Good
Clinical
Data
Management
Practices
Committee

Good Clinical Data Management Practices

Version 4
October 2005

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

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

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

Good Clinical Data Management Practices

Executive Summary

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

> "The review and approval of new pharmaceuticals by federal regulatory agencies is contingent upon a trust that the clinical trials data presented are of sufficient integrity to ensure confidence in the results and conclusions presented by the sponsor company. Important to obtaining that trust is adherence to quality standards and practices. To this same goal, companies must assure that all staff involved in the clinical development program are trained and qualified to perform those tasks for which they are responsible.
>
> The discipline of Clinical Data Management includes paper and electronic case report form (CRF) design, clinical trials database design and programming, data acquisition and entry into the clinical trials database, data review, validation, coding and database finalization. Independent of how individual companies perform these tasks within their company each company is obligated to ensure that the individuals performing these tasks follow Good Clinical Practices. However, currently there are not any published good clinical practice guidelines specific to the discipline of Clinical Data Management. As the professional organization representing Clinical Data Management professionals in North America, SCDM is in a position to develop and publish GCDMP guidelines and to educate CDM professionals on any related regulatory guidelines which have been published. SCDM will continue to educate the industry on the fact that CDM is valuable because we produce quality data."

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

Another two years have passed since Version 3 of the GCDMP was completed. During that time, the GCDMP Committee focused on the stability and future of the document and established a Lifetime Maintenance Plan (LMP) to document the processes that guide changes. In an effort to keep the document current in a changing Industry, this plan defines a formal process and timeline for review by the committee, the SCDM Board of Trustees, the international community, which is represented by the International Network of Clinical Data Management Associations (INCDMA), and the users. Four working subcommittees are defined in the LMP to assist in the maintenance of the document and the LMP itself.

In addition to planning for, writing and putting in place the Lifetime Maintenance Plan, the GCDMP committee finalized a new section (Metrics for Clinical Trials) for the document, as well as reviewed and updated five sections from Version 3. These updated sections will be released when the complete review process has been completed.

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

Acknowledgments
As the committee chairperson, I would like to acknowledge the expertise, dedication and hard work of the document authors. The following individuals have contributed to one or more versions of the GCDMP: Susan Bornstein, Letitia Bowen, Sally Cassells, Anthony J. Costello, Wendy Cuthbert, Bernadette Farrell, Kaye Fendt, Lisa Freeman, Volker Freiman, Imogene Grimes, Marysasser Hedrick Holloway, Susan Howard, Becky Kush, Angel Lazarov, Terrence Loding, Meredith Nahm, Armelde Pitre, Don Rosen, Barbara Tardiff, Lisa Taylor, Beth Wilson. In addition, Lisa Freeman continues to lend her knowledge and skills as Chief Editor, for which we are most grateful. Carol Garvey shared her skills by spearheading the effort to write and put in place the Lifetime Maintenance Plan, Susan Howard led the review and update effort and Meredith Nahm directed the collection and review of comments from users. Kaye Fendt, who initially took the idea of this committee to the Board of Trustees and to interested members of SCDM and who served as Board and FDA Liaison in its early years has continued to lend her expertise to this committee as an innovator, an author, an editor, a supporter, and a motivator. Susan Bornstein led the committee during its formation and coordinated the creation of the CDM Task List, which served as the basis for the organization of this document. Meredith Nahm chaired the committee through 2001, served as Board Liaison through 2004, and has continued to contribute to the review process. Anthony Costello, current liaison to the Board of Trustees, has brought a new focus on exposure and training of the document and new excitement to the committee.

Special acknowledgements are extended to the users who offered helpful comments, the SCDM Board of Trustees and the INCDMA members who participated in the review process. Without their continued interest and support, the GCDMP would not exist. Administrative help (which includes providing the technical expertise needed to post the document and the Lifetime Maintenance Plan) was provided by SCDM's management organization, including Kim Breitbach, Samantha Bordeaux, Rachel McCormick and David Wood.

