SCDM SCDM SCDM SCDM

DATA BASICS

Volume 10
Number 2
2004 Summer

Promoting Clinical Data Management Excellence

A NEWSLETTER SUPPORTED BY AND FOR THE MEMBERS OF THE **SOCIETY FOR CLINICAL DATA MANAGEMENT, INC.**

Letter from the Editors

Dear SCDM Members,

To quote SCDM's Good Clinical Data Management Practices, "It is important that Data Management professionals take a proactive approach in setting appropriate standards for acceptable data quality levels, methods for quantifying quality, and practices to assure quality." The effort to achieve an error-free database is neither practical nor realistic. We would be better served to redirect this effort to understanding, identifying and correcting those errors that would have an impact on study results.

Based on the overwhelming response from authors and Spring Forum session moderators, it is clear that this topic is one of intense interest. So, in order to provide the full breadth of information, two issues of Data Basics will be devoted to the theme, Data Quality. This issue will explore quality concepts and the SCDM Data Quality Survey. The next issue will focus on different aspects of data management (CRF design, software, regulatory requirements and Data Quality Research Institute) and their contribution to the quality of data.

It is our hope that the information provided will be a help and stimulate discussion. Please let us know your thoughts about this and other topics by contacting Kit Howard at kit@kestrelconsulting.com or Lynda Hunter at HunterLynda@PRAIntl.com.

Lynda Hunter and Kit Howard ■

The Case for Clinical Data Quality

Kaye H. Fendt, DQRI

Introduction

Over the last two decades, as regulatory requirements have increased and health-related research has focused more on rare diseases and conditions, the number of trials, patients and clinical procedures required to demonstrate significant differences has also increased. As a result, we have seen a large increase in the amount of data requiring corporate staff to electronically capture, review, analyze, interpret, and summarize in clinical research reporting, whether reported in peer reviewed journal articles or in support of a marketing application with a format regulated by the various biopharmaceutical regulatory agencies. Unfortunately, methods and metrics for assessing the quality of these data and the impact of that quality on the data-based decisions made have not developed at the same pace. As data management has matured over the last 40 years, increasing emphasis has been placed on the procedure to eliminate errors primarily from the point of database entry to the extraction of data from the database. Efforts to design quality into the procedures have developed in a variety of ways across the research community. Today there is an emerging need to define, measure, and evaluate the quality of data and the impact of the quality of data used for decision making throughout the entire life-cycle of the data. Also, there is a growing need for further research into how these methods can be standardized, reported in a way to allow responsible use of the data in the future, cost-contained, and implemented in an efficient manner.

The need for high quality standards for data used for global decision making in regulated industry is

Continued on page 4

This Issue

SCDM SCDM SCDM SCDM

DATA BASICS Editorial Policies

Editorial Board

(also known as Publications Committee)

Kit Howard, Co-editor
Kestrel Consultants, Inc.
Phone: 734-576-3031
E-mail:
kit@kestrelconsultants.com

Lynda Hunter, Co-editor
PRA International
Phone: 913-577-2972
E-mail:
hunterlynda@praintl.com

Cherie Stabell, Editorial Board Chair
Genentech, Inc.
Phone: 650-225-7672
E-mail: stabell@gene.com

Felicia Ford-Rice
Acambis, Inc
Phone: 678-336-2713
E-mail:
Felicia.FordRice@Acambis.com

Alec Vardy
CV Therapeutics, Inc.
Phone: 650-384-8519
E-mail: alec.vardy@cvt.com

Belinda Field
Pfizer
Phone: 860-732-9880
E-mail:
Belinda_field@groton.pfizer.com

Elena Mullin
Cell Therapeutics, Inc.
Phone: (206) 272-4381
E-mail:
emullin@ctiseattle.com

Larry Stark
Cell Therapeutics, Inc.
Phone: (206) 272-4186
E-mail: lstark@ctiseattle.com

Tamela Blackstone
Allos Therapeutics
Phone: (720) 540-5274
E-mail: tblackstone@allos.com

Submission Deadlines (Articles and Advertising Art Work)

Our quarterly publication schedule for the next three issues requires the following input deadlines:

Volume 10, #1 (Spring)	27 February 2004
Volume 10, #2 (Summer)	26 April 2004
Volume 10, #3 (Fall)	27 July 2004
Volume 10, #4 (Winter)	25 October 2004

Publication Policy

We welcome submission of previously unpublished materials for publication in *Data Basics.* Materials should preferably be submitted in electronic form (MS Word). Acceptance of materials for publication will be at the sole discretion of the Editorial Board. The decision will be based primarily upon professional merit and suitability (i.e. publication may be edited at the discretion of the Editorial Board.

Niether SCDM nor the Data Basics Editorial Board endorses any commercial vendors or systems mentioned or discussed in any materials published in *Data Basics.*

Advertising Policy

AD RATES**	x1	x2	x3	x4
FULL Page	$675 each	$640 each($1280)	$610 each ($1830)	$575 each($2300)
HALF Page	$470 each	$445 each($890)	$425 each($1275)	$400 each($1600)
QTR Page	$285 each	$270 each($540)	$255 each($765)	$240 each($960)

**Ads are net, non-commissionable.

Advertisers purchasing multiple ad packages will have the option of placing those ads anytime within the 12-month period following receipt of payment by SCDM.

Quarter Page = (3 5/8 inches x 4 7/8 inches)

Half Page-vertical = (3 5/8 inches x 10 inches)

Half Page-horizontal = (7 ½ inches x 4 7/8 inches)

Full page = (7 ½ inches X 10 inches)

MECHANICAL REQUIREMENTS: Do not send logo/photos/images from word processing software, presentation software, or Web sites. Files should be saved in the native application/file format in which they were created. Photos/images should be high resolution and received in the file size you wish to have it printed.

600 dpi or higher for black and white, 300 dpi or higher for color/grayscale

Acceptable file formats include AI, EPS, and high resolution PDF, PSD, JPEG, and/or TIFF

PAYMENT: Payment must be received with advertising. Space reservations cannot be made by telephone. There is NO agency discount. All ads must be paid in full.

CANCELLATIONS: Cancellations or changes in advertising requests by the advertiser or its agency 5 days or later after the submission deadline will not be accepted.

GENERAL INFORMATION: All ads must be pre-paid. Publisher is not liable for advertisement printed from faulty ad materials. Advertiser agrees to hold SCDM harmless from any and all claims or suits arising out of publication on any of his/her advertising. SCDM assumes no liability, including but not limited to compensatory or consequential damages, for any errors or omissions in connection with any ad. The SCDM does not guarantee placement in specific locations or in a given issue. SCDM reserves the right to refuse or pull ads for space or content.

Please submit all forms, artwork, and payments to:

Society for Clinical Data Management
555 E. Wells St.
Suite 1100
Milwaukee, WI 53202

Phone: 1 (414) 226-0362
Fax: 1 (414) 276-3349
E-mail: info@scdm.org
Web site: www. scdm.org

The Case for Clinical Data Quality

Continued from cover

not new. As early as the 1970s, the public health service recognized and provided support by funding a major research university supplemental graduate curriculum to educate Research Data Managers and to conduct research on methods of improving efficiency and quality of data management procedures for longitudinal multi-site research data. Today, the need continues and has become even more important as government agencies and other regulatory bodies rely more and more on the evaluation of electronically collected, stored, transmitted, and archived data for critical data-based decision making.

The Impact of the Quality of Regulated Clinical Data

What is data quality and why should we care? In a landmark workshop report in this area, the Institute of Medicine (IOM) defined data quality in 1999 as "data that support conclusions and interpretations equivalent to those derived from error free data".[1] Larry English breaks this definition into *inherent* and *pragmatic* quality. Inherent quality refers to the "correctness or accuracy of data" and pragmatic quality is "the value that accurate data has in supporting the work of the enterprise".[2]

Along with the increases in data quantities, technology has advanced to meet the needs of the Information Age. However, theories and concepts to make the best use of these technological advances have not developed as rapidly. As dynamic data becomes a reality for the regulatory world, this becomes even more important. The current explosion of data and information management tools will increase the reliance on data and reduce the time to detect errors that have been generated. Thus, it is time to develop focused methods for looking at the quality of data in the entire drug development process.

While quality research (e.g. Deming) is highly developed within fields such as manufacturing, this research is not currently directly applicable to the subtleties of clinical data. Beginning in the 1970's, public health leaders recognized a need for high quality data in longitudinal studies. At this time, a grant was funded to develop an entirely new concept: the quality of data from the beginning to end of the data life cycle. This endeavor begins with the question definition and does not end until the end of data archival. The founding names in the field of data emerged from this funding period.

While data management has existed as a discipline since the 1950s, data quality research did not emerge until the 1970's. Until then, data management had been a very manual process and dealt with smaller amounts of data. The development of high quality data managers, however, led to the demand for more like themselves, as their value was now perceived in a broader arena. Increased demand encouraged the integration of professionals from other fields in the field of data quality, and, as a result, the field changed. Today data management can be the responsibility of a trained statistician, a nurse, or a clerk. As the professional responsibility for data management became less streamlined, the breadth of its scope was narrowed. The scope of data management had diminished to solely the data residing in the clinical database, rather than the full spectrum of data activities. As a result of these two phenomena, the study of the quality of data through its entire life cycle became a rarity.

In laying the foundation for data quality research, the Council for Excellence in Government (CEG), CDISC, and the GCDMP all played integral roles. The CEG program served as the launching pad for the GCDMP, CDISC, and ultimately the Data Quality Research Institute (see elsewhere in this issue for further details). ICH-E2B established some electronic storage standards, while CDISC established standards for data transfer. The GCDMP began the definition of standard processes, providing the link between regulatory guidance and data quality management in practice.

With these standards in place, the missing piece has become the definition, measurement, and reporting on the quality of data in the clinical research environment. Here in the 21st century, almost all decisions are now based on huge amounts of data in complex databases, and those data are manipulated in a myriad of ways. Techniques and approaches developed in the 1970's are simply no longer adequate to assess the quality of data in today's databases. While new techniques have been developed, they have not been assessed or validated in a coherent

Continued on page 4

SCDM Committees

The following are currently active Committees within the Society for Clinical Data Management.

Certification Committee
Chair : Armelde Pitre
Phone: (203) 226-6456
E-mail: pitrea@snet.net

GCDMP Committee
Chair : Christine Little
Phone: (919) 408-8000
E-mail: clittle@rhoworld.com

Membership Committee
Chair : Brenda Hoeper
Phone: (513) 984-0450
E-mail:
Brenda.Hoeper@quintiles.com

Publication Committee
Chair: Cherie Stabell
Phone: (650) 225-7672
E-mail: stabell@gene.com

Web Site Committee
Chair: David Borbas
E-mail: dave@borbas.net

Web Sites to Check Out

ACDM - www.acdm.org.uk
CDISC - www.cdisc.org
FDA - www.fda.gov
ICH - www.ich.org

There are more links to be found on our web site!
SCDM - www.scdm.org

Please let the Web site Committee know about any other "hot" web sites that you feel would be of interest to the SCDM membership.

The Case for Clinical Data Quality

Continued from page 3

manner. In addition, the techniques that have been developed are designed to look at small subsets of the data out of the context of the data lifecycle.

With the framework of E2B, CDISC, and the GCDMP in place, the effects of data quality on health decisions can now be researched. There is a need for work process improvements and resource maximization surrounding the topic of the quality of clinical trials information. Regulated industries and regulatory bodies need efficient and useful standards to evaluate the quality of data used in decision making. To have metrics and procedures for evaluation of data quality could enable regulatory decision makers to evaluate the confidence level of decisions made based on the data.

Specifically, at the IOM meeting in April 1998, Ken Shine of the Academy of Science identified the "quality and validity of clinical trials in regulatory decision-making" and "GRDMP" (Good Regulatory Data Management Plans) as high-priority needs in meeting our objective of collecting relevant, high-quality data at a reasonable cost. The regulated industries need more transparency and clarity about the data quality process. Today, quality information research must return to the larger view and evaluate the impact of data quality in the system that begins with the protocol design and extends through the analysis and reporting to the archiving of clinical data.

Current State of the Quality of Biopharmaceutical Data

Data used to make decisions in clinical trials must be accurate. Decisions about dosages, risk of adverse events, and risk-benefit profiles of the treatments are made using these data. Regulated industry expends major resources to check and correct the data prior to any analysis. This includes checking the original source document to the data input, checking the values to make sure they make sense in the field, and looking for outliers. This checking is very expensive and time consuming. Similarly, regulated industry allocates resources to assure the quality of their computer systems to make sure the data are not compromised by the systems processing the data. This checking includes a detailed software development life cycle for software developed in-house, vendor evaluation, and installation and operational qualification for purchased software. The combined effort for the data checking and computer systems validation is huge.

Standards

Appropriate standards positively impact quality. Currently, standards exist for software development, data management, and data transfer formats. With standard data elements, we can improve the efficiency of the drug development documentation and review process. With standard data elements, we can reduce the CRF (case report form) design and development time, improve consistency across trials and reduce confusion. In short, with standard data elements we can reduce the cost of trial design, conduct, submission and review.

The International Organization for Standardization (ISO) provides a set of standards used in software engineering including ISO 12207 (a framework for software life cycle processes) and ISO 9003 (a quality management system for software.

The Good Clinical Data Management Practices (GCDMP) (www.scdm.org) provides guidance and thus establishes standards on accepted practices for the many areas of Clinical Data Management (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 Clinical Data Management. The GCDMP recommends guidelines and provides practical suggestions and proven means for meeting the guidelines.

The GCDMP addresses the CDM areas of responsibility in ten sections. Each section provides Minimum Standards as well as Best Practices. These summarize the main recommendations of each section in bulleted form. Each section also contains recommended Standard Operating Procedures. Given that data management tasks are often technical and specialized, and in the absence of CDM regulatory standards, it is important for experienced, professional data managers to provide thought leadership on accepted data quality levels, practical methods of achieving them and implications of new technology on the CDM tasks.

Continued on page 5

The Case for Clinical Data Quality

Continued from page 4

The Clinical Data Interchange Standards Consortium (CDISC) (www.cdisc.org) is an open, multidisciplinary, non-profit organization committed to the development of industry standards to support the electronic acquisition, exchange, submission, and archiving of clinical trials data and metadata for medical and biopharmaceutical product development. The mission of CDISC is to lead the development of global, vendor-neutral, platform-independent standards to improve data quality and accelerate product development in our industry.CDISC has developed data exchange standards for use in Clinical Research, and many organizations participating in data interchange activities have requested expert assistance for implementing these standards for specific applications.

Models for submission of data to the FDA have been completed and are currently being piloted by the FDA; models for sharing data across the industry are also in late stage development and testing

On May 20, 2002, Dr. Elias A. Zerhouni, began his tenure as the 15th Director of the National Institutes of Health and initiated the creation of a new research vision for the NIH that focuses the attention of the biomedical research community on new pathways of discovery, research teams for the future and the re-engineering of the clinical research enterprise. This initiative provides an opportunity for the Industry and government to focus on the need for, the impact of, and the risk/benefit of research into how to improve the quality of clinical research data and potentially reducing the cost or providing good data-based health decisions at the same time. The time has come to bring all the pieces together and create the definitive body of clinical data quality knowledge.

The author would like to thank Sue Carroll and Kay Obenshain for their help with this paper. ■

References

1 *Institute of Medicine Roundtable Report, April, 1999, Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making, Workshop Report, Institute of Medicine, National Academy Press, Washington, D.C. 1999.*

2 *English, Larry P. Improving Data Warehouse and Business Information Quality, New York: Wiley, 1999.*

Summary of 2004 SCDM Spring Forum Session A: GCDMP- Measuring and Assuring Data Quality

Facilitated by Anita Walden, Duke Clinical Research Institute and Renee Pridgen, Duke Clinical Research Institute

Objective: To generate discussion around measuring and assuring quality data and how real life practices compare with GCDMP (Good Clinical Data Management Practices) Minimum Standards and Best Practices.

The GCDMP Measuring and Assuring Data Quality Minimum Standards and Best Practices are guides for the Data Management industry but are they sufficient standards and who is following them?

Company Practices

The sessions began with an exercise that allowed us to get to know each other and find out how different aspects of the Measuring and Assuring Data Quality sections of the GCDMP are viewed by different companies. The participants were separated into small groups where they reviewed the GCDMP Minimum Standards and Best Practices and discussed how they fit into their current organizations. They were asked to keep the following questions in mind during the discussions:

- Does your company follow these practices?
- Are Minimum Standards enough?
- Are Best Practices too much?

The session participants came with a broad range of perspectives regarding measuring and assuring data quality. While some worked at organizations that don't do database audits, asking "What do we gain?", others worked at organizations where most if not all of the GCDMP Best Practices are followed. Some participants felt database audits should be done before database lock, others after database lock, and still others – not at all.

There were some who felt that the minimum standard were not complete. Others felt they were not the right set of standards.

Ideal Standards and Measurements

Since everyone had an opinion on whether or not the GCDMP standards were the right standards, the next question was to ask the industry leaders what they thought the standards should be. In the next exercise, each group was asked to list the steps that they, as a group, felt were necessary to ensure a quality database at the time of database lock. Some groups thought that steps only Data Management could control such as the database design, processing and closeout procedures should be part of the GCDMP. Other groups thought that the GCDMP standards should include everything that impacts the data quality from the protocol to the designing of the CRF to the monitoring plan.

Aside from the scope of the data manager's responsibility on data quality, most data managers agreed on what assured a quality database with very few exceptions. The exceptions included whether CRF to Database Audits were necessary and if so, when to have them; the use of error rates, who decides on the error rate and how to calculate it; and Electronic Data Capture (EDC), whether performing single entry at the site level with no audit when we traditionally perform double data entry with an audit, is adequate.

Most Important Standards and Measurements

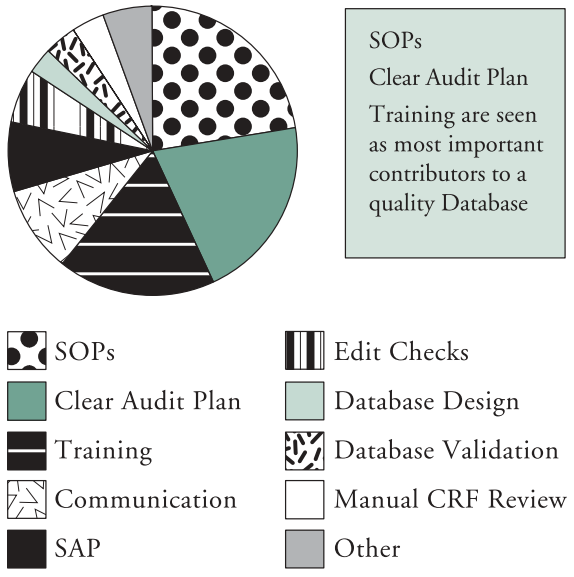
The results of each team's list of components that make up a quality database were then reported out to the larger group so that one comprehensive list was compiled.

The final part of the session was for each participant to identify the things from the group list that they felt were most important to data quality. Everyone received three votes to use however they wanted so that if one thing was extremely important to them, they were free to give that item more than one vote.

Results

According to the session participants, having solid SOPs and standard processes came out as the number one thing we can do as Data Managers to ensure a quality database at database lock. This was followed closely by consistent training and having a clearly defined audit plan *at the beginning* of the trial. Other notable things that the participants agreed are important are clear communication plans across functional groups, early defined Statistical Analysis Plans (SAPs), and predefined edit checks.

What is notable about the results is that two items – training and communication – are not mentioned anywhere in the Best Practices



SCDM Spring Meeting 2004 Results of the Most Important Items that Contribute to Assuring and Measuring a Quality Database

Continued on page 7

Summary of 2004 SCDM Spring Forum Session A: GCDMP- Measuring and Assuring Data Quality

Continued from page 6

or Minimum Standards for GCDMP Assuring or Measuring Data Quality. Are these implied? Maybe. Should they be specifically mentioned? The participants in our session felt the answer to this question is yes. As a group it was clear that the GCDMP standards are a reflection of what individuals think should measure and assure quality data with only two critical items missing.

Conclusion

The results of this session will be given to the GCDMP committee to decide if modifications should be made to the Quality Audit sections. Since more people are aware and reading these guidelines it is important that the minimum standards are adequate to assure a quality database.

More to Come

It was apparent from facilitating these sessions that this topic has generated a lot of discussion. Some of the questions are generating discussions that could change in the near future the way data management measures and assures data quality.

Some of the questions…

- CRF to Database Audit – how critical is it, how much weight should it carry in determining data quality and does it really measure data quality accurately?
- Who are the minimum standards for and if they produce a quality database why have best practices?
- Do best practices change trial results?
- Error rates – who decides what is an acceptable error rate? Can we come up with a standard? ■

Spring Forum Moderators *Front Row from left to right: Jonathan Andrus, Judy Pyke,* ***Anita Walden, Renee Pridgen****. Back Row from left to right: Greg Dziem, Anthony Costello, Abdel Oualim.*

Pick A Number, Any Number?

Hugh Donovan, Aventis Pharmaceuticals

When one thinks of the statistical profession in the context of the pharmaceutical industry, the number 0.05 instantly springs to mind. This is the Holy Grail, the sine qua non of clinical trials. The industry's reliance, more importantly the regulatory authorities' acceptance, of this one number has given statisticians legitimacy within clinical development. Despite the issues of multiple comparisons, and assumptions about the underlying distributions, 0.05 has helped statisticians gain a valued place at the clinical development table. Should we in Clinical Data Management strive to develop a similar standard number that might instantly enhance our status?

If so, what area of our work should we draw attention to quantitatively? When determining the approvability of a submission, two of the three key internal performance indicators, time and cost, are irrelevant to the reviewers. The third, namely quality, is the aspect of the regulatory review that most closely relates to our contribution, therefore this is where we should logically focus any efforts to come up with the magic number, the universal indicator of acceptable or unacceptable quality. As confirmed by recent communications in the SCDM's Discussion Forum, there is currently no standard. Therefore we would have to create one. Sounds simple, but what should the number be, how would we derive it, and, most importantly, would it help us?

Calculating the so-called 'error rate' is the easiest part of the problem. Simply check all, or a sample, of data values, determine the number of errors, divide it by the number of values checked, multiply by 100 and, there you have it, the percentage error rate that is usually quoted. For details of the real complexity of such an exercise, see the GCDMP document, but this is adequate for the purposes of this discussion. What does the error rate tell us? There is perhaps a misconception that it reflects the quality of the study itself, whereas it just reflects the accuracy of the entry and subsequent processing of the data as reflected in the paper case report forms or collected electronically through an EDC system. It does touch upon the accuracy of the transcription from source to paper/screen. Also it does not give any indication of the quality of the study conduct; were the right subjects enrolled, was the data collected according to the protocol, to GCP, etc. These components of quality are at least as important as the error rate we calculate and quote.

Assuming that, despite this limitation, it is still considered valuable to have one universal number to judge, in a binary fashion, the acceptability of the database, and, therefore, the acceptability of our work, how do we determine what the standard should be? A major obstacle to this is the lack of a common definition of an error. It is not a controversial topic, and not one that has been much discussed, because, without an industry standard number, it is irrelevant. However, it can have a very large bearing on what the number is. For example, is every missing comma or insignificant typo in a comment field going to be given the same weight as an error in the primary endpoint? Reaching agreement on this and other related issues could result in as much debate as the CDISC initiative has generated, particularly if taken in the context of the overall acceptability, or otherwise, of a study database.

Having one value may give us instant name recognition; perhaps more correctly, 'number' recognition, but is it something that we want to be associated with? We do not want to be associated with an unacceptable error rate. True, if it is initially too high, we can take corrective action, until, like a good limbo dancer, we can just get under the barrier, but is that what we want, both for the profession, and for our companies? First there could be the stigma of the fact that it was initially a failure and the perception that the data had to be subsequently manipulated to reach the right value. Then there is the danger of complacency. If the value is less than some arbitrary level at the first attempt, do we send out the usual congratulatory e-mails and rest on our laurels? Usually error rates are based on samples rather than 100% checks of the data. The checking process is subject to error itself. Also, possible danger signals seen during the checking can be ignored because the database was 'accepted'. One or two major mistakes, discovered during a review, can have a major impact on the results; even if they do not change the safety and efficacy conclusions, they can cast doubt upon the integrity of the data, whether or not the initial acceptability criterion was met.

Determining industry wide standards, accepted by regulatory authorities, is an onerous task, and one that should only be undertaken if the benefits justify it. Data standards are such an example, and the CDISC initiative, although complex and time-consuming, will ultimately pay off if submissions can be more readily assembled and reviewed. In contrast, the value of determining one number that would be used as the yardstick against which all study databases are measured seems dubious. It only reflects one aspect of the overall quality of the study and associated data. It would require a large effort to reach consensus on the methodology and definitions. Also, it could lead a false sense of security that could ultimately do more harm than good, not only to the reputation of the specific company, but to the overall reputation of the CDM profession, which would be the exact opposite effect of the one we were trying to achieve. Therefore, the recommendation is to accept the status quo and, in this particular instance, to resist the CDM tendency, valuable as it is in most cases, to standardize.

What, then, is the value of the error rate that we all love to quote, particularly if it is less than our company's standard? Assuming that there is a standard methodology and standard definitions are used across studies, across projects, then there is a value in comparing rates across studies and over time, to see if the lessons have been learned and there has been measurable improvement. However, the qualitative aspect of a database audit is more important than the quantitative aspect. Regardless of the error rate, there may be findings that need to be rectified, not only for the particular database, but which should be applied across all ongoing and future projects. Too much dependency upon an arbitrary number may take away the opportunity to benefit from such action. In conclusion, the error rate is a useful tool, but one which has to be seen as a means to continuous improvement rather than an end in itself. It is certainly not the CDM profession's admission ticket to the top table in the pharmaceutical industry. ■

Fit for Reporting

Katherine R. Ciambrone, GlaxoSmithKline

What are the Data?

When Words and Expectations Contradict

In the past twelve years and in multiple clinical data management environments, I have noticed that clinical data management (CDM) has been and always is, stuck between a rock and a hard place. The rock represents the high expectations of our customers and partners in Clinical and Statistics. The hard place is the reality of the constant process and technology changes and seemingly unlimited tasks that must be completed within certain time constraints by qualified individuals. Overall, what we do in CDM is not horribly complicated or unobtainable. We almost always manage to fulfill our obligations, meet the timelines and please our customers. So why is there always a perception and reality that we are treading water to survive. This dilemma has long perplexed me. I had to ask why.

Setting Expectations

In searching for the answer, I began to look at what we do in CDM. We clean data. I immediately realized the word 'clean' is the root of many of our issues. We use the word clean not only as a verb, in that we *clean* the data, but when we complete the cleaning process, we deem the database *clean*. CDM uses this term, both internally and externally, with our customers to explain our process. It is, unfortunately, the wrong word and sets incorrect expectations. Not only is CDM not setting clear expectations internally and for our customers, but this confusion is compounded by federal regulations which also do not clearly identify the minimal acceptable data quality levels. It is important for CDM to take a proactive role in setting appropriate standards for acceptable data quality levels, methods of quantifying data quality, and practices to assure data quality[1].

In trying to be proactive, however, we have taken the simple idea of cleaning data and have complicated and misconstrued it by adding so many checks and balances to assure the data is beyond the statistically acceptable quality parameters. This has manifested itself in a deplorable plethora of inefficient procedures purported to ensure a defendable level of data quality in clinical trials[2]. Rather than focusing on true value-added procedures, which are not only common sense, but also good business sense, we have instilled a culture in CDM departments that perfection is the goal. Instead, we should focus on preventing questionable practices; not ensuring the data is entirely free of error. Some errors will remain undetected and uncorrected regardless of quality assurance, editing and auditing[3].

Therefore, we are, in fact, sending conflicting messages by using phrases such as; the data is clean, clean enough, and meets a certain level of quality. A common obstacle in communication on technical matters is the use of terminology that is not understood by one of the parties[4]. The word 'clean' means free from impurities, free from wrongdoing, even, regular and thorough, and complete. Our clinical data definitely does not meet all those definitions all the time.

I suggest to my CDM community that we strike the word 'clean' in our day-to-day tasks.

Alternatives

In the last couple of years, many of us are trying to make the word 'quality' fit into our daily rhetoric. The definition of the word 'quality' that applies is the degree or grade of excellence. Remember that our data cannot be totally error free; therefore, if we use the word quality we must always preface it with a description that reflects the state of quality such as 'acceptable quality'.

So where are we? Basically, we are lacking adequate instructional guidance from the government. We have too many inefficient non-value added processes with the goal of perfection. It is impossible to get an error-free database and CDM has customers and staff with unrealistic expectations of what we do.

Solution

In CDM, we need to change how we refer to the state of the data to set clear expectations for our deliverables. I suggest we use the word 'Fit' to describe our data. We need to remove from CDM terminology key phrases we have all heard too often, such as: the data is clean, acceptable data quality levels, appropriate data quality for intended purpose, and assurance that data is of an agreed quality.

'Fit' means to be the correct size and shape, to be appropriate and to be or make suitable. Use of the word 'Fit' demonstrates a balanced approach to meet the quality, cost, and efficiency needs of today's measured environment. Other FDA areas use the word 'fit' such as in 7 CFR Part 1434: Metal containers must meet the requirements of the Federal Food, Drug and Cosmetic Act, as amended, and regulations issued thereunder must be generally fit for purpose for which they are to be used. The Information Technology area uses the word 'fit' for general expressions, which can be useful to ensure that solutions are appropriate for your organizations. There are Fitness for Duty programs to provide reasonable assurance that personnel will perform tasks in a reliable and trustworthy manner (10 CFR Part 26).

In CDM, we can use the word 'fit' to set the right expectations for our main deliverable - a final database. The Fit for Reporting (FFR) definition could be used to reflect the state of our data. FFR would mean that critical variables meet expectations, no statistically significant errors exists, reliability in data handling exists, processes are ethical, risk management processes are in place and followed, and an appropriate time and cost investment to achieve the desired outcome of a study has been applied.

In Conclusion

I suggest we strike the word 'clean' from all CDM activities, our documentation and our training programs. We use the word 'Fit' to set the appropriate expectations, standards, methods and practices supporting overall credible and accurate reporting. We incorporate 'Fit for Reporting' as a milestone replacing the Database Release or Freeze, which describes an activity performed in CDM, not the state of the data. ■

Continued on page 11

Fit for Reporting

Continued from page 10

References

1 *Good Clinical Data Management Practices, Society of Clinical Data Management, Version 2, 2002.*
2 *Lorstad MH, '"Too much, Too late" to "Right the first time": Quality Guru's Deming's Advice for Clinical trials'. Drug Inf J. 2000. 34:1319-1328.*
3 *Kannerud GL, Rockhold FW, George SL, Barton FB, Davis CE, Fairweather WR, Honohan T, Mowery R, O'Neill R. 'Guidelines for quality assurance in multicenter trials: a position paper'. Control Clin Trials. 1998; 19:477-493.*
4 *Karten, Naomi, Managing Expectations: working with people who want more, better, faster, sooner, now! New York: Dorset House. 1994.*
5 *Webster's II New Riverside Dictionary*
6 *ICH E5*

We've Moved!

The SCDM office has changed location. Please direct correspondence to the following address:

Society for Clinical Data Management
555 E. Wells Street
Suite 1100
Milwaukee, WI 53202-3823

Good Clinical Data Management Practices (GCDMP) is available to download from the SCDM website, www.scdm.org. To date, over 800 people have downloaded GCDMP from all over the world.

SOCIETY FOR CLINICAL DATA MANAGEMENT

2004 FALL CONFERENCE

2004 Fall Conference Clinical Data Management in a Global Environment

October 3-6, 2004 • Fairmont Royal York Hotel • Toronto, Canada

Watch for Preliminary Program to be mailed in June
For more information, please visit www.scdm.org/events/

Data Quality Survey Results

Meredith Nahm, Duke Clinical Research Institute; Greg Dziem, Amgen; Kaye Fendt, DQRI; Lisa Freeman, Corus Pharma, Inc.; Joann Masi, Wyeth; Zoe Ponce, Wyeth

Introduction

In both industry and government funded research, there are different definitions of Data Quality, as well as different approaches to measuring, reporting and acting on data quality information. There is little aggregate information available within the industry on the data quality practices of organizations. Two surveys on data quality have been previously conducted. One was conducted in 1994 by the Association for Clinical Data Management (ACDM) and the second survey was conducted in 1997 by the Society for Clinical Data Management (SCDM). The 1994 ACDM survey was conducted in a European population. The results were summarized in the Drug Information Association (DIA) Journal in a 1999 paper by McEntegart, *et.al.*[1] The 1997 SCDM survey was conducted in a primarily US population, and was published in DataBasics.[2] Both surveys showed that a majority of organizations perform data quality audits.

Regulatory guidance in the area of Data Quality is sparse. ICH E6 section 5.1.3 states that quality control must be applied to each stage of the data handling process to assure that data are processed accurately.[3] The stated purpose of 21 CFR Part 11, the electronic record and electronic signature rule is to assure that electronic data are as good as data recorded on paper.[4] The associated FDA guidance, Computerized Systems Use in Clinical Trials provides more specificity, stating that data must be Attributable, Legible, Contemporaneous, Original, and Accurate (ALCOA).[5] The available regulations and guidance documents seem to specify everything but an acceptable quality level and acceptable methods to measure it. Therefore, organizations have turned to industry standards. The Good Clinical Data Management Practices (GCDMP) document[7] includes an entire section on best practices, minimum standards and methods for measuring data quality. In the absence of regulations, organizations want not only information on suggested methods and evidence of their effectiveness, but also current information on industry practices to aid in internal decision making

The Survey

In March 2004, the Society for Clinical Data Management conducted a data quality survey to gain insight into current industry practices and approaches to quantifying Data Quality. Three main aspects of Data Quality were investigated, definition of data quality, sample sizes used in data quality audits, and error rate calculation. The survey was validated in the following manner: Content validity was achieved by developing and reviewing the survey with a group of subject matter experts. The survey was then pre-tested by an individual with skills and background similar to the anticipated respondents. Quantitative validation was approached through inclusion of internally redundant questions and questions where the information obtained could be compared for consistency. Our inter-question consistency was >90% for all internally redundant questions. Benchmarks with previous surveys are discussed throughout this paper.

The survey was sent to 1412 individuals selected as a convenience sample from the society mailing list. The response rate was 7% (93/1412 individuals) which is considered an artifact of the survey design. Respondents were asked to forward the survey to the individual in their organization responsible for the Data Quality function. This was done to discourage multiple responses from the same organization. Therefore, an organization with several SCDM members would have only 1 response.

We realize that some large organizations have practices that differ between locations due to mergers and acquisitions, or the different locations performing different types of work. Our design allowed for responses from separate locations while effectively discouraging multiple responses from the same organization. Therefore, we counted two responses from large organizations as coming from different locations. Seven of 93 responses were in this category (two responses from the same organization). Two of 93 had three responses, which were from organizations with only one location. These are counted as repeaters. One organization had more than three repeats. Our rate of duplicate organizations was 3%. The respondents were from the following industry sectors: CRO, Academia, Biotechnology, Pharmaceuticals, Medical Devices or Other (Table I).

Sector	Percent of Respondents
CRO	35%
Academia	4%
Biotech	19%
Pharmaceuticals	30%
Device	8%
Other	5%

Table I: Responding Organizations

Definition of Data Quality

The Institute of Medicine, IOM, defines data quality as “data that support conclusions and interpretations equivalent to those derived from error free data”.[6] Although this definition is precise and accurate, it is difficult if not impossible to implement, and requires simulations to do so.[7] Most likely, for this reason, definitions of data quality across the industry differ. We asked questions to identify whether or not organizations are quantifying data quality, using pre-defined acceptance criteria, and if so, the different type of the criterion that organizations use.

Continued on page 14

Data Quality Survey Results

Continued from page 13

Most organizations, about 70% of the respondents, perform CRF-to-database comparison audits as an indication of data quality. In addition, 65% of those who routinely perform CRF-to-database inspections quantify the results, for example, in an error rate. In many organizations, acceptance criterion are set, and a decision is made whether or not to lock a database based on the error rate being below the acceptance criterion. Sixty-nine percent of our respondents that perform CRF-to-database comparison audits use some type of pre-defined acceptance criterion. We asked questions to identify the distributions of different criterion that organizations used. The most popular acceptance criterion for overall database error rate were 0.10% and 0.50%, or 10 and 50 errors per 10,000 fields respectively (Figure 1). The acceptance criterion for critical data was bi-modal with 33% of the respondents using zero errors in critical fields, and 25% of the respondents using 0.10% or 10 errors per 10,000 fields.

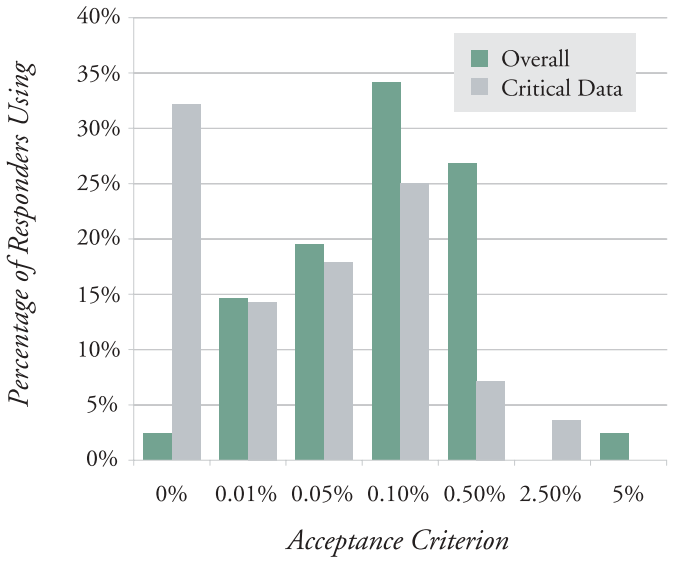


Figure 1: Acceptance Criterion for Critical Data and Whole Database

For our respondents, on average, a database would have sufficient data quality to lock if: (1) There were no errors in critical fields and less than 10 errors in 10,000 fields overall or (2) There were less than 10 errors per 10,000 fields in critical variables and less than 50 errors per 10,000 fields overall.

Simulations in the scientific literature[1] have shown that where edit check programs exist, error rates in critical variables (those used in programming the analysis) of up to 0.5%, or 50 errors per 10,000 fields caused loss of statistical power less than 2% in most cases examined, and less than 5% in all cases. Error rates of 0.1% or 10 errors in 10,000 fields resulted in less than 1% statistical power loss across all cases examined.[1] In a second simulation, for error rates ranging from 2.4% to 9.8%, the effective power loss ranged from 2%-6%.[10] These are conservative estimates because in the simulations, all of the errors were in the analysis field/s. A third simulation by Mullooly assessed effects of error rate on attenuation of the correlation coefficient, finding for all cases covered by the simulation, an error rate of 0.5% resulted in attenuations less than 6%.[11] The results varied in all simulations according to aspects such as extent of "Range Checking", study sample size and assumed size of the treatment effect. Please see the actual papers for details of the situations for which the results are applicable.

Audit Scope

A quality system audit encompasses all of the systems and procedures within an organization that assure a quality product. These include everything from an organization's quality policy and evidence that the organization has systems in place to implement the policy, to the quality control procedures implemented on each product. Quality system audits are often performed to certify vendors or as pre-award assessments. Process audits usually are an assessment that all of the procedures necessary for producing a quality product were performed appropriately. Process audits often include assessments of the quality of the actual product. In our case, the product is the CRF-to-database comparison audit.

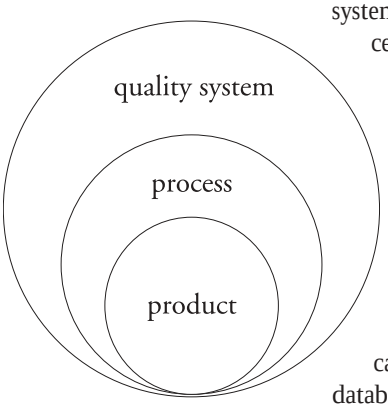


Figure 2: Quality System, Process and Product Audits

Continued on page 15

CALENDAR OF EVENTS

October 3-6, 2004	March 13-15, 2005	October 9-12, 2005	October 8-11, 2006	September 16-19, 2007
Fall Conference	2005 Spring Forum	2005 Fall Conference	2006 Fall Conference	2007 Fall Conference
Royal York Hotel	Grand Hyatt Buckhead	Sheraton San Diego Hotel & Marina	Wyndahm Palace Resort & Spa	Hyatt Regency
Toronto, Canada	Atlanta, GA	San Diego, California	Orlando, Florida	Chicago, Illinois

Data Quality Survey Results

Continued from page 14

In our industry, there appeared to be variability with respect to the type of audit performed on trials, product audits, process audits or some hybrid. To identify the industry practices in this area, we examined the components of a process audit, and asked respondents which of these components their organization performed on each trial.

From the responses, about 70% of organizations responding to the survey, perform a CRF-to-database comparison of each trial. Most (77%) organizations responding to this question, (51% of survey respondents) also include an assessment of the trial specific documentation including work instructions, data handling guidelines, data management plans, etc. to assure that the required documentation is present, compliant with organizational SOPs, and current.

	Usage Rate	
Process audit component	Percent of responders to question	Percent of responders to survey
CRF-to-database audit	100%	66%
Calculation of error rate for the trial	82%	54%
Audit of trial documentation to assure that it is present, compliant with organizational procedures and current	77%	51%
Validation documentation for trial specific programming	59%	39%
Training documentation	44%	29%
Vendor Audits for vendors performing CDM tasks	31%	20%
Compliance to industry standards, i.e. CDISC	20%	32%

Table II: Process Audit Component Usage Rate

To a lesser extent, organizations include other components of process audits like completeness of training documentation, vendor audits and compliance of trial data or procedures to industry standards.

On a more detailed level, audit scope is also a question of what data handling processes are included in CRF-to-database comparison audits. As discussed in the GCDMP[7] and elsewhere, there are multiple steps in most data handling processes. These include documenting a patient encounter in a medical record, transcription of the data onto a CRF or other trial specific data collection tool, data entry, identification and resolution of data discrepancies, and programming of databases and analysis datasets. These steps occur in paper-based clinical research. Each of these steps has the potential for error.[9] Depending on the scope of the CRF-to-database comparison, the errors associated with these steps may or may not be included in the error rate. We hypothesized that industry practices here differ also. For example, organizations performing National Cancer Institute, NCI, funded research compare the database to the source document for a representative sample of patients at that site, across all trials. These comparisons cover the majority of the data handling steps, and therefore have the broadest scope of the CRF comparison audits. Most industry funded research utilizes a CRF-to- database comparison with a narrower scope, omitting CRF transcription, but performed on a representative sample of patients from each trial.

To identify practices, we again, examined components of the data handling process including: CRF transcription, data entry, data cleaning, and programming. The respondents indicated the following data handling processes are performed during the quality control audit of each trial: Database Entry (84%); Query (78%), Programming (66%) and CRF Transcription errors (51%).

The majority of the responders actively included data entry and cleaning processes in their audits. Sixty-six percent of the respondents had procedures in place to identify programming errors during a data quality control audit. A surprisingly high, 51% of the respondents identified CRF transcription errors in their audits. It was not clear if these errors were "designed into database audit procedures" or whether the organizations had other procedures in place to compare the database or CRF to the source documents at clinical investigational sites. In addition, organizations may have considered the source document verification step usually performed by monitors or Clinical Research Associates as the assessment of CRF transcription errors. However, it is clear that the majority of respondents who perform CRF-to-database audits include data entry, cleaning, and database programming in the scope of their audit.

Sample Selection

Audit sample selection includes both what data is audited as well as the sample size considerations. A question often asked in the industry is, "Do you audit entire patient datasets, selected variables or both?" This is an area where practices were unclear. For example, was all data for selected patients audited, was a subset of a patient's data selected, or did practices include both methods, (e.g., all data on a certain percentage of patients followed by critical variables on an additional percentage). To more clearly understand data selection practices, we asked, "What type of quality inspections does your company use?" There were four mutually exclusive choices, entire patients only, selected variables, combination of selected variables and entire patients, or none of the above.

From the responses, it is clear that a majority of the organizations select data for audits as a combination of patients where all of the data are audited, and a percentage of patients where only selected variables, for example critical variables, are audited.

Continued on page 16

Data Quality Survey Results

Continued from page 15

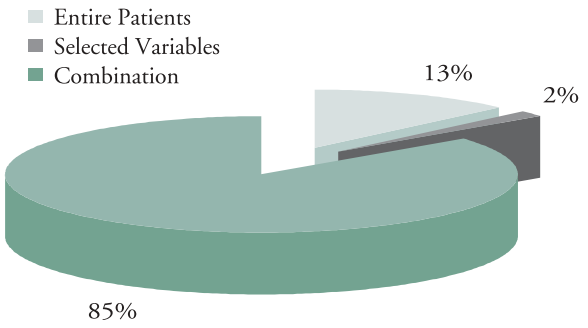


Figure 3: Data Selection Methods

There are many ways sample sizes for data quality audits are calculated.[7] Some of the more popular methods described in the GCDMP are a fixed percent, fixed percent within a valid range, square root of n+1 patients, square root of n+1 fields, standard sampling tables, and having a statistician calculate the sample size for audits. When asked how they arrived at their sample size, 79 % of the respondents use a fixed percent sample size. Half of those qualify the fixed percent sample size by citing a range of database sizes over which the fixed percent sample size is applicable (i.e., there is adequate power for decision- making based on the audit results, and/or the sample size does not require unnecessary audits in the cases of large databases).

Method for Arriving at Sample Size	Percentage of Respondents to the question
Fixed %	42%
Fixed % with Range	37%
SQRT (n patients + 1)	19%
SQRT (n values + 1)	2%
Standard Sampling Tables	0%
Statistician Calculates	13%
Other	0%

Table III: Method for Arriving at Sample Size

Reportedly, there are differences in industry practices regarding sample sizes by phase of study, and between regulated and non-regulated research. To understand the practices, we asked respondents to indicate their sample size by phase of research. In an effort to assess comparability of industry versus other funding sources, i.e. government, professional society research not run under an IND or IDE, we added a government funded category and a Registry / not regulated category.

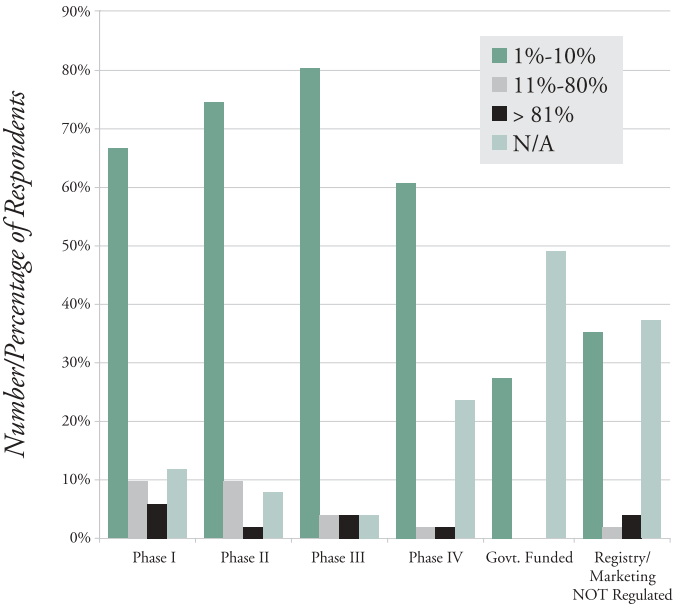


Figure 4: Sample Size by Phase

Less than 10% of the respondents use a sample size above 10%. In addition, the number of respondents auditing over 10% was non existent for government funded studies. A higher number of respondents audited a larger percentage of the data for Phase I and Phase II trials than for Phase III, Phase IV and non-regulated studies.

Because many organizations audit selected variables on part or on their entire audit sample, we also assessed the sample size for critical versus non-critical variables by phase. The responses showed that a majority of the organizations audited from 1-10% of their non-critical data and >81% of their critical variables for all phases and categories of studies.

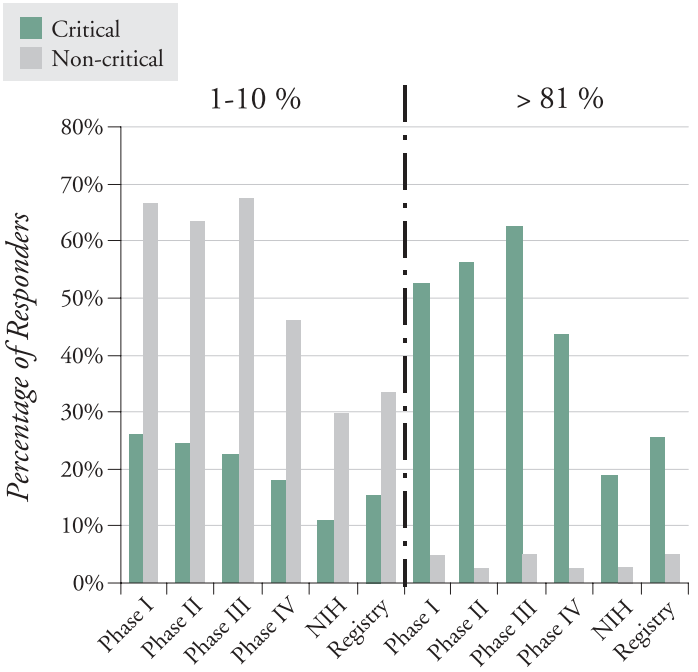


Figure 5: Critical/Non-critical

Continued on page 17

Data Quality Survey Results

Continued from page 16

Based on the responses, a common scenario for auditing data is to sample 1-10% of the patients reviewing all of their data, followed by a manual review of just the critical variables on 80 or more percent of the patients. Although a manual review of "more than a sample" of any data point is more of a data cleaning activity, it is being performed by an audit group.

The Numerator, The Denominator and the Error Rate:

Of the organizations that perform CRF-to-database audits, 65%, said that they quantified the results of their quality inspections. There are several components to quantifying the data quality that vary across the industry. They are definition of an error, method of arriving at the denominator and method of calculating the error rate. To assess the industry variability in these components, we provided an example data listing, with errors identified and asked respondents to calculate their numerator, denominator, and error rate. The example data listing was from two "tables" VITALS and DEMOG. The example and directions appear in Table IV.

Vital Signs Table/Dataset

Protocol	RecID	PtID	SiteID	Visit	SYSBP	DIABP
CHECK	XG15	32	001	1	122	55
CHECK	PB53	56	001	2	140	X
CHECK	AR36	29	003	1	153	75
CHECK	AD95	54	004	1	135	88
CHECK	TP62	16	002	2	115	69

Demography Table/Dataset

Protocol	RecID	PtID	SiteID	Visit	DOB	Age	GEND	ETHN
CHECK	XG15	32	001	1	2/7/61	42	1	2
CHECK	AR36	29	003	1	11/15/58	45	2	4
CHECK	AD95	54	004	1	4/15/42	61	2	
CHECK	AD96	16	002	1	1/2/73	31	1	2

- Protocol and RecID are system generated variables.
- PtID and Site ID are entered once for each patient and propagated throughout the database.
- Visit is saved in the database according to the visti that the entry operator selects.
- Age is calculated based on DOB and Enrollment date.
- Null fields or blanks are confirmed as being left blank by sites on the CRF.
- SYABP and DIABP are data entered from the CRF.
- The X is a field that was erroneously left blank during Data Entry.
- The green text fields are confirmed data entry errors.

Table IV: Data Tables used for Error Rate Survey Questions

Continued on page 18

Data Quality Survey Results

Continued from page 17

Calculation of the numerator depends on an organization's definition of an error. The results in Table V below show the percent of respondents to the question who chose each "numerator" from a mutually exclusive list of six possibilities and confirm the variability in the methods.

Number of Errors Counted	Percent of responders
1 error	0%
2 errors	30%
3 errors	59%
4 errors	8%
5 errors	3%
other	16%

Table V: Number of Errors Counted

The survey question was designed to identify variability in the counting methods. The errors created in the tables and the directions / additional information provided scenarios of different counting methods. For example the patient identifier was entered once per each patient, and marked as a field confirmed as "in error". Some organizations count all the data entered under the patient as "in error" if the data were entered under the wrong patient number or if the patient number itself was entered erroneously. Other organizations count one error for this type of occurrence. The entered field "DOB" was "in error". Some organizations count this as one error, while others count it as two because the derived age would also not reflect the actual age of the patient. The field, "DIABP" was described as erroneously left blank during data entry, some organizations count this as an error just because it is missing. We did not design this question to definitively identify counting rules, and couldn't because of the impact of possible errors made in taking the survey itself, and because of the number of questions we would have to have asked. We did, however confirm that different counting rules were used by the respondents and these results indicate that industry practices for counting rules vary.

Likewise, we hypothesized that there is a large amount of variability in the method of counting the fields in the denominator for the error rate. Examples of the possible differences in the survey question were, not counting system derived variables like "PROTOCOL" and the record identifier in the denominator at all, counting them only once per database, once per patient, or once per table. Another possibility was not counting null fields. Differently counting fields entered once per visit, but stored in all tables was also a possibility. This question was not designed to differentiate between the specific counting rules, but was designed to detect variability in counting rules among the respondents, if present. There was not a majority of respondents that chose any method.

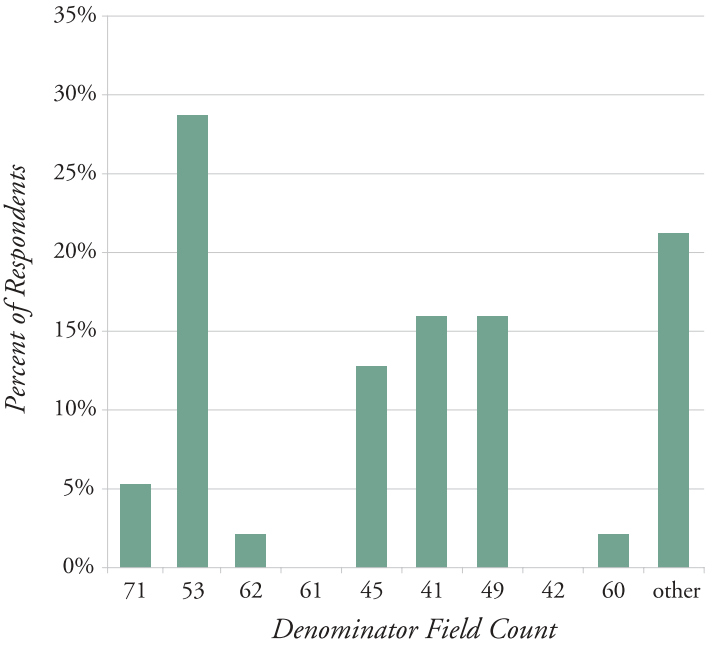


Figure 6: Variability in Field Counting Rules (error rate denominator)

Some of the variety in the industry originates in the counting of system generated fields, fields that are entered once for a patient or patient visit, and derived variables. For example, PROTOCOL and RecID are system generated variables. They can be counted once per patient, once per visit, once per study, or not at all. The patient identifier can be counted once for the study, once per visit, or not at all. The visit identifier, which for this question was "saved in the database according to the visit the entry operator chose to enter the data under" could be counted once per visit, once per record, or not at all since it was not an entered variable. Some of the methods do not significantly change the denominator, and thus the error rate. However, differences between whether or not to count the system generated variables or nulls can alter the error rate by a factor of two or more.

If differences in the error counting rules and differences in the denominator calculation are present, error rates using the different methods will not be comparable. Within this survey, the error rates calculated by different organization ranged from 1% to 7% and depending on the counting rules used, may or may not be comparable. In a worst case scenario combination of numerator and denominator choices in the survey, the error rates for the data shown in the example would have differed by a factor of seven.

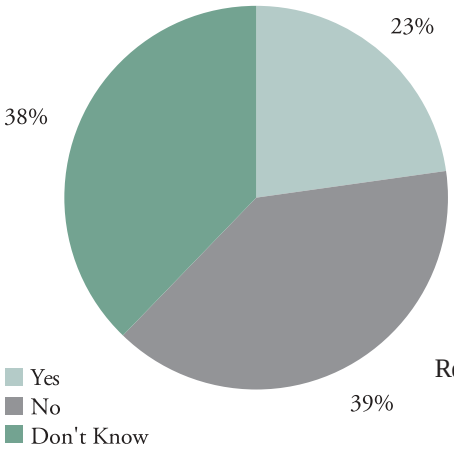
Reporting

There are many ways to report the results of a data quality audit. One way is to note discrepancies and provide a listing of those to the process owner. Although the listing facilitates correction of the errors, it provides no information on the error rate, and does not facilitate decision-making regarding the database, and fitness of data for analysis. A second level of reporting is to summarize the discrepancies as an error rate and report that to the process owner. Ninety-four percent of the respondents produce a report of the audit results.

Continued on page 19

Data Quality Survey Results

Continued from page 18



The highest level of reporting is to provide the database error rate in the Clinical Study Report for those reviewing the data. Only 23% of the respondents provided their database error rate in the Clinical Study Report (Figure 7).

Figure 7: Communication of Error Rate in Clinical Study Report

In conclusion, this survey represents the current practices of the 93 respondents. Although survey results may not be generalizable to the Clinical Trials industry, it is our hope that providing a sampling of industry practices is helpful to our membership.

References

1. *Mcentegart, Damian J., Jadhav, Simon P., Brown, Tanya, Channon, Edward J., Checks of Case Record Forms Versus the Database for Efficacy Variables when Validation Programs Exist. Drug Information Journal, Vol. 33, pp.101–107, 1999.*
2. *DataBasics, Society for Clinical Data Management, 1997.*
3. *International Conference on Harmonization. ICH Expert Working Group: Guideline for Good Clinical Practice, E6, Section 5.1.3; 1996. Available from the US Food and Drug Administration.*
4. *Code of Federal Regulations, Title 21, Part 11, Electronic Records; Electronic Signatures. Food and Drug Administration. US Government Printing Office.*
5. *Guidance for Computerized Systems Used in Clinical Trials, Food and Drug Administration, April 1999.*
6. *Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making, Workshop Report, Institute of Medicine, National Academy Press, Washington D.C., 1999.*
7. *GCDMP Version 3.0, Society for Clinical Data Management, 2003.*
8. *Anderson, Pam, Wortman, Bill, CQA Primer, Quality Council of Indiana, 1997.*
9. *Helms, Ronald W. PhD. Data Quality Issues in Electronic Data Capture. Drug Information Journal, Vol. 35, 2001.*
10. *Day, Simon, Fayers, Peter, Harvey, Derek. Double Data Entry: What Value, What Price? Controlled Clinical Trials, Vol. 19, pp. 15-24, 1998.*
11. *Mullooly, John P. The Effects of Data Entry Error: An Analysis of Partial Verification. Computers and Biomedical Research, Vol. 23, pp. 259-267, 1990.*

Please contact info@scdm.org if you have questions about registration for upcoming meetings, advertising, renewal of membership, or if you need to provide updated mailing/contact information.

Society for Clinical Data Management, Inc.
555 E. Wells Street
Suite 1100
Milwaukee, WI 53202-3823
Phone: 414-226-0362
Fax: 414-276-3349
E-mail: info@scdm.org

Visit us at
www.scdm.org

92-0504-002

Society for Clinical Data Management, Inc.
555 E. Wells Street
Suite 1100
Milwaukee, WI 53202-3823

FIRST CLASS PRSRT
U.S. POSTAGE
PAID
PERMIT NO. 2066
EAU CLARIE, WI 54701