Web-based Electronic Data Capture (EDC) Implementation and Study Start-up

David Eade,[a] Maxine Pestronk,[b] Ralph Russo,[c] Muthamma Muthanna,[d] Derek Johnson,[e] Denise Redkar-Brown,[f] Olivia Montaño,[g] Shweta Kerkar,[h] Meredith Zozus,[i]

a: United BioSource Corporation Kansas City, MI; b: Independent, Cambridge, MA; c: Pfizer, New York, NY; d: Accenture, Bangalore, India; e: Inflamax Research, Fargo, ND; f: Independent, Elkton, MD; g: PROS Foundation, San Diego, CA; h: Premier Research, Houston, TX; i: University of Texas Health Sciences Center, San Antonio, TX

1) Abstract

Web-based Electronic Data Capture (EDC) is a mainstay of form-based data collection and management in clinical studies. This chapter reviews the implementation and start-up tasks for clinical studies using web-based EDC for form-based data collection and management (hereafter EDC). Topics covered include designing, developing, testing, and implementing workflow and data flow in clinical studies using EDC systems. Topics of focus include data collection, workflow and data flow associated with data collection and processing, data processing such as exchange, integration, cleaning and coding in EDC systems, implementation at study sites including training and account management, and working with EDC system vendors. The chapter emphasizes common responsibilities of Clinical Data Management (CDM) professionals in the implementation of an EDC application (study).

After reading this chapter, the reader should understand:

- the regulatory basis for practices in EDC implementation and study start-up
- similarities and differences between paper and web-based data collection
- basic and common features of fields, forms and form groupings in EDC systems
- common dynamic workflow and data flow options within web-based EDC systems
- special considerations for data processing when using web-based EDC
- common steps in system set-up and testing
- methods for managing system access and privileges
- common practices for training clinical investigational sites in the use of EDC
- considerations and business models for using vendor-hosted EDC systems

2) Introduction

In the first clinical studies, data were collected on paper forms called Case Report Forms or CRFs. The structured forms served to assure complete and consistent data collection for each study participant. Since the early 1990s, we have documented best practices for designing paper CRFs. (Spilker 1991, Kennedy 2002, McFadden 2007, Avey 2000) Design considerations focused on graphical lay-out within the confines of a paper page CRFs and visual cues to aid the form filler such as use of boxes versus circles to indicate 'check all that apply' versus 'check one'. Instructions were printed on the forms and there were rules governing the type of writing instruments such as black ballpoint pens and rules for correcting data such as use of a single

line cross out and providing the corrected value, the date, and the initials of the person making the change. Yet there was nothing to facilitate the workflow of data collection or to prevent writing discrepant or errant values on the paper form.

In the days of paper-based data collection, how the data were entered into electronic format, typically in a Clinical Database Management Systems (CDMS), was not the most important consideration. Clinical study data were usually double data entered. Entry operators were trained in handling exceptions and allowable corrections. Clinical data management systems automated the workflow of data entry, integration of external data, cleaning and coding, and provided automation for tracking data entry, discrepancy identification, and discrepancy resolution. However, the benefits of these advances were largely limited to in-house data management groups in Sponsor organizations.

At the turn of the century, technology leveraging the internet such as web-based EDC systems opened the possibility of extending the benefits of information systems in clinical research to multiple roles on research teams, investigational sites, and study participants. Commercial and academic advances were sporadic. (Helms 2001) EDC systems shifted the work of data entry to clinical investigational sites and contained less data management functionality than the Clinical Data Management Systems of the time. For example, integrating external data and medical coding were challenging or altogether not supported in early EDC systems.

There are four ways through which information systems can add value to organizations: automation, connectivity, decision support, and data mining. (Stead and Lin 2009) Today's EDC functionality offers some of this potential. However, achieving broad benefit across clinical study design, conduct, and reporting for clinical investigators, research teams, and study participants is dependent on how EDC functionality is leveraged for each clinical study and how the data are used for better decision-making within and across studies. (Kush 2003) This chapter focuses on realizing value from EDC functionality through design and implementation of workflow and data flow within clinical studies.

3) Scope

This chapter provides information on the design, development, and implementation concepts related to setting-up a study (application) in an EDC system. Practices, procedures, and recommendations are proposed for clinical data managers to design and implement EDC facilitated workflow and data flow for automation, connectivity, decision support, and data mining within and across clinical studies.

While many of the tasks described in this chapter may be joint responsibilities between different functional areas of an organization, those tasks associated with the collection, processing and storage of data are covered here. These responsibilities are the core of the Clinical Data Management profession. As such, the clinical data manager is usually responsible for the overall implementation of any study application.

Recommendations for EDC system selection were covered in the chapter "Electronic Data Capture Selecting an EDC System". Recommendations for study conduct and study closeout using EDC are addressed in the chapter “Electronic Data Capture – Study Conduct, Maintenance and Closeout”.

4) Minimum Standards

As a mode of data collection and management in clinical studies, EDC systems have the potential to impact human subject protection as well as the reliability of trial results. Regulation and guidance are increasingly vocal on the topic. The **E6(R2) Good Clinical Practice: Integrated Addendum** contains several passages particularly relevant to use of EDC systems in clinical studies.

Section 2.8 “Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective tasks.”

Section 2.10, “All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification.”

Section 5.0 states that, “The methods used to assure and control the quality of the trial should be proportionate to the risks inherent in the trial and the importance of the information collected.”

Section 5.1.1 states that “The sponsor is responsible for implementing and maintaining quality assurance and quality control systems with written SOPs to ensure that trials are conducted and data are generated, documented (recorded), and reported in compliance with the protocol, GCP, and the applicable regulatory requirement(s).” Section 5.1.3 states that, “Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly.”

Section 5.5.1, “The sponsor should utilize appropriately qualified individuals to supervise the overall conduct of the trial, to handle the data, to verify the data, to conduct the statistical analyses, and to prepare the trial reports.”

Section 5.5.3 “When using electronic trial data handling and/or remote electronic trial data systems, the sponsor should, a) Ensure and document that the electronic data processing system(s) conforms to the sponsor's established requirements for completeness, accuracy, reliability, and consistent intended performance (i.e., validation).”

Section 5.5.3 addendum “The sponsor should base their approach to validation of such systems on a risk assessment that takes into consideration the intended use of the system and the potential of the system to affect human subject protection and reliability of trial results.” and in the addendum b) states the requirement, “Maintains SOPs for using these systems.”

Section 5.5.3 addendum c-h introductory statement states that, “The SOPs should cover system setup, installation, and use. The SOPs should describe system validation and functionality testing, data collection and handling, system maintenance, system security measures, change control, data backup, recovery, contingency planning, and decommissioning.”

Section 5.5.4 under Trial Management, Data Handling and Recordkeeping, states that, “If data are transformed during processing, it should always be possible to compare the original data and observations with the processed data.”

Similar to ICH E6 R2, **Title 21 CFR Part 11** also states requirements for traceability, training and qualification of personnel, and validation of computer systems used in clinical trials. Requirements in 21 CFR Part 11 Subpart B are stated as controls for closed systems (21 CFR Part 11 Sec. 11.10), controls for open systems (21 CFR Part 11 Sec. 11.30), Signature manifestations (21 CFR Part 11 Sec. 11.50), Signature/record linking (21 CFR Part 11 Sec. 11.70). Requirements for electronic signatures are stated in in 21 CFR Part 11 Subpart C.

Recommendations in *Section A* of the 2007 **Guidance for Industry Computerized Systems Used in Clinical Investigations** state that, “Each specific study protocol should identify each step at which a computerized system will be used to create, modify, maintain, archive, retrieve, or transmit source data.”

Section B of the CSUCI guidance states expectations with respect to Standard Operating Procedures (SOPs); “There should be specific procedures and controls in place when using computerized systems to create, modify, maintain, or transmit electronic records, including when collecting source data at clinical trial sites” and that, “the SOPs should be made available for use by personnel and for inspection by FDA.”

Section C reiterates document retention requirements under 21 CFR 312.62, 511.1(b)(7)(ii) and 812.140. Further, *section C* of CSUCI goes on to state that, “When source data are transmitted from one system to another …, or entered directly into a remote computerized system … or an electrocardiogram at the clinical site is transmitted to the sponsor’s computerized system, a copy of the data should be maintained at another location, typically at the clinical site but possibly at some other designated site.” And that “copies should be made contemporaneously with data entry and should be preserved in an appropriate format, such as XML, PDF or paper formats.”

Section D further specifies 21 CFR Part 11 principles with respect to limiting access to CSUCT, audit trails, and date and time stamps.

Section E likewise provides further detail regarding expectations for security, e.g., “should maintain a cumulative record that indicates, for any point in time, the names of authorized personnel, their titles, and a description of their access privileges” and recommends that, “controls be implemented to prevent, detect, and mitigate effects of computer viruses, worms, or other potentially harmful software code on study data and software.”

Section F addresses direct entry of data including automation and data standardization; data attribution and traceability including explanation of, "how source data were obtained and managed, and how electronic records were used to capture data"; system documentation that, identifies software and hardware used to, "create, modify, maintain, archive, retrieve, or transmit clinical data"; system controls including storage, back-up and recovery of data; and change control of computerized systems.

Section G addresses training of personnel as stated in 21 CFR 11.10(i) that those who, "develop, maintain, or use computerized systems have the education, training and experience necessary to perform their assigned tasks", that training be conducted with frequency sufficient to, "ensure familiarity with the computerized system and with any changes to the system during the course of the study" and that, "education, training, and experience be documented".

The Medicines & Healthcare products Regulatory Agency (MHRA) **'GXP' Data Integrity Guidance and Definitions** covers principles of data integrity, establishing data criticality and inherent risk, designing systems and processes to assure data integrity, and also covers the following topics particularly relevant to EDC.

Similar to ICH E2 (R2), MHRA *Section 2.6* states that, "Users of this guidance need to understand their data processes (as a lifecycle) to identify data with the greatest GXP impact. From that, the identification of the most effective and efficient risk-based control and review of the data can be determined and implemented."

Section 6.2, Raw Data states that, "Raw data must permit full reconstruction of the activities."

Section 6.7 Recording and Collection of Data states that, "Organisations should have an appropriate level of process understanding and technical knowledge of systems used for data collection and recording, including their capabilities, limitations, and vulnerabilities". And that, "The selected method [of data collection and recording] should ensure that data of appropriate accuracy, completeness, content and meaning are collected and retained for their intended use." And further, that, "When used, blank forms … should be controlled. … [to] allow detection of unofficial notebooks and any gaps in notebook pages."

Section 6.9 Data Processing states that, "There should be adequate traceability of any user-defined parameters used within data processing activities to the raw data, including attribution to who performed the activity." And that, "Audit trails and retained records should allow reconstruction of all data processing activities…"

The General Principles of Software Validation; Final Guidance for Industry and FDA Staff (2002) states guidance regarding proper documentation expected of software utilized in a clinical trial.

Section 2.4 "All production and/or quality system software, even if purchased off-the-shelf, should have documented requirements that fully define its intended use, and information against which testing results and other evidence can be compared, to show that the software is validated for its intended use."

*Section 4.7 (Software Validation After a Change), "*Whenever software is changed, a validation analysis should be conducted not just for validation of the individual change, but also to determine the extent and impact of that change on the entire software system"

*Section 5.2.2 "*Software requirement specifications should identify clearly the potential hazards that can result from a software failure in the system as well as any safety requirements to be implemented in software."

Good Manufacturing Practice Medicinal Products for Human and Veterinary Use (Volume 4, Annex 11): Computerised Systems (2011) provides the following guidelines when using computerized systems in clinical trials. Though the guidance is in the context of manufacturing, it is included to emphasize the consistency of thinking and guidance relevant to use of computer systems in clinical trials across the regulatory landscape.

Section 1.0 "Risk management should be applied throughout the lifecycle of the computerised system taking into account patient safety, data integrity and product quality. As part of a risk management system, decisions on the extent of validation and data integrity controls should be based on a justified and documented risk assessment of the computerised system."

Section 4.2 "Validation documentation should include change control records (if applicable) and reports on any deviations observed during the validation process."

Section 4.5 "The regulated user should take all reasonable steps, to ensure that the system has been developed in accordance with an appropriate quality management system."

Section 7.1 "Data should be secured by both physical and electronic means against damage. Stored data should be checked for accessibility, readability, and accuracy. Access to data should be ensured throughout the retention period."

Section 7.2 "Regular back-ups of all relevant data should be done. Integrity and accuracy of backup data and the ability to restore the data should be checked during validation and monitored periodically."

Section 9.0 "Consideration should be given, based on a risk assessment, to building into the system the creation of a record of all GMP-relevant changes and deletions (a system generated "audit trail"). For change or deletion of GMP-relevant data the reason should be documented. Audit trails need to be available and convertible to a generally intelligible form and regularly reviewed."

Section 10.0 "Any changes to a computerised system including system configurations should only be made in a controlled manner in accordance with a defined procedure."

GAMP 5**: A Risk-based Approach to Compliant GxP Computerized Systems (2008)** suggests scaling activities related to computerized systems with a focus on patient safety, product

quality and data integrity. It provides the following guidelines relevant to GxP regulated computerized systems including systems used to collect and process clinical trial data:

Section 2.1.1 states that, "Efforts to ensure fitness for intended use should focus on those aspects that are critical to patient safety, product quality, and data integrity. These critical aspects should be identified, specified, and verified."

Section 4.2 "The rigor of traceability activities and the extent of documentation should be based on risk, complexity, and novelty, for example a non-configured product may require traceability only between requirements and testing."

Section 4.2 "The documentation or process used to achieve traceability should be documented and approved during the planning stage, and should be an integrated part of the complete life cycle."

Section 4.3.4.1 "Change management is a critical activity that is fundamental to maintaining the compliant status of systems and processes. All changes that are proposed during the operational phase of a computerized system, whether related to software (including middleware), hardware, infrastructure, or use of the system, should be subject to a formal change control process (see Appendix 07 for guidance on replacements). This process should ensure that proposed changes are appropriately reviewed to assess impact and risk of implementing the change. The process should ensure that changes are suitably evaluated, authorized, documented, tested, and approved before implementation, and subsequently closed."

Section 4.3.6.1 "Processes and procedures should be established to ensure that backup copies of software, records, and data are made, maintained, and retained for a defined period within safe and secure areas."

Section 4.3.6.2 "Critical business processes and systems supporting these processes should be identified and the risks to each assessed. Plans should be established and exercised to ensure the timely and effective resumption of these critical business processes and systems."

Section 5.3.1.1 "The initial risk assessment should include a decision on whether the system is GxP regulated (i.e., a GxP assessment). If so, the specific regulations should be listed, and to which parts of the system they are applicable. For similar systems, and to avoid unnecessary work, it may be appropriate to base the GxP assessment on the results of a previous assessment, provided the regulated company has an appropriate established procedure."

Section 5.3.1.2 "The initial risk assessment should determine the overall impact that the computerized system may have on patient safety, product quality, and data integrity due to its role within the business processes. This should take into account both the complexity of the process, and the complexity, novelty, and use of the system."

The FDA guidance, **Use of Electronic Health Record Data in Clinical Investigations**, emphasizes that data sources should be documented and that source data and documents be retained in compliance with 21 CFR 312.62(c) and 812.140(d).

Section V.A states that, "Sponsors should include in their data management plan a list of EHR systems used by each clinical investigation site in the clinical investigation" and that, "Sponsors should document the manufacturer, model number, and version number of the EHR system and whether the EHR system is certified by ONC".

Section V.I states that, "Clinical investigators must retain all paper and electronic source documents (e.g., originals or certified copies) and records as required to be maintained in compliance with 21 CFR 312.62(c) and 812.140(d)".

Similarly, the FDA's guidance, **Electronic Source Data Used in Clinical Investigations** recommends that all data sources at each site be identified.

Section III.A states that, each data element should be associated with an authorized data originator and goes on to state, "A list of all authorized data originators (i.e., persons, systems, devices, and instruments) should be developed and maintained by the sponsor and made available at each clinical site. In the case of electronic, patient-reported outcome measures, the subject (e.g., unique subject identifier) should be listed as the originator."

Section III.A.3 elaborates on Title 21 CFR Part 11 and states, "The eCRF should include the capability to record who entered or generated the data [i.e., the originator] and when it was entered or generated." and "Changes to the data must not obscure the original entry, and must record who made the change, when, and why."

Section III.A.5 states that the FDA encourages "the use of electronic prompts, flags, and data quality checks in the eCRF to minimize errors and omissions during data entry".

Section III.C states that, "The clinical investigator(s) should retain control of the records (i.e., completed and signed eCRF or certified copy of the eCRF)." In other words, eSource data cannot be in sole control of the sponsor.

As such, we state the following minimum standards for the implementation and study start-up using EDC systems.

- Document requirements for all aspects of the eCRF and data collected, processed, or stored by or in the EDC system.
- Document sources of data at each site including explicit statement that the EDC system is used as the source where this is the case.
- Ensure data values can be traced from the data origination through all changes and that the audit trail of all data changes is immutable, preserved, and available for review.
- Use electronic "prompts, flags, and data quality checks in the eCRF to minimize errors and omissions during data entry".

- Establish and follow SOPs to ensure that testing including user acceptance testing (UAT) of the study-specific EDC application is commensurate with the assessed risk.
- Establish and follow SOPs to ensure that testing is completed and documented prior to implementation and deployment to sites.
- Establish and follow SOPs to ensure that all users have documented training prior to using the system.
- Establish and follow SOPs to limit data access and permissions to authorized individuals and to document data access and permissions.
- Establish and follow SOPs for the process of setting up an EDC system for a study.

5) Best Practices

- Develop eCRFs with cross-functional teams including clinical operations, monitoring, clinical data management, statistics, regulatory affairs, quality assurance, pharmacovigilance/drug safety, and medical leadership. [VI]
- Ensure adequate attention to the collection, processing, and routing of safety data. [VI]
- Develop the eCRF and edit checks concurrently; finalize eCRF specifications prior to finalization of edit check specifications. [VI]
- Ensure that all data required for study-related safety decision making at the clinical investigation site (e.g. toxicity management) are available to sites within the EDC system. [VI]
- Ensure the eCRF design is intuitive and user-friendly for data entry and that instructions and references are readily available. [VI]
- Ensure eCRFs do not introduce bias into the data by containing leading questions or forcing responses. [VI]
- Ensure that comprehensive help (written, live, or otherwise) including eCRF entry guidelines, study data definition, and dynamic functionality behavior for all fields/forms/visits are up-to-date and readily available to sites. [VI]
    - Help should be available during work days and times of all regions included in the study.
    - Help should support the number of languages including local dialects needed to communicate with all EDC system users.
- Consider use of available data standards. [VI]
- Where data standards are used, ensure that the eCRF conforms to the standards so that detail (information content) is not lost in downstream mapping to such standards for submission or data sharing. [VI]
- Ensure site personnel have reliable access to the internet and the EDC application. [VI]
- Ensure that system users are informed in advance of anticipated down time such as planned system maintenance and that procedures and alternate contact information are in place for quickly communicating unplanned outages. [VI]
- Ensure that alternate procedures and tools are in place for maintaining critical study operations during EDC system downtime. [VI]

6) What it Means to Design a Study Application Within an EDC System

Filling in an electronic form is quite different than completing a paper form, where the advantages of EDC technology are leveraged. For example, in an interventional cardiology study, if a transfusion is entered, a new form is conditionally generated to collect hemoglobin and hematocrit values as well as assessment information for peri-procedural bleeding. At the same time, an email notification is automatically sent to the study safety desk. On-screen checks are run to flag out-of-range and logically inconsistent lab values, and the investigator's assessment of relatedness to the study drug is required; a discrepant data flag is attached to the form until the investigator's assessment of relatedness is populated. In this simple scenario, the EDC system added new forms relevant to the patient, provided greater control of data entered in the form, facilitated study workflow, automated tracking of discrepant data, and decreased the gap between the site and central team managing the study – all in real-time. This functionality is not available when collecting data on paper forms and due to the time-lag in processing and entering paper forms, immediate action by the central study team is not possible. Thus, in this simple example, the EDC system as implemented for the study provided significant value over paper data collection through automation, connectivity, and decision support.

EDC systems offer the opportunity to define and enforce workflow of data collection in addition to the data to be collected. While the extent to which workflow and data flow can be automated within an EDC system depends on the functionality offered by each system, the basic functionality described in the interventional cardiology study example is available in most EDC software.

It is the increase in connectivity, work, and data flow automation and decision support that makes use of EDC different from collecting data on paper forms. An EDC-enabled study goes beyond implementation of new technology to re-engineer processes and improve decision-making during study conduct. (Kush 2003) For this reason, setting-up a study in an EDC system is referred to as "*building a study*" rather than creating a study database. Using EDC means that for each data value collected on an eCRF, there is an additional choice of what if anything the system should do in response to entry of each of the possible entry values. In today's EDC systems, there are often many options. Thus, the data manager should have a thorough understanding of workflow and data flow automation and decision support in addition to EDC system functionality to optimize data-related aspects of study conduct. [VI]

How a human works with a paper form and writing instrument is different from working with a computerized system and related input devices. Because EDC is often more invasive than paper data collection in individual working practices and institutional process flow, these interactions become important in design, testing and implementation. While professionals field-tested paper forms prior to their use on studies, true human centered design, usability testing, or implementation monitoring are often appropriate if not necessary in use of EDC.

7) eCRF Design

Most study builds start with the static aspects of the electronic CRFs (eCRFs), i.e., the data elements or fields to be collected, their definition, valid response values, lay-out on the screen, and their organization into forms and visits. EDC systems facilitate different ways of grouping of data to be collected or displayed on eCRFs. Such groupings include grouping of data collection fields into modules, modules into forms, and forms into visits similar to such groupings on paper data collection forms. (Figure 1) Likewise, the fields in modules may differ from form to form and the contents of forms may differ from visit to visit. Just as with paper data collection forms such module-to-module, form-to-form, or visit-to-visit variability increases the development cost and must be weighed against ease of use and data quality. [VI] In EDC systems, grouping of data elements into modules, forms, or visits may be part of data definition and affect data storage as well as and layout. Thus, eCRF design requires a thorough understanding of the relationship between data definition, grouping, layout, and data storage structure in the specific EDC system. [VI]

When designing an eCRF, it is often not known what type of computer(s) will be used for data entry by the end-user. The size and type of screen and input devices such as keyboard and mouse can easily differ across users and at the same time can affect the data entry process. For example, fields “below the scroll” may be more easily missed. Many EDC systems have the capability of allowing for longer or wider forms, as well as multiple ‘forms’ within one eCRF presentation. However, it is good practice to take into consideration the smallest screen available on the market when deciding grouping and lay-out of data fields on eCRF screens even if doing so may reduce the amount of data collected on a single eCRF screen. [VI] Static aspects of eCRF design that are not specific to EDC and are covered in the CRF design chapter of the GCDMP. There are, however, static aspects of eCRF design that are specific to EDC. These are covered in subsequent sections of this chapter.

A well-designed eCRF should assist site staff with study conduct including data collection. Most EDC systems have functionality that can be leveraged to guide site staff in data entry and actions to be taken in response to entered data. The aspects of the EDC systems that pertain to the EDC system response to data entered by the user and other user actions taken within the EDC system are collectively called dynamic behavior. Dynamic behavior is apparent in workflow and data flow implemented in and facilitated by the EDC system.

Dynamic behavior can be triggered by an individual field or by some relationship between multiple fields or forms. Further, dynamic behavior may act on an individual field or multiple fields. Consider handling of dates. A site in Europe may prefer entering dates using the “dd-mmm-yyyy” format, whereas a site in the United States may prefer using “mmm-dd-yyyy”. Some EDC systems allow site or user-specific settings so that a user can enter dates in their preferred format and the system converts and stores the data in a standard date format. Further dynamic behavior pertaining to dates includes functionality to prevent or facilitate entry of partial dates that may include alerts to the user, or flagging values within the database and alternate processing of flagged partial dates. In the date examples, the dynamic behavior is

triggered by and acts on single fields (the dates). In this case, the dynamic behavior includes facilitating different entry formats, how the data will be processed by the EDC system, and the workflow and data flow associated with entry of valid values and exceptions. The bleed example above is an example of dynamic behavior triggered by indication of a transfusion (an individual field) and acting on multiple fields through creation of a new bleed form with fields for bleed-related lab values and bleed assessment details. In the bleed example, the dynamic behavior includes triggering a new form in response to an entered data value. Thus, when designing an eCRF, the designer evaluates each field for whether it is a trigger for dynamic behavior either alone or in concert with other fields and whether the desired behavior pertains to an individual field or to multiple fields. [VI] Dynamic aspects of EDC study builds are covered in subsequent sections of this chapter.

8) Basic Form Features

The most basic function of EDC software is the ability to build and deploy web-based electronic forms for the entry of data and to store the entered data. In most EDC systems, data elements are associated with a data collection structure when they are first added to the system. Every time the data element is implemented as a field in a form, the associated data collection structure is used, standardizing it throughout the study. Common data collection structures in EDC systems include free text, many options for semi-structured text, radio buttons, dropdown lists, and checklists. The use of pre-defined answer choices such as those in radio buttons, checklists, and dropdown lists provides constraints during entry and along with on-screen edit checks are associated with higher data quality. (Zozus in Richesson 2018)

Free text fields allow the user to type in character strings. Free text fields often require specification of a length where the length is sometimes limited by the system or by functionality in downstream data systems. Consistency of responses is challenging with completely free text fields and for this reason, they are rarely used for safety and efficacy endpoints. Free text fields are used when constraining the possible responses is not desired, for example, comment fields or collection of site explanations of protocol deviations.

Semi-structured text fields, however, are used often and have many variations. For example, response characters can be limited to alpha or numeric characters. Integer data can be collected and limited to a number of integers. Floating point, i.e., numbers without a fixed number of digits before or after the decimal, and fixed point, i.e., numbers with a fixed number of digits after the decimal, can be used where fractional parts are expected. Semi-structured fields may also constrain the format of entered data, such as parentheses around a phone number area code and a dash between the third and fourth digit, or specification of a date format. Semi-structured text fields constrain entered data and increase consistency in the collected data. Semi-structured fields should maximally constrain entered data while accommodating entry of all possible accurate response options. [VI] For example, assuming the system has the ability, the numeric data element heart rate for adult humans in beats per minute should be constrained to an integer value. While there will be disagreement about the range of values possible, the fastest reported human heart rate is 480 beats per minute.

(Chhabra 2012) A range of between zero and 250 beats per minute would not be unreasonable for a field constraint for a study in normal human adults. The aforementioned recommendation requires balancing clinical representativeness for the rare case such as 480 bpm or zero at death with the error-prevention benefit of a tighter range. Consideration should also be given to fields of mixed type. For example, many lab values are typically a numeric field, but some test results may be reported with a > , <, or + symbol, or as 'positive', 'negative', 'few' or 'trace'.

Radio buttons allow selection of only one choice from a usually short list of options. Radio buttons provide the maximum constraint possible. As such, this data collection structure should be used when the response options are known and standardized. [VI] Implementation of radio buttons should include a mechanism to de-select, i.e., un-select, a previously selected response. [VI]

Like radio buttons, *checklists* are usually used for shorter lists because all response options are displayed on the screen. Checklists differ from radio buttons in that they allow the user to select more than one option, i.e., check all that apply. Also like radio buttons, checklists should be used when the response options are known and standardized. [VI] Implementation of checklists should include a mechanism to de-select, i.e., un-select, a previously selected response. [VI]

Dropdown lists allow the selection usually of only one choice and also provide a significant amount of constraint. A drop-down list can be used in any situation appropriate for radio buttons but require one of more additional clicks; thus, for short lists, radio buttons are preferred. [VI] Dropdown lists are often used for longer lists to save space on the screen. However, implementing dropdown lists where a scroll is required should be approached with caution because items "below the scroll" may be more likely to be missed. [VI] A variation on both the free text and dropdown list is type ahead functionality where the choice options are restricted by matching the options to the characters typed by the user. Type-ahead functionality in combination with a dropdown list may allow use of very long lists including some clinical controlled terminology sets such as the International Classification of Diseases, or Current Procedural Terminology. Implementation of checklists should include a mechanism to de-select, i.e., un-select, a previously selected response. [VI]

Radio buttons, checklists, and dropdown lists collect discrete response options. The underlying data, however, may not always be discrete. For example, an eCRF may collect the following data element, "Does the participant have hypercholesterolemia?" with the response options yes and no, rather than collect the raw cholesterol value. This discretization also represents a clinical diagnosis which admittedly may take into account more than a single lab value, the discretization reduces the information content of the data. The yes/no response would be useless if a different definitional range for hypercholesterolemia were to be applied. Continuous ratio or interval data should be discretized to ordinal or nominal data only after careful consideration. [VI]

9) Required verses not required fields,

A field stated as required in a study protocol may or may not be implemented as such in an EDC system. Certain data may be critical to safety or efficacy endpoints of a study. However, there may be times when the data are legitimately not available. Functionality for implementing a field as required in an EDC system differs across systems. In some systems, marking a field as required means that the user can't save the form or move past the field or form until a value is supplied. This is often called a "hard stop" or a "hard required". Many EDC system offer an override feature where a user is prompted to provide a missing value and allowed to still move forward. This is sometimes called a "soft required" meaning that a value for the field is expected and if the needed values are not entered, the user can acknowledge or override the alert. The missing value may or may not be tracked depending on the functionality offered by the EDC system. Whereas paper-based data collection mandated a query be sent to the site after missing data were discovered, in EDC, this can be a one-step process, with the user confirming that the value is missing, and possibly a reason why. Where data is expected to be missing, it may be appropriate to include response options for the user to indicate a reason the data are not provided, e.g., "sample not collected", "assessment not done", "not applicable", "data not available/not retrievable", "asked but unknown", "asked but subject refused to answer", "actual value invalid", etc. [VI]

Some data are conditionally required. For example, a pregnancy test result is often conditionally required based on gender, and the specifics of an adverse event are required when an adverse event is indicated as having occurred. Many EDC systems include functionality to implement hard or soft constraints for conditionally required data. Implementing such functionality decreases user data entry time and frustration and should be implemented where it exists. [VI]

10) Calculated Derived fields

Most EDC systems have the ability to derive fields using basic calculations and algorithms. For example, Body Mass Index (BMI), is a calculation dependent on the subjects' Height and Weight. A field may be placed in the eCRF that automatically calculates and displays the BMI during entry for the site. Other examples include unit conversions, calculating weight-based drug dosing and scoring rating scales, and applying eligibility and other study criteria to raw data. This functionality is useful in providing decision support to sites. Similar to calculated fields, there are times when it is helpful for a site to have the ability to 'see' the data that was entered at a previous study time point, form, or visit. The user should not be able to change the re-displayed data copied into the current visit. [VI] If the original data changes, the calculated or copied value should automatically be updated. [VI] How copied or calculated fields work should be emphasized in training and a mechanism should be in place to indicate to the user how and why data has appeared in a form. [VI]

Algorithms to calculate values may be run at the time of entry or afterward and stored but not displayed. Algorithms to calculate values should be run at the time of entry and calculated

values should be displayed if they are used in decision-making at the clinical investigational site or if edit checks are based on the calculated value. [VI] If calculated derived fields are to be used, they should not be editable by the site, to ensure a consistent calculation is performed across the study population. [VI] Calculated derived values for the purpose of data analysis are usually not programmed into EDC systems, rather they are programmed as part of building analysis data sets. Calculations are the result of algorithms; like other computer programs, procedures should exist to determine the extent of testing and the processes by which testing and documentation of testing should occur. [I] (Title 21 CFR Part 11)

11) Dynamic Fields

Many EDC systems have the ability to conditionally display a field (or not), for example, based on a previously entered data value. In this case, the field on which the condition depends is referred to as a trigger field, and the conditional logic is referred to as the trigger. This is often referred to as 'skip logic", 'skip patterns", 'dynamic branching', or 'dynamic fields'. For example, if collecting whether a procedure was performed, a lead-in question might be asked, "Did the subject complete the procedure?" If the answer is "Yes", then the eCRF will display questions specific to the procedure. If the answer is "No", then the eCRF will show only a drop-down select list or text box for the site to record the reason the procedure was not performed. Displaying or activating fields only when a response is valid is a form of constraint and prevents discrepant data from being entered. Such constraints should be implemented where feasible. [VI]

Some systems perform this feature in real time while other systems apply such rules once the form is saved. In the latter case, the functionality is limited to fields on subsequent forms. If the EDC system supports complex dynamic field branching in 'real time', sometimes called multi-layered dynamics, then use of the feature to control entry is recommended to catch errors at the earliest possible point or altogether prevent them. [VI] If the EDC system does not have real-time branching functionality, clear instructions for completion as well as edit checks to catch logically discrepant data should be used. [VI] Dynamic field behavior should be emphasized in training and a mechanism should be in place to indicate to the user how, why and when a dynamic field appears. [VI]

12) Dynamic Forms

In EDC systems, fields are associated with forms (or item groups or modules of forms), and forms may (Figure 1a) or may not (Figure 1b) be associated with a study visit. Similar to dynamic fields, many EDC systems support dynamic forms, i.e., forms that appear only when a subject meets a certain criterion such as entry of a particular data value. A common example of a dynamic form is a form for prostate cancer screening that only needs to be completed for male participants. Because dynamic forms do not always appear to be available in the EDC system, e.g., the prostate screening form will not appear in the system if a participant is female, they have the potential to confuse users. Dynamic form behavior should be emphasized in

training and a mechanism should be in place to indicate to the user how, why, and when a dynamic form appears. [VI]
Use of dynamic forms requires considering how the data on dynamic forms activated in error are handled when the form is subsequently inactivated. For example, consider the case where the sex of a patient is incorrectly entered as female, generating dynamic gender-specific forms; afterwards, the gender is subsequently corrected to male. Different EDC systems handle this scenario differently. Because the origin and all changes to data should be recorded and immutable, (Title 21 CFR Part 11, ICH E6(R2)) the removal of dynamic forms generated in error and any data entered on them should be permanently tracked by the system. [I]

Dynamic forms such as repeat, event-driven, or unscheduled assessments can be automatically triggered as just described or they can be manually triggered. Some EDC systems support repeat form functionality where a form can be set-up to allow site users to manually trigger a new instance of the form. For example, some studies may allow for or require repeat assessments for abnormal vital sign, laboratory, or ECG results. These could be implemented as a repeat form if the EDC system supports this functionality. Using built-in system functionality for repeat forms often also automatically maintains the association of the repeated assessment results records with the visit in which the original result was measured. However, there is substantial variability in if and how EDC systems support repeatable forms; for example, some systems only allow for repeatable forms in association with unscheduled visits. Manually triggered dynamic forms such as repeatable forms are usually used when the necessity of the additional form instance is (1) dependent on the participant's course or resulting data and (2) only occurs for a subset of the participants.

These manually or "site user-triggered" forms require consideration at set-up for how to maintain the association of the data with the appropriate time point or visit. In other words, where the EDC system doesn't or cannot support automatic association of dynamic form data with the needed time point or triggering event, special provisions for referential integrity must be made such as requiring entry of the number or date of the event or visit with which the data on the new form should be associated. [VI] Further, the appearance of the new instance on the system should be clearly distinct from a visit. [VI]

Many EDC systems have built in functionality for the user to enter the actual date of data collection and associate the date with the data collected. This data is both 'meta-data' (data about the data), as well as part of the clinical data in the EDC system. It is often assumed that an assessment or collection date can be derived from the visit date if needed. Thus, physical exam or vital signs forms often lack assessment or collection date, presumably to decrease data entry burden on the site. However, some studies permit assessments to occur within a window of the scheduled or actual visit. These are often implemented as dynamic forms. Omitting a visit date for dynamic forms or forms within the visit results in information loss by losing the association of the assessment data with the actual date on which it was collected. We recommend explicit association of the data to the collection date for better traceability as articulated in ICH E6R2. [VI]

Determining how to best implement dynamic forms depends on capabilities of the EDC system and complexity of the study. When used multiple times within a study or program, how dynamic forms are triggered and completed and how the content and behavior appears to site users should be consistent to decrease the potential for user confusion. [VI] For example, consistently using dynamic forms in similar situations such as all repeatable forms or event-driven forms, and using the same process for triggering, completing and correcting them across all form instances will go a long way toward usability.

13) Dynamic Visits

Similar to forms, some visits are expected for a study, i.e., scheduled per protocol, whereas others are conditional, that is some visits are only needed for a subset of the participants and driven by events that occur during the study. Expected visits are usually described as such in the study visit schedule within the protocol and implemented as such in the eCRF design and EDC system. Like event-driven forms, event-driven visits may or may not occur and are created as needed. Event-driven or otherwise unscheduled visits are not expected to occur for every participant. Many EDC systems support event-driven visits by facilitating their display (or not) based on data entered for a patient or by manual triggering. In an oncology trial for example, when a patient meets a certain criterion he or she may move to a different treatment group with a different set of visits. In some EDC systems, the ability to auto-skip or hide visits is an option. For example, a field is answered in a visit form stating the subject discontinued from the study and the remaining visits are skipped or hidden from the visit schedule and are no longer expected for that subject.

Similar to their form counterparts, referential integrity for repeatable (sometimes called multi-occurring) and event-driven visits often requires special consideration. For example, expected visits have a minimum count of one and a maximum count of one while event-driven visits may or may not occur and they may occur multiple times for a patient within a study; i.e., a minimum count of 0 and often a maximum count greater than one. Providing for referential integrity means that the data collected on dynamic visits can be associated with the correct triggering event as well as the correct time point. To accomplish this, usually these visits must be declared and set-up as such. Dynamics may impact data entry efficiency and system speed so clinical data managers should weigh the benefit versus the possibility of overloading sites with confusing or complicated dynamic functionality when considering dynamic visits. [VI] Like dynamic forms, dynamic visits should be taken into consideration in data status reporting.

Figure 1: Varieties of Alignment of Fields, Forms, and Visits Commonly Supported by EDC Systems

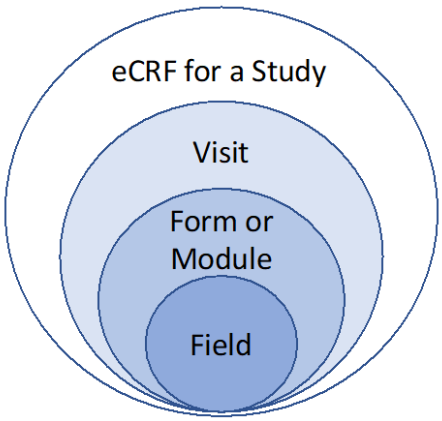

Figure 1a: EDC systems often reinforce an alignment between fields, forms (or modules), and visits. Sometimes it is a strict hierarchy.

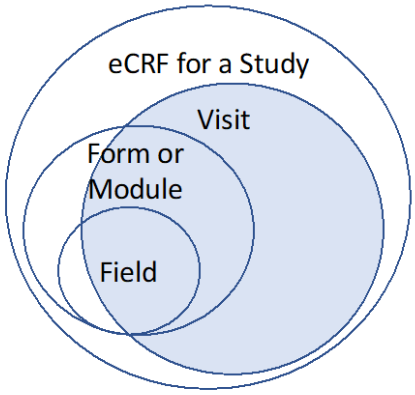

Figure 1b: Some systems may not require fields, forms, and visits to be associated in a strict hierarchy and may support collection of data outside the study visit structure; i.e., forms that are not associated with visits such as log-based forms that span the study like concomitant medication logs or study withdraw forms.

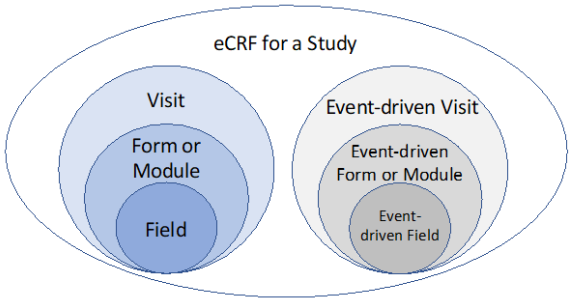

Figure 1c: Some EDC systems may support event-driven forms and may also have mechanisms or requirements for associating them with visits, other forms, or fields. Examples include repeat assessments or unscheduled visits.

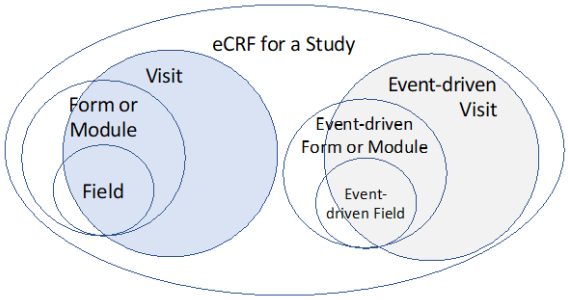

Figure 1d: Some EDC systems may support both event-driven forms and visits as well as forms that occur outside of a visit context. Examples include event-driven forms and log-based forms that span the study such as a log of protocol violations or a log of clinical events.

14) Form Instructions

From a cognitive engineering standpoint, forms serve as an extension of an individual's thinking. (Zhang 1997) Good form design minimizes cognitive load on the user, i.e., the number or complexity of mental operations that a form completer needs to perform. Thus, the best form completion instructions are those that are not needed because the structure of the form makes correct completion obvious and prevents incorrect completion. In the remote collection of study data as in the case with EDC, such extensive constraint is often not possible. For example, where a site has only year or month and year for a medical history item, studies

would rather have lower resolution than no data. Sometimes the flexibility is needed to account for expected variability while in other situations the lack of system functionality requires it. Form completion instructions fill the gap between the ideal of complete constraint and the reality of needing some flexibility on the user interface and are usually required. Most EDC systems offer options for form completion instructions that go far beyond those available to studies using paper forms. For example, most EDC systems provide additional opportunities to co-locate instructions with the fields to which they pertain via mouse-over or other field-specific and user-activated help. Field-specific instructions should be as close as possible to the field to which they pertain and available to the user with minimal barriers to access. (Wickens and Hollands 2016) While this leans away from provision of form completion instructions as a separate document, instructions are the minimum required to support consistency where options exist.[VI] A separate instruction document is better than optionality without clarification. [VI] There is a trade-off between (1) the need to cover options with instructions and (2) the amount of time needed to specify and add them into an EDC system. For more information on form completion guidelines, refer to the CRF Design and CRF Completion Guideline GCDMP chapters,

15) Data Integration Set-up

Data independence is the ability to change data values and logical or physical structure of the data without changing the software application that uses the data. (Earley 2011) Maintaining independence of data from software has been a commonly followed best practice since Edgar F. Codd, a pioneer of the relational model for databases, defined rules for database management system to be considered relational. (Codd 1970) Most EDC software today utilizes an independent underlying database management system. Practitioners referring to the "clinical database" usually mean the data as stored in the database management system utilized by the EDC software. Thus, the clinical database usually always contains data entered through the EDC system.

Most EDC systems offer functionality to import, integrate, process, and display externally managed data such as data from central clinical labs, core labs, ePRO systems, and central reading centers. However, functionality for the acquisition and management of external data vary widely from real-time data feeds with all data processing functionality applicable to the imported data to systems that require manual imports of external data and those that do not support standard data processing functions such as discrepancy identification, discrepancy resolution, and change tracking for imported data. The decision to import and integrate external data depends on the study needs, the functionality available in the EDC system and the resources required to apply that functionality. Clinical data managers should understand how data collected or maintained outside an EDC system will be used, who will use it, and for what purpose. The answers to these questions help determine the extent and timing of data integration.

Some vendors involved in the collection and management of data from central clinical labs, core labs, ePRO systems, and central reading centers offer real-time access to an information

system where the external data can be viewed by sites. However, these often require a separate login. Integration of external data into an EDC system may be required if the data have direct impact on clinical decisions or study management. Examples include where EDC-based randomization uses scores on patient-completed assessments or when doses are adjusted based on results from an external lab. When data are needed and expected to be used by site users of the EDC system, they should be integrated into the EDC system. [VI]

Because integration of external data usually includes reconciliation and cleaning of the data, integration of external data into the EDC system also facilitates interim analysis and database lock by these checks having been conducted in an ongoing manner throughout the study. Where it obviates the need for manual entry of data, integration of external data likely saves time and increases data quality. Maintaining a study blind is an additional consideration in integration of external data. Data with the potential to unblind a blinded study, for example, a lab result that might give away the treatment assignment, may require separate treatment or access controls.

Setting up an EDC system to receive imported data usually requires creating data fields within in EDC system to receive and store the data to be integrated as well as the algorithms through which the incoming data are parsed, transformed if necessary, and written to the destination fields. Because data integration requires algorithmic or manual manipulation of data, the planned data integration should be fully specified, tested, and traceable. [I] (Title 21 CFR Part 11) Detailed considerations and practices for designing, specifying, managing, and assuring the quality and compliance of externally managed data are provided in the GCDMP chapter titled integration of external data.

16) Data Validation Checks (Edit Checks)

Data validation checks are algorithms that are used to screen data for invalid, questionable, or anomalous values. They are sometimes referred to as edit checks, query rules, error checks. Data validation checks that run and identify issues as data are entered, as is the case with EDC, are referred to as on-screen checks.

The use of edit checks in an EDC system offers clinical data managers the opportunity to screen data for invalid, questionable, or anomalous values and alert the enterer earlier than with paper-based data collection processes. Use of on-screen edit checks is associated with data quality similar to that of double entered data. (Zozus et al in Richesson 2018) Further, some data errors can't be corrected after time has passed; e.g., a clinical assessment at 30 days post dose can't be recreated after the passage of time. While assessment the next day may be acceptable, the validity of the observations as reflective of the 30-day timepoint decreases with the passage of time. (Zozus 2017) On-screen checks gives the enterers the ability to address the flagged values sooner if not immediately, ideally during the assessment or when the source of the information is at hand. Preventing errors or catching data problems earlier reduces costs. It is widely accepted that there are significant increases in total cost the further downstream errors are caught. This concept has become known as the 1-10-100 rule and is described in

three stages, as error prevention is ten-fold less than correction where correction is yet again ten-fold less expensive than remediation of failures due to uncorrected errors. (Labovitz and Chang, 1993) Thus, in a risk-based approach, costs associated with prevention can be weighed against cost of correction and damages from failures due to uncorrected errors. On-screen checks should be used with EDC to the extent that benefit outweighs cost associated with, for example, human safety, re-work, and regulatory delays. [VI]

Operationally, on-screen checks in EDC systems increase the immediacy with which Data Managers, study Monitors, or in-house Study Coordinators or Site Managers can become aware of and review unresolved discrepancies and interact with investigational sites to resolve them. Such data-driven contact by phone with site staff promotes an active approach to decreasing elapsed time to complete and clean data. Getting data in and clean faster has always been a major part of the value proposition of EDC. At the same time, because EDC broadened the number and variability in users from internal personnel to users at all of the clinical sites involved in a study, the requirements for system training and usability are significantly increased.

a) Types of Edit Checks

Most EDC systems use a rules-based approach to identification of discrepant data and have functionality for authoring, storing, managing, executing the rules and tracking the lifecycle of identified discrepancies. Edit checks in EDC can be classified into two broad categories, "hard" edits and "soft" edits. Soft edits identify discrepant data and usually prompt the site for data correction but allow the data to be confirmed as is and saved so that entry can continue. Whereas hard edit checks also identify discrepant data but prevent the identified data from being saved. In some systems, the form itself can't be saved with open hard edits. In other systems, hard edits do not produce an alert that a user can "confirm as is" or override. Thus, hard edits are sometimes called non-actionable because the user can't acknowledge the check and proceed; the only permissible action is to enter data that conform to the requirements. Data type checks (sometimes called browser checks because they almost always run real-time in the browser) are commonly implemented as hard edits. For example, if a user attempts to enter an alphabetical character in a numerical field, the check will not accept the data and if the field is required, the form will not save until conformant data are entered. Another type of a hard edit is a property check. Property checks prevent entering data that do not match form and/or item property settings of the field documented during system set-up. For example, when a field requires a number with 2 decimals, a value of "3" cannot be entered. Instead, a number with two digits to the right of the decimal must be entered to satisfy the property requirement. Without satisfying the property requirement, if the field is required, the form will not save until conformant data are entered. For this reason, hard edits are usually used for non-feasible scenarios such as physically impossible values while soft edits are used to identify data values that are unexpected or unlikely but which could occur. These considerations are more important in the context of EDC because the data enterer is at a clinical investigational site. Because failure of a hard edit prevents forward progress with the task of data entry, users are

incentivized to enter a data value that will "pass the check". Thus, we do not recommend use of hard edits in EDC. [VI] Many systems have evolved and now allow all checks to be implemented as soft edits and allow entry of otherwise invalid data along with a reason for the non-conformant data.

b) Lifecycle Documentation and Management of Edit Checks

Because data entry is done by investigational sites with EDC, the user interface and usability becomes more important. If a discrepant data value is identified by an edit check, a real-time indicator such as a color change, an alert, haptic feedback, or a change in iconography on or near the discrepant data is most helpful to the user. [VI] Similarly, an explanation of the discrepancy should be readily available. [VI] In addition to real-time cues to the user, a lifecycle record for all detected discrepancies best meets the traceability requirements as stated in ICH E6(R2). [I] (ICH/FDA 2018) Such a record provides a mechanism through which changes to data can be reconstructed from the original entry, a prompt (or not) regarding a discrepancy, and changes to data. In the absence of such a record, it is not possible to distinguish prompted versus unprompted changes to data following the initial entry. Further, a record of open discrepancies facilitates reporting and active management of data collection and cleaning. EDC system functionality for lifecycle documentation and management of discrepancies varies.

Usability and lifecycle documentation and management of data discrepancies can be disrupted where EDC functionality does not support complex multivariate checks. For example, complex rules such as those that cross multiple forms or visits or those with logic requiring extensive programming may not be able to be implemented real-time on the user interface or, in some cases, implemented at all within the EDC system. EDC systems vary and a check may be considered complex by one system and easy for another. Where such complex checks are considered required for the study and cannot be implemented within the EDC system, they must be developed and implemented outside the EDC system. Consequences of developing and implementing edit checks externally include inability to execute them in a real-time manner on the user interface, cost and time to maintain the external systems, additional resources required to track and report externally identified discrepancies, and challenges providing comprehensive status and work-facilitating reporting during the study. For example, how will the check results be communicated to the users via the EDC user interface if they are not themselves implemented within the EDC system? Without an interface, manual re-entry of queries into the EDC system is required so that site-based users can use the EDC system to resolve queries. The resources needed to manage this activity should be considered. These realities erode the benefit of EDC.

In addition to edit check complexity, other factors such as data availability and system performance may prompt consideration of implementation of edit checks external to the EDC system. For example, if an edit check uses coded terms but coding occurs external to the EDC system, unless data are coded by the site users within the EDC system, the edit check cannot run in real-time. Further, for edit checks to be run and tracked within the EDC

system the coded data must be imported into the EDC system or be available to the EDC system through an interface. A second factor that commonly prompts implementation of edit checks outside the EDC system is system performance in the presence of numerous checks or complex checks programmed in EDC. For example, suppose a there is a need to check that the last date of subject contact is the last chronological date in the database. In this case, the edit check should pull all dates from each module in the database and compare those dates against the newly entered date of last contact, or prepopulate it directly. This type of edit check might access the underlying database thousands of times and noticeably degrade system response time. Optimization of system performance may require balancing running complex checks real-time, system response time and infrastructure cost and most often require collaboration with Information Technology professionals because of the interplay between hardware and software. [VI] The trade-offs between real-time identification and resolution of discrepancies and data availability and system performance may also erode the benefit of EDC.

Usability and lifecycle documentation and management of data discrepancies can be disrupted where EDC functionality does not support graceful lifecycle management of the edit check rules themselves. To prevent frustrating and time-consuming rework from data discrepancies identified in data that previously appeared to sites as complete or clean, edit checks should be available when the system is moved to production. [VI] While this recommendation is straightforward when starting a study, mid-study changes bring challenges and trade-offs. EDC systems have different limitations when adding or altering edit checks after data have already been entered. For example, while some systems have the ability to re-trigger edits on existing data edit checks that are added mid-study, other systems may only apply new edit checks to new or modified data. Therefore, the clinical data manager should consider how existing data will be checked and may need to provide for checking previously entered data by means such as programming a listing to identify issues with existing data. In this example, sites should also be informed that they may be required to resolve issues identified in earlier visits.

Similarly, usability and lifecycle documentation and management of data discrepancies can be disrupted where EDC functionality does not support control over when checks are run. Edit checks may generate queries due to the order of data entry. For example, consider an edit check that compares the date for visit 3 to the date from visit 2 to assure that visit 3 occurs after visit 2 in time. If data for Visit 3 is entered prior to visit 2, i.e., out of the expected sequence, the check triggered from entry of the visit 3 date may not run on entry without the comparator record for visit 2 and may not run when visit 2 is subsequently entered because the check is triggered from visit 3. These scenarios depend on the functionality supported by the EDC system. Some systems may allow manual re-execution of all checks to ameliorate this problem whereas others may not.

EDC system functionality for developing and managing edit checks vary. For example, some systems handle Univariate checks in the data element or screen definition process. Univariate checks, those that apply to a single data value, specify valid values of a data

element and thus are sometimes viewed as properties of the data element. Examples of Univariate checks include data-type checks, missing checks, value options for enumerated data elements, and maximum, minimum, or range checks for numerical data elements. Univariate checks are often specified and managed as part of the data element definition process within EDC systems. The resulting data definition is then leveraged to offer automated checking of entered data against the valid value constraints in the data definition. The type and extent of definitional information entered and the extent to which EDC systems leverage it for checking data vary across systems. Importantly, edit checks that rely on properties or similar metadata do not require computer programming and thus data definition-based checks are not subject to the requirements of software validation; i.e., they do not need to be tested for each study set-up within an EDC system once the relevant functionality is validated. Not all univariate checks can be supported by definitional metadata, for example conditional univariate checks such as those that apply in some situations but not in others. Univariate checks that have to be programmed as edit checks must be tested just like any other computer program. [I] (21 CFR Part 11)

Multivariate check functionality varies even more. Recall that multivariate checks are those that compare multiple data values, for example, comparing subject weights over time to use unlikely weight changes to identify potentially errant data. These checks usually require writing rules (logic-based algorithms). EDC systems vary in the extent of support for authoring and managing such rules with some systems merely storing executable SQL code written for the system's data model. Such rules are custom computer programs and should be tested as such. [I] (21 CFR Part 11) Guidance on rules-based approaches to data cleaning and methods for developing, testing and managing rules can be found in the GCDMP Chapter titled Edit Check Design.

To define and review edit checks prior to production release of an EDC study, clinical data managers coordinate activities with clinical, IT, quality control, quality assurance, and/or other groups. Because of the aforementioned trade-offs and impact on study operations at sites, the approach to data cleaning should be discussed during development of the EDC study specification, and in consultation with all stakeholders involved in data validation, especially sites and team members that work directly with sites [VI].

17) Medical Coding Set-up

During the eCRF development process, all data fields to be coded and the controlled terminologies with which they will be coded should be identified. [VI] Decisions about coding of medications, adverse events, procedures, and other study data should be documented in organizational or study-specific procedures or guidelines. [I] (ICH/FDA 2018) This is particularly relevant in medical coding because coding tasks are often shared between algorithms and humans in processes that leverage autoencoding technology followed by use of a human coder to handle those terms not codable by the algorithm.

EDC system functionality for medical coding varies considerably with three main process variants. The first (Figure 2a) involves use of type ahead functionality on verbatim term fields to facilitate use of the controlled terminology by the data enterer at the clinical investigational site. Because controlled terminologies can be very large, for example MedDRA or SNOMED contain between 70,000-100,000 terms, this model requires optimized architecture and infrastructure to assure adequate system response time. Further, the structure of controlled terminologies can vary widely from a list to a single hierarchy taxonomy to a poly-hierarchical taxonomy to a poly-hierarchical system of multiple relationships, i.e., an ontology. Supporting this model means that the EDC system also must have functionality to parse these controlled terminologies for terms, to store them and to update them with new releases of the terminology. For these reasons, while the model in Figure 2a, i.e., having sites review automatically applied codes may be attractive, it is not often the case. A second model involves the type ahead and dictionary management functionality described in 2a, but also allows for non-matching verbatim terms to be saved and coded later by a central medical coder (Figure 2b). This model requires that the infrastructure for medical coding is available at the time of data entry in addition to functionality to support central coding and issuing coding-related queries to sites through the EDC system. While this model decreases coding-related queries via the type ahead matching and real-time review of the matched term by the sites, it requires the infrastructure of both front-end and back-end coding. The third model is the traditional post-processing model and involves the clinical investigational site entering verbatim terms that are later coded centrally (Figure 2c). This model relieves the pressure of system response time associated with type ahead coding. However, a mechanism of communicating coding-related queries to the sites, preferably through the EDC system interface, is required.

Optimizing the coding quality and system usability by the clinical investigational sites requires a good understanding of the capability of the EDC system to support one or more of the coding models.

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. [VI] Unless coded terms are included in queries, to avoid confusion, it is recommended not to display coded terms back to the site user. [VI] Clinical data management should work with team members trained in controlled clinical terminology to determine how data coding should be handled. [VI] Ensure the clinical team understands who will be coding terms that don't match or otherwise autoencode and how clinical review of coding, where deemed necessary, will occur. [VI] Documentation of the coding process should include training or guidelines for assigning codes, the frequency of coding and clinical review, procedures, timing for any data imports or exports required, and management of dictionaries used in the coding process.

Most terms that code will have additional code/vales associated, a mechanism to re-associate the codes with this additional information should be built into the DB. For example a MedDRA term will always have an associated System Organ Class (SOC).

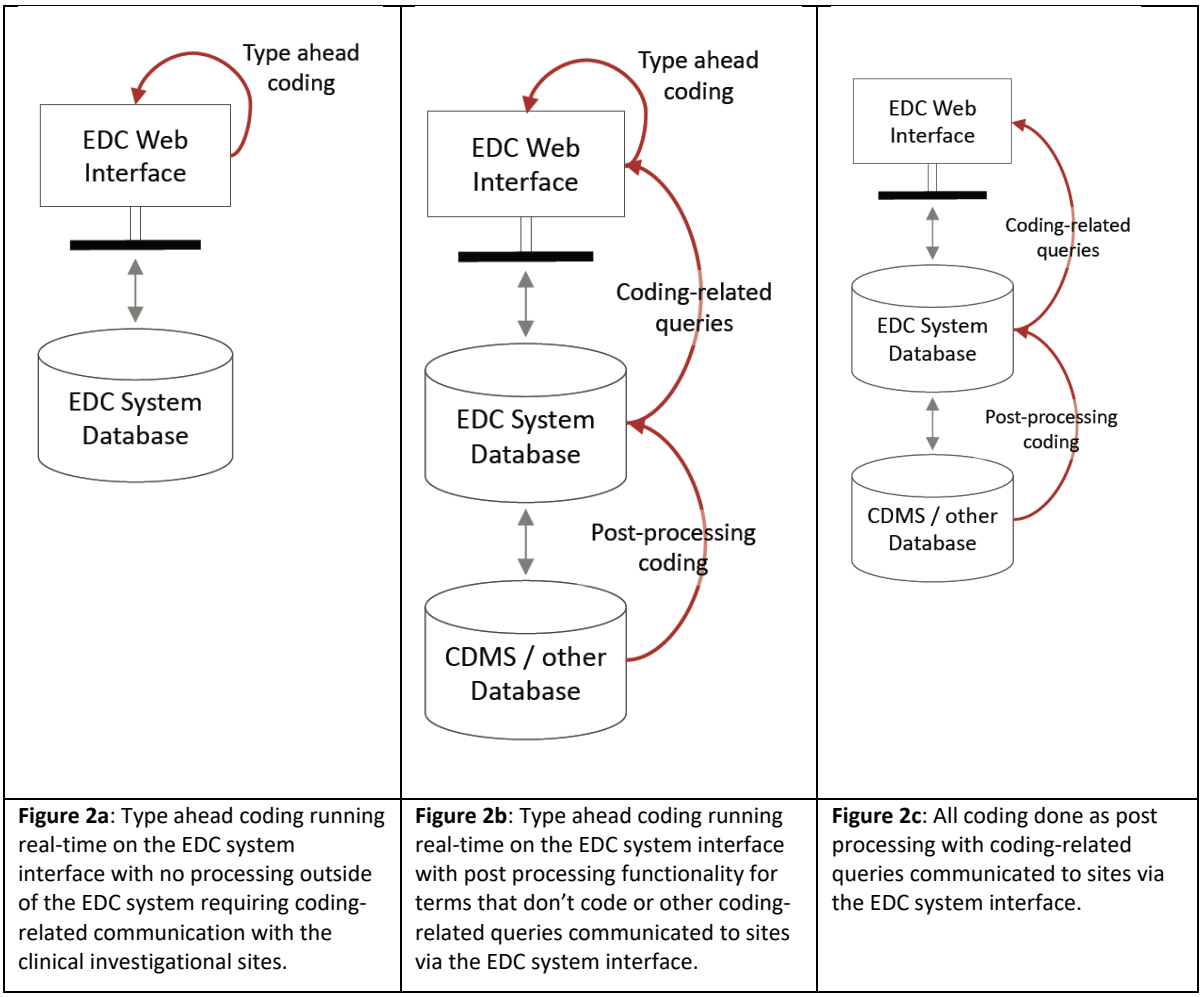

Figure 2a: Type ahead coding running real-time on the EDC system interface with no processing outside of the EDC system requiring coding-related communication with the clinical investigational sites.

Figure 2b: Type ahead coding running real-time on the EDC system interface with post processing functionality for terms that don't code or other coding-related queries communicated to sites via the EDC system interface.

Figure 2c: All coding done as post processing with coding-related queries communicated to sites via the EDC system interface.

18) Developing and Testing a Study Within an EDC System

EDC functionality with respect to building a study varies widely. Some systems require computer programs to be written to create data entry screens and the corresponding logical structures in which data are stored. However, most EDC systems have tools that decrease or altogether eliminate custom programming to set up entry screens and data storage. For example, some EDC systems accept a spreadsheet of data elements by screen and their properties such as the data type, whether a response is required, the prompt to be displayed on the screen, the data collection structure to be used, structure-specific specification of valid values, preceding data element on the screen, and grouping to which the data element belongs. The EDC system then builds the screen according to the spreadsheet. Other systems offer less automation and sometimes more flexibility in screen set-up through using graphical user interfaces where different data collection structures are added to a screen and properties are added to the data collection structure. Similarly but often to a lesser extent, most EDC systems have tools to facilitate importing and exporting data as well as for the development of edit

checks and other rule-based system features such as dynamic visits, forms and fields, screen tab order and skip patterns.

To the extent custom computer programming is required, professionals trained in relevant programming language, style and tools are required. [I] (Title 21 CFR Part 11) Further, to the extent that custom programming is required, so is a documented process for specification, development, and validation of programmed components. [I] (Title 21 CFR Part 11)

User testing with comprehensive test cases is strongly recommended for EDC studies. [VI] Because the users are external, problems can be more impactful and harder to remediate than problems in a system used by internal data management staff. Errors in rule specification can cause equally serious problems such as rules never firing or firing in false positive manner. Fixing problems with rules often requires site users to go back and address newly fired discrepancies on data previously thought complete and clean. For this reason, each rule should be tested with at least one boundary with a case that causes the rule to fire and a test case that should not cause the rule to fire. [VI] Where rules are tested in a manner that does not address each logic path in the rule, rules should be monitored once in production to identify rules that fire too frequently and rules that have not yet fired. [VI] The more data accrued, the better the ability of such monitoring to identify rules likely to be malfunctioning. Active monitoring finds problems sooner and prevents sites from receiving queries from errant rules fixed late in the study. Studies should not collect production data until a User Acceptance Testing (UAT) has been performed and documented. [VI] The extent of UAT, i.e., the number and type of test cases for screens and rules can and should be risk-based. [VI]

While tools and functionality obviating custom programming can save time and resources, they do not eliminate the need for testing. A system with absolutely no custom programming in study set-up should be tested. [VI] This is because errors can occur during set up and unintended consequences can result from errors in set-up. For example a spreadsheet listing fields to be displayed on a screen can contain an error in the data type, prompt, data collection structure, or valid values. Such errors result in systematic data quality problems because they most often impact every value entered in the affected field. Further, a system that functions perfectly according to specifications can cause unintended problems once in use by humans and at multiple institutions. For example, the set-up specifications for a study on which data were to be entered in-house contained different response order for questionnaire data and for clinical observations. The inconsistent display order of yes no radio buttons in a study resulted in an error rate of over 200 errors per 10,000 fields. (DataBasics, Nahm) The problem was discovered when the study chair and statistician reviewed the draft tables, found a particular result clinically unlikely, and investigated. (DataBasics, Nahm) An astute tester may have detected the problem before the system was released. Other unintended problems include a screen so long it requires scrolling causing sites to miss fields at the bottom, other field lay-out that causes fields to be consistently missed, and misleading prompts that cause inconsistent data entry. Testing in-house may catch some problems. Testing at investigational sites will likely catch more problems. Thus, some testing of “zero-programming” or configuration-only set-up is recommended. [VI]

Regardless of the type and amount of testing done, observing a system's operation once in production is recommended. [VI] System observation can take many forms including review of system error logs, distributional and conditional comparison of entered data, queries, query response, and operational metadata across visits, forms, data elements, sites, and users. Routine and ongoing system observation may serve as a trigger for risk-based activities including site calls, monitoring, investigation, and auditing. These activities also help meet the intent of ICH E6 section 5.1.3, "Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly". (ICH E6R2, 2018) Probably most importantly, frequent ongoing observation is pro-active and catches problems earlier than relying on downstream processes to identify things that look odd, for example, during analysis programing or table and listing review. "Ironically, there is a major difference between a process that is presumed through inaction to be error-free and one that monitors mistakes. The so-called error-free process will often fail to note mistakes when they occur." (Arndt 1994) For these reasons, ongoing systematic observation of system performance is recommended. [VI] Such monitoring may itself be risk-based in terms of the frequency and extent of the observations and the type of items monitored. [VI]

19) Study Start

End User Preparation (Site)

Good clinical practices advise site assessments. In addition to reinforcing the Title 21 CFR Part 11 requirement that individuals involved in conducting a trial should be qualified by education, training, and experience to perform their respective task(s), the introduction to quality management section 5.0 states that the, "sponsor should implement a system to manage quality throughout all stages of the trial process" and is followed by a description of risk identification and control. (ICH E6 R2 2018, Title 21 CFR Part 11) In that same section 5.0.2, risk identification, states that the, "sponsor should identify risks to critical trial processes and data", that risks should be considered at both the system and trial levels. (ICH E6 R2 2018) To meet the intent of regulation and guidance, a site assessment should confirm a site's ability to access and use the EDC system prior to initiation of the study at the site. [VI] Such an assessment may include personnel qualification prior training and experience, institutional infrastructure, and system training and demonstration of competence in preparation for a study.