I am most grateful for your help.

Christine Little, Committee Chair

Anthony Costello, Board of Trustees Liaison

Table of Contents

Executive Summary ..... **ii**

Acknowledgments ..... **iii**

Table of Contents ..... **iv**

Introduction ..... **v**

Data Acquisition **1**

Data Privacy **7**

Electronic Data Capture Principles **12**

CRF Printing and Vendor Selection **16**

Preparation and Preservation of CRF Completion Guidelines **20**

CDM Presentation at Investigator Meetings **24**

Data Storage **27**

Database Validation, Programming and Standards **31**

Data Entry and Data Processing **36**

Laboratory and Other External Data **44**

Dictionary Management **52**

Safety Data Management and Reporting **57**

Serious Adverse Event Data Reconciliation **72**

Measuring Data Quality **76**

Assuring Data Quality **86**

Database Closure **95**

Clinical Data Archiving **97**

Training **103**

Vendor Management **111**

Metrics for Clinical Trials **114**

Glossary **122**

Appendix A: Sample CRF Printing Specifications Checklist **133**

Appendix B: Sample SAE Data Reconciliation Form **134**

Appendix C: Potential Metrics and Definitions **135**

Introduction

The purpose of this document is to provide guidance on accepted practices for the many areas of Clinical Data Management (CDM) that are not covered by existing regulations and guidance documents. The intent is to remain consistent with regulatory practices in related areas of clinical research and to apply the concepts contained in those regulations and associated guidance documents to Clinical Data Management. It is also the intent of this document to provide practical suggestions and proven means of meeting the guidelines recommended in this document. This document is written to serve the needs of multiple audiences: Data Managers, Data Processors, Statisticians, Site Personnel, Clinical professionals, Compliance Auditors, Regulatory Affairs Personnel and all Clinical Research Professionals making decisions regarding or using clinical trial data.

This document addresses the CDM areas of responsibility in 20 sections. Each section provides Minimum Standards and Best Practices. The Minimum Standards and Best Practices summarize the main recommendations of each section in bulleted form. For those wanting an executive summary or an overview of the document, read the Minimum Standards and Best Practices for each section. The Minimum Standards are those that assure that data are complete, reliable and processed correctly, otherwise known as data integrity. The Best Practices are those that offer higher efficiency, quality, and function and lower risk in addition to assuring data integrity. The body of each section provides rational, technical detail and often, discussion of alternate or common practices. References are provided at the end of each section to provide the reader with additional resources. Each section also contains recommended Standard Operating Procedures (SOPs). Whether the SOPs are departmental or institutional in nature, it is the data manager's responsibility to ensure that all data management concerns are addressed.

In the absence of CDM regulatory standards, it is important for experienced, professional data managers to provide thought leadership on accepted data quality levels, practical methods of achieving them and implications of new technology on the CDM tasks. Data Management tasks are often technical and specialized. Therefore, it is crucial that Data Management professionals take a proactive and forward thinking role in setting appropriate expectations and standards for data quality, methodology for quantifying data quality, and auditing practices to assure data quality as the industry utilizes new technologies.

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

Data Acquisition

Introduction

There is arguably no more important document than the instrument that is used to acquire the data from the clinical trial with the exception of the protocol, which specifies the conduct of that trial. The quality of the data collected relies first and foremost on the quality of that instrument. No matter how much time and effort go into conducting the trial, if the correct data points were not collected, a meaningful analysis may not be possible. It follows, therefore, that the design, development and quality assurance of such an instrument must be given the utmost attention.

Scope

This document is intended to address issues pertaining to the systems used in the acquisition or collection of clinical trial data. Acquisition or collection of clinical trial data can be achieved through various methods. Such methods may include, but are not limited to the following: paper or electronic medical records, paper forms completed at a site, interactive voice response systems, local electronