributed to a constant system of chance causes). See also **in-control process**.

**original medical record**

See **source documents**.

**P****Pareto Principle / 80-20 rule**

An observation that 20% of the input creates 80% of the result

**phase I - IV**

Refer to the FDA glossary ([clinicaltrials.gov](http://clinicaltrials.gov)).

**predicate rule**

The overarching regulations that the industry must follow for GxP (Good “*Anything*” Practice or any collection of quality guidelines).

**production environment**

The location (e.g., website, server, EDC) where real clinical data is entered and stored.

**protocol**

A document that describes the objective(s), design, methodology, statistical considerations, and organization of a trial. The protocol usually also gives the background and rationale for the trial, but these details could be provided in

other protocol-referenced documents. Throughout the *ICH GCP Guideline*, the term “protocol” refers to protocol and protocol amendments.

**protocol amendment**

A written description of a change (or changes) to, or formal clarification of, a protocol.

**protocol deviation**

Any alteration/modification to the IRB-approved protocol. The protocol includes the detailed protocol, protocol summary, consent form, recruitment materials, questionnaires, and any other information relating to the research study. (Partners Human Research Committee; <http://healthcare.partners.org>)

**protocol violation**

Any protocol deviation that is not approved by the IRB prior to its initiation or implementation. (Partners Human Research Committee; <http://healthcare.partners.org>)

**Q**

**quality assurance (QA)**

All those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with Good Clinical Practice (GCP) and with the applicable regulatory requirement(s).

**quality control (QC)**

The operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled.

### **quality assurance/quality control**

(ASQ) Two terms with many interpretations because of the multiple definitions for the words “assurance” and “control.” For example, “assurance” can mean the act of giving confidence, the state of being certain, or the act of making certain. “Control” can mean an evaluation to indicate needed corrective responses, the act of guiding, or the state of a process in which the variability is attributable to a constant system of chance causes (for a detailed discussion on the multiple definitions, see *ANSI/ISO/aSQC a3534-2, Statistics—Vocabulary and Symbols—Statistical Quality Control*<sup>6</sup>). One definition of quality assurance includes the following: all the planned and systematic activities implemented within the quality system that can be demonstrated to provide confidence that a product or service will fulfill requirements for quality. One definition for quality control includes the following: the operational techniques and activities used to fulfill requirements for quality. Often, however, “quality assurance” and “quality control” are used interchangeably to discuss the actions that ensure the quality of a product, service, or process.

### **quality audit**

(ASQ) A systematic, independent examination and review to determine whether quality activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve the objectives.

### **query rule**

See edit check.

## **R**

### **random sampling**

(ASQ) A commonly used sampling technique in which sample units are selected in such a manner that all combinations of  $n$  units under consideration have an equal chance of being selected as the sample.

### **randomization**

The process of assigning trial subjects to treatment or control groups using an element of chance to determine the assignments. Used to reduce bias.

### **regulatory authorities**

Bodies having the power to regulate. In the *ICH GCP Guideline*, the expression “regulatory authorities” includes the authorities that review submitted clinical data and the authorities that conduct inspections (see Section 1.29<sup>1</sup>). These bodies are sometimes referred to as “competent authorities.”

### **research misconduct**

Falsification of data in proposing, designing, performing, recording, supervising, or reviewing research or in reporting research results. Falsification includes acts of omission and commission. Deliberate noncompliance with the regulations can be considered misconduct but is secondary to falsification of data. Research misconduct does not include honest error or differences of opinion.<sup>7</sup>

## **S**

### **safety database**

A database typically used by Drug Safety or Pharmacovigilance departments to collect adverse event data.

### **SAS transport file**

A machine-independent file that allows you to move a SAS data set from one operation system to another. (<http://kb.iu.edu/data/aevb.html>)

### **serious adverse event (SAE); serious adverse drug reaction (serious ADR)**

Any untoward medical occurrence that at any dose:<sup>2</sup>

- Results in death;
- Is life-threatening;
- Requires hospitalization or prolongs hospitalization of a subject;
- Results in persistent or significant disability/incapacity; or
- Is a congenital anomaly/birth defect.

### **Service Level Agreement (SLA) - from the Vendor chapter**

An SLA is part of a service contract where the level of service is formally defined.

### **SLA**

Service Level Agreement.

### **source data**

All information that is necessary for the reconstruction and evaluation of the trial, including information about clinical findings, observations, or other activities in a clinical trial. Source data are contained in source documents such as original records or certified copies of original records.

### **source documents**

Original documents, data, and records (e.g., hospital records, clinical and office charts, laboratory notes, memoranda, subjects' diaries or evaluation checklists, pharmacy dispensing records, recorded data from automated instruments, copies or transcriptions certified after verification as being accurate copies, microfiches, photographic negatives, microfilm or magnetic media, x-rays, subject files, and records kept at the pharmacy, at the laboratories, and at medico-technical departments involved in the clinical trial).