As part of operating a validated system, the sponsor or designee is responsible for ensuring that sites are qualified to use hardware or software required by the EDC system. (Title 21 CFR Part 11) In many parts of the world, access to the internet and associated infrastructure are virtually ubiquitous; however, there may still be sites that have connectivity, hardware, or software challenges. For example, a site's internet browser or browser version, may not be compatible with the EDC system, or the local area may have less than ideal electrical power quality. Internet-based test sites will suffice in many situations; i.e., "if you can access this site, you will be able to use the EDC system". In rural areas or parts of the world lacking consistent electrical power or internet access, more consideration should be given

to a site's ability to use EDC. Site evaluation and qualification with respect to EDC systems by the sponsor or designee should occur during start-up activities prior to subject screening. [VI]

20) EDC Account Management

Setting System Rights Determined by Roles and Privacy

Title 21 CFR Part 11 requires, "Limiting system access to authorized individuals" including use of, "authority checks to ensure that only authorized individuals can use the system, electronically sign a record, access the operation or computer system input or output device, alter a record, or perform the operation at hand". (Title 21 CFR Part 11) System access and privileges within the system need to be considered for all roles using the EDC system. Management of system access and privileges begins with enumeration of the roles and the responsibilities and tasks to be associated for each role within the EDC system. [VI] Available roles, tasks, and allowed associations vary across EDC systems. Factors to be considered when defining user roles include the following:

- Data Entry Rights—it is important to understand which users will need access to each form or groups of forms within the study. In most clinical trials, site users will be the most common user who will need data entry permissions; however, in some studies, call center, central reading center, and core lab users or ePRO completers may need more limited data entry rights. In some scenarios the sponsor or their designee's staff may need entry or edit rights. For example, in EDC systems with limited coding functionality, dictionary coding requires that sponsor/CRO staff be able to enter or modify verbatim term fields on a form. To ensure that integrity and reliability of data are maintained, sponsors should carefully consider which fields will be modifiable by the sponsor team.
- Data Management Review (DM Review) or other custom rights—some EDC systems are configured to have other workflows such as DM Review, Medical Monitor (MM) review, etc. If these workflows are available as part of the EDC system and turned on for a particular study, it is imperative to have certain users set up with the appropriate permissions and process documentation outlining the workflow and necessary steps.
- Source Data Verification (SDV) rights—CRA or other clinical operations staff may have SDV rights to indicate source-verified fields and to enter queries to the site where discrepancies are noted.
- Read-only access—Some roles may require read access to some fields; for example, a research pharmacy filling an order or a central reading center viewing data associated with an event under review.
- Creating manual queries—CDM, CRAs, Drug Safety, clinical/medical coders, etc. may all have the ability to create different types (CRA, DM, etc.) of manual queries.
- Answering or resolving queries (manual or system)—sites will always have the ability to answer manual or system queries but some EDC systems may allow other configured users (DM, Drug Safety, etc.) to respond to queries as a part of the data cleaning process.

- Closing queries (manual or system)— CRAs may only be able to close or resolve queries created by a CRA user group, while CDMs can close system or manual DM queries after reviewing site responses. In some EDC configurations, CRAs and DMs could share responsibility for closing one another's queries.
- Report creation, generation, or view-only access at both the site and by the sponsor or designee should be considered. Some possible scenarios include limiting access so that each site can only generate reports for their subjects or CRAs can generate reports for subjects at their sites or the entire study depending on the user permissions, limiting report generation across countries or regions, or limiting report creation to CDM staff who have received more advanced training.
- Data extraction should be similarly limited to prevent unintended disclosure of data.

Documentation and tracking over time of access and privileges in the system supports auditability of procedures.

21) User IDs and Passwords

User credentials such as user identifiers and passwords are essential to the control required for non-repudiation by Title 21 CFR Part 11. As such, Part 11 section 11.10 requires the, "establishment of, and adherence to, written policies that hold individuals accountable and responsible for actions initiated under their electronic signatures, in order to deter record and signature falsification". This requirement in Part 11 section 11.100 requires written certification to the FDA that electronic signatures, "are intended to be the legally binding equivalent of traditional handwritten signatures". [I] (Title 21 CRF Part 11) In addition, electronic signatures must be unique to one individual and should not be reassigned and the identity individuals using electronic signatures must be verified. [I] (Title 21 CRF Part 11)

Processes for dissemination of user credentials such as user identifiers and passwords should be established. [VI] These processes should include tracking that users have been properly trained prior to receiving access to the system. [I] (Title 21 CRF Part 11) To support non-repudiation by keeping user credentials secure, the EDC system should force users to change their password at first log-in. [VI] and training or system documentation should educate users as to the rules and regulations regarding keeping user ID and password information confidential, as well as requirements for changing their passwords. [VI] Lastly, the training materials should instruct users on what to do should they lose or forget their ID and/or password. [VI] Thus, site users should have an individual and not a shared email account to receive user IDs and passwords for EDC applications. [VI] Some institutional sites may use a shared email account for operational purposes. This can be problematic if the EDC system uses email address to uniquely identify user accounts.

22) Account Management

The account management process may be defined with cross-functional input and should be maintained by a function with knowledge of and close communication with the sites. [VI] This

supports site user training as well as validation of an individual's identity and detecting personnel changes requiring changes in system access and privileges. Consideration could be given to linking the Clinical Trial Management System (CTMS) to the account creation and activation system, thereby eliminating the need to transfer user information between systems. [VI] A secure process for managing access and privileges will minimize the number of manual steps that are included and employ separation of duties. [VI] An example of a typical account activation process is enumerated below.

1. A user is trained and authorized to be granted access to the system for a specific role.
2. The sponsor or designee confirms that EDC training has been completed by the user.
3. An account is created, and access provided.
4. Account use is monitored for aberrant behavior and site staffing is monitored for changes necessitating discontinuation of access and onboarding new site personnel.
5. Accounts are disabled as the access need diminishes when individual patients, visits, or the database are locked.

23) Training Prior to System Access

Title 21 CFR Part 11 requires a determination that, "persons who develop, maintain, or use electronic record/electronic signature systems have the education, training, and experience to perform their assigned tasks". [I] (Title 21 CFR Part 11) Following a risk-based approach, training for site users with previously established system education, training, or experience may be less extensive than for site users lacking relevant education, previous training, or experience with the system. Similarly, training for an open-label extension or similar trial with similar data collection in the same system may be significantly reduced. On the other hand, for inexperienced sites or new system functionality or processes, study-specific training in the EDC system may be more extensive and include an assessment of competence. [VI]

Documentation of training completion should be maintained in the Trial Master File (TMF). [I] (ICH E6 R2 2018) Training documentation may also be given to trainees and used to support qualification on future studies. [VI]

User training on both the system and study application is important. There are varying views on the extent to which these two components should be included in training. At a minimum, all users should have competency in basic system functionality available through their permissions. For a site user, these usually include how to login, how to navigate to patients and visits in the system, how to enter and update data, and how to respond to system generated data discrepancy notices. [VI] Often, studies add dynamic behavior; in these cases, study specific training covering how the study eCRF responds to different user actions and input may be required. Because of the increased interaction between data and form behavior in EDC it may be effective to combine training on the study eCRF with training on data collection such as training on guidelines for where in the medical record to find needed data and what value should be chosen in the case of multiple conformant values. [VI]

User training can be provided through different methods, including

- Self-study of reading or e-learning materials followed by demonstration of competency using sample forms in a training environment
- Demonstrating competency in training environments that provide training exercises with examples that are generic or customized to the study-specific workflow
- Web-based instruction or decentralized/remote demonstration followed by demonstration of competency using sample forms in a training environment
- Face-to-face training for users in a central training facility, such as at investigators' meetings or other centralized training meetings.

Consideration should be given to issues posed by language barriers to 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.

The training requirements articulated in Part 11 also apply to individuals who build, test, and maintain the study eCRF and those who manage accounts, privileges and study data within the EDC system. [I] (Title 21 CFR Part 11) Individuals with these responsibilities should have documented training corresponding to their roles and responsibilities. [I] (Title 21 CFR Part 11)

24) Study Considerations and Start-up Timelines

Because the study database should become active prior to enrollment of the first subject, study start-up activities are critical for EDC studies. While in paper-based studies, data entry screens could be finalized prior to edit checks being ready, this practice is not recommended in EDC studies. [VI] As described earlier in this chapter, it is best to run checks during entry. Thus, all eCRF development, edit checks, and dynamic behavior should occur prior to the first subject screening. [VI] Many of the typical CDM start-up activities for EDC studies include finalization of the study protocol, eCRF form design, edit-check specifications, data exchange and integration set-up, study user acceptance testing (UAT), report development, and site training. Many of these activities occur as the study operational plans are being finalized and changes can occur. In most organizations start-up activities are highly cross-functional. Managing the set-up of the EDC system as a project in and of itself is recommended to assure that the system is ready when needed. [VI] SOPs often govern the sequence of and responsibilities for start-up activities adherence as well as documentation of activities performed.

a) Sponsor/CRO EDC Vendor Responsibilities

Though Implementation of the Study application may be performed by contracted vendors, the Sponsor is ultimately responsible for the adherence to regulatory considerations, and final acceptance of the study implementation. A sponsor may choose to build the study in-house, using tools provided by an EDC Vendor, or outsource the build to a third party such as a CRO or independent contractor. In some cases the EDC vendor may be contracted for the study build. When several companies are involved with the database build, it is still necessary for them to have frequent communication and guidance from the Sponsor. At

minimum, the sponsor should retain signatory approval of the EDC build or components of the EDC build such as the eCRF design, edit checks, testing, initiation on production use of the system documentation of the aforementioned activities for the Trial Master File. [VI]

b) International Study Considerations

EDC systems are routinely used in international studies. Many EDC systems have the ability for presenting the EDC interface in multiple languages or collecting the data in multiple languages. CDMs should work with stakeholders to understand language and time zone needs of the study or any components of the eCRF. Issues to consider include the following:

- Whether the local language can be used in a multi-national study. Many coordinators speak more than one language. Asking this simple question or challenging the status quo in this area can avoid unnecessary work.
- Planning enough time for eCRFs that must be translated, rendered in multiple languages, and undergo back-translation.
- Ensuring that the eCRF completion guidelines are available in appropriate languages.
- Understanding how time zone differences will affect time and date stamping of the EDC audit trail, and external data that may be collected in other time zones.
- Consideration of the wording of electronic and manual queries to ensure they will be understood by speakers of other languages.
- Ensuring that helpdesk support has sufficient language coverage to assist sites with system issues in their local language and time zones.
- Understanding how data collected in different languages will be interpreted and used for analysis.