**special causes**

(ASQ) Causes of variation that arise because of special circumstances. These causes are not an inherent part of a process. Special causes are also referred to as assignable causes. See also **common causes**.

**specification**

(ASQ) A document that states the requirements to which a given product or service must conform.

**sponsor**

An individual, company, institution, or organization that takes responsibility for the initiation, management, and/or financing of a clinical trial.

**sponsor-investigator**

An individual who both initiates and conducts, alone or with others, a clinical trial, and under whose immediate direction the investigational product is administered to, dispensed to, or used by a subject. The term does not include any person other than an individual (e.g., it does not include a corporation or an agency). A sponsor-investigator must fulfill the obligations of both a sponsor and an investigator.

**standard operating procedures (SOPs)**

Detailed instructions written to achieve uniformity of the performance of a specific function.

**statistical process control (SPC)**

(ASQ) The application of statistical techniques to control a process. Often the term “statistical quality control” is used interchangeably with “statistical process control.”



**statistical quality control (SQC)**

(ASQ) The application of statistical techniques to control quality. Often the term “statistical process control” is used interchangeably with “statistical quality control,” although statistical quality control includes acceptance sampling as well as statistical process control.

**sub-investigator**

Any individual member of the clinical trial team designated and supervised by the investigator at a trial site to perform critical trial-related procedures and/or to make important trial-related decisions (e.g., associates, residents, research fellows). See also **investigator**.

**subject/trial subject**

An individual who participates in a clinical trial, either as a recipient of the investigational product(s) or as a control.

**subject identification code**

A unique identifier assigned by the investigator to each trial subject to protect the subject’s identity and to be used in lieu of the subject’s name when the investigator reports adverse events and/or other trial related data.

**T****trial site**

The location(s) where trial-related activities are actually conducted.

**trigger**

An event that precipitates other events.

### **Type I error**

(ASQ) An incorrect decision to reject something that is acceptable, such as a statistical hypothesis or a lot of products.

### **Type II error**

(ASQ) An incorrect decision to accept something that is unacceptable.

## **U**

### **UAT**

User Acceptance Testing.

### **unexpected adverse drug reaction**

An adverse reaction, the nature or severity of which is not consistent with the applicable product information (e.g., investigator's brochure for an unapproved investigational product or package insert/summary of product characteristics for an approved product). See the *ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting*.<sup>2</sup>

## **V**

### **variable**

See also field.

### **VCL**

Virtual Central Lab

### **vulnerable subjects**

Individuals whose willingness to volunteer in a clinical trial may be unduly influenced by the expectation, whether justified or not, of benefits associated with participation, or of a retaliatory response from senior members of a hierarchy in case of refusal to participate. Examples are members of a group with a hierarchical structure, such as medical, pharmacy, dental, and nursing students, subordinate hospital and laboratory personnel, employees of the pharmaceutical industry, members of the armed forces, and persons kept in detention. Other vulnerable subjects include subjects with incurable diseases, persons in nursing homes, unemployed or impoverished persons, subjects in emergency situations, ethnic minority groups, homeless persons, nomads, refugees, minors, and those incapable of giving consent.

## **W**

### **well-being (of the trial subjects)**

The physical and mental integrity of the subjects participating in a clinical trial.

### **WHOdrug**

WHO Drug is a dictionary of medicinal product information. It is used to identify drug names and provides information about a drug's active ingredients and its therapeutic use(s).

## **X**

### **XML**

Extensible Markup Language is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

## References

1. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E6*. Geneva, Switzerland: Author; 1996.
2. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E2*. Geneva, Switzerland: Author; 1996.
3. International Conference on Harmonisation. *Guideline for Good Clinical Practice, E3*. Geneva, Switzerland: Author; 1996.
4. Juran JM. *Juran's Quality Control Handbook*, 4th ed. New York, NY: McGraw-Hill; 1988.
5. US Food and Drug Administration. New Drug Application (NDA) Process Web site. Available at: <http://www.fda.gov/cder/regulatory/applications/nda.htm>. Accessed on February 27, 2007.
6. American Society for Quality. *ANSI/ISO/ASQC A3534-2-1993: Statistics—Vocabulary and Symbols—Statistical Quality Control*. Milwaukee, WI: American National Standards Institute; 1993.
7. Woolen S. Office for Good Clinical Practice Web site. October 2001. Available at: <http://www.fda.gov/oc/gcp/>. Accessed on February 27, 2007.

## Chapter Revision History

Publication Date	Comments
September 2000	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
October 2013	Revised with the addition of approximately seventy-five (75) terms.