25) Recommended Standard Operating Procedures

ICH E6 states that, “During protocol development the Sponsor should identify processes and data that are critical to ensure human subject protection and the reliability of trial results.” (ICH E6 R2 section 5.0.1) This implies that organizations should map out the processes involved in study design, start-up, conduct, and closeout and make explicit decisions about which are considered to impact human subject protection and the reliability of trial results. Organizational processes may be partitioned differently leading to different scope and titles for SOPs. We provide the following as a list of processes commonly considered to impact human subject protection and the reliability of study results. Organizations may differ as to how these processes are covered in SOPs.

- Data Management Plan Creation and Maintenance
- Document Control (ICH E6 R2 8.0)
- Software Development Lifecycle (Title 21 CFR Part 11)
- System validation and functionality testing including how study eCRFs will be specified, developed, or configured and tested (Title 21 CFR Part 11, ICH E6 R2 5.5.3 b)
- Data collection (ICH E6 R2 5.0)

- Data processing including how medical coding, data review and validation, and integration of external data will be handled (ICH E6 R2 5.0)
- System maintenance (ICH E6 R2 5.5.3 b)
- System change control (ICH E6 R2 5.5.3 b)
- System security measures (ICH E6 R2 5.5.3 b)
- Data backup and recovery (ICH E6 R2 5.5.3 b)
- Contingency planning (ICH E6 R2 5.5.3 b)
- System decommissioning (ICH E6 R2 5.5.3 b)
- Vendor selection and management (Title 21 CFR 312.52, ICH E6 R2 5.0)
- User Access Creation, Modification, and Revocation (Title 21 CFR Part 11)
- User training and support (Title 21 CFR 312.52, ICH E6 R2 5.0)
- Specification, development, and testing of study status reports (ICH E6 R2 5.0)

26) Literature Review details and References

This revision is not based on a systematic review of the peer-reviewed literature. Recommendations from the literature and writing group have been included in the chapter and graded according to the GCDMP evidence grading criteria in the table below.

Evidence Level	Criteria
I	Large controlled experiments, meta, or pooled analysis of controlled experiments, regulation or regulatory guidance
II	Small controlled experiments with unclear results
III	Reviews or synthesis of the empirical literature
IV	Observational studies with a comparison group
V	Observational studies including demonstration projects and case studies with no control
VI	Consensus of the writing group including GCDMP Executive Committee and public comment process
VII	Opinion papers

Regulations, Guidance, and Standards

1. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Guidance for industry Part 11, Electronic Records; Electronic Signatures-Scope and Application August 2003
2. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Title 21 CFR Part 312.62 - Investigator recordkeeping and record retention.
3. Medicines & Healthcare products Regulatory Agency (MHRA), 'GXP' Data Integrity Guidance and Definitions. Revision 1: March 2018. Accessed June 2, 2018, available from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf

4. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER), E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R2) Guidance for Industry, March 2018. Available from https://www.fda.gov/downloads/Drugs/Guidances/UCM464506.pdf
5. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), The General Principals of Software Validation; Final Guidance for Industry and FDA Staff. 2002
6. European Commission, EudraLex, The rules governing medicinal products in the european union Volume 4 Good Manufacturing Practice: Medicinal Products for Human and Veterinary Use, Annex 11: Computerised Systems. 2010.
7. GAMP 5 Guide: A risk-based approach to Compliant GxP Computerized Systems. International Society for Pharmaceutical Engineering (ISPE), February 2008.
8. U.S. Department of Health and Human Services Food and Drug Administration, Use of electronic health record data in clinical investigations guidance for industry. July 2018. Accessed August 8, 2018. Available from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM501068.pdf
9. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Title 21 CFR Part CFR 812.140 Investigational Device Exemption Records.
10. U.S. Department of Health and Human Services Food and Drug Administration, Guidance for industry electronic source data in clinical investigations. September 2013. Accessed August 8, 2018. Available from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf
11. United States Food and Drug Administration (FDA), Title 21 CFR Part 312.52 Transfer of obligations to a contract research organization.
12. GCP Inspectors Working Group (GCP IWG) European Medicines Agency (EMA), EMA/INS/GCP/454280/2010, Reflection paper on expectations for electronic source data and data transcribed to electronic data collection tools in clinical trials, June 9, 2010.
13. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Guidance for industry computerized systems used in clinical Investigations. May 2007
14. International Conference on Harmonisation. ICH Expert Working Group: Guideline for Good Clinical Practice. Geneva, Switzerland: Author; 1996.
15. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Code of Federal Regulations, Title 21, Part 11, Electronic Records; Electronic Signatures. Washington, DC: US Government Printing Office; 1997.
16. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Guidance for Industry: Computerized Systems Used in Clinical Trials. Washington, DC: US Department of Health and Human Services; 2007.
17. Code of Federal Regulation, Title 21, Volume 1, Part 11: Electronic Records, Electronic Signatures. Washington, DC: Government Printing Office; 1998.

Referenced in Text

1. Spilker B and Schoenfelder J., Data Collection Forms in Clinical Trials. Raven Press, New York, 1991. Chapters 1 - 3.
2. Kennedy D and Hutchinson D, CRF Designer, Canary Publications, 2002. ISBN 0-9531174-7-2
3. McFadden E, Management of Data in Clinical Trials, 2nd Ed. 2007, John Wiley & Sons, Hoboken NJ.
4. Avey M, Case report form design, Chapter 3, pp. 47-74 in Rondel RK, Varley SA, Webb CF, Clinical Data Management 2nd Ed. 2000. John Wiley & Sons, West Sussex.
5. Helms RW, Data quality issues in electronic data capture. Drug Information Journal. 2001 35: 827–837.
6. Stead and Lin, National Research Council. 2009. Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Washington, DC: The National Academies Press. https://doi.org/10.17226/12572.
7. Kush RD, Bleicher P, Kubick WR, Kush ST, Marks R, Raymond SA, Tardiff B, eClinical Trials: Planning and Implementation 1st Edition. Thompson CenterWatch, Boston MA, 2003.
8. Zozus NM, Kahn M, Weiskopf N, "Clinical research data quality." Clinical Research Informatics, Health Informatics. 2nd ed. Ed. R.L. Richesson and J.E. Andrews. London: Springer-Verlag, 2018. 175-201.
9. Chhabra, Indian Pacing Electrophysiol J. 2012 Jan-Feb; 12(1): 32–35. Published online 2012 Jan 31. doi: 10.1016/s0972-6292(16)30463-6 , PMCID: PMC3273956, PMID: 22368381 Mouse Heart Rate in a Human: Diagnostic Mystery of an Extreme Tachyarrhythmia, Lovely Chhabra, MD,1 Narender Goel, MD,1 Laxman Prajapat, MD,2 David H Spodick, MD, DSc,2 and Sanjeev Goyal, MD3
10. Zhang, J., The nature of external representations in problem solving, Cognitive Science, Volume 21, Issue 2, April–June 1997, Pages 179-217 https://www.sciencedirect.com/science/article/pii/S0364021399800226.
11. Wickens CD, Hollands JG, Banbury S, Parasuraman R, Engineering Psychology & Human Performance 4th Edition. Routledge, New York, 2016.
12. Earley SS, Data Management Association (DAMA) – Dictionary of Data Management Framework. Technics Publications, LLC., New Jersey 2011.
13. E.F. Codd, A relational model of data for large shared data banks, Commu. ACM, 13 (1970), pp. 377-387
14. Zozus MN, The Data Book: Collection and Management of Research Data. Taylor & Francis/CRC Press, Boca Raton FL, 2017.
15. Labovitz GH and Chang YS, Appendix E, Quality costs: the good, the bad and the ugly, 1987 Organizational Dynamics, Inc., Burlington Mass., in Labovitz GH, Chang YS, and Rosansky V, Making Quality Work: A Leadership Guide for the Results-Driven Manager. 1993, John Wiley & Sons, Inc. Hoboken, NJ.
16. Nahm M. Data gone awry. Data basics, Soc Clin Data Manag. 2008. 14(2).
17. Arndt S, Tyrrell G, Woolson R, et al, Effects of Errors in a Multicenter Medcal Study: Preventing Misinterpreted Data. J. psychiat. Res Vol. 28. No 5, pp. 447 459, 1994

For Further Reading

1. 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
2. Electronic records; electronic signatures. Federal Register. 1997:62;13429–13466.
3. Marks, RG. Validating electronic source data in clinical trials. Controlled Clinical Trials. 2004;25:437–446.
4. 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.
5. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. Perspect Clin Res. 2014;5(4):159–166. doi:10.4103/2229-3485.140555
6. Stokes T. Computer systems validation (6-part series). Applied Clinical Trials. September 1996, January 1997, February 1997, April 1997, June 1997, August 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.
8. 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.
9. US Food and Drug Administration. Guidance for Industry: Electronic Source Data in Clinical Investigation. Washington, DC: US Department of Health and Human Services; September 2013.
10. US Food and Drug Administration. Guidance for Industry: Oversight of Clinical Investigations – A Risk-based Approach to Monitoring. Washington, DC: US Department of Health and Human Services; August 2013.
11. 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. Washington, DC: US Department of Health and Human Services; June 2008.
12. Prokscha S. The data management plan. Practical guide to clinical data management. 3rd ed. Boca Raton, FL: 2012:3-8.
13. Khosla R, Verma DD, Kapur A, Khosla S. Efficient source data verification. Indian Journal of Pharmacology. 2000;32:180–186.
14. Schuyl, Malcolm L, Engel, Thim. A review of the source document verification process in clinical trials. Drug Information Journal. 1999;33:789–797.
15. CDISC (Clinical Data Interchange Standards Committee) web site. Available at: http://www.cdisc.org/. Accessed on March 31, 2008.
16. Nahm M. "Clinical Research Data Quality." Clinical Research Informatics, Health Informatics. Ed. R.L. Richesson and J.E. Andrews. London: Springer-Verlag, 2012. 175-201.
17. Society for Clinical Data Management, eSource implementation in Clinical Research: A Data Management Perspective. June 12, 2014. Available from http://scdm.org/wp-

content/uploads/2016/03/eSource-White-Paper-06122014-FINAL.pdf accessed June 10, 2016.
18. National Research Council. 2009. Stead WW and Lin HS eds. Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Washington, DC: The National Academies Press. https://doi.org/10.17226/12572.
19. http://www.appliedclinicaltrialsonline.com/electronic-data-capture-clinical-trials
20. http://www.clinicalinformaticsnews.com/edc_market_model.aspx
21. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4527052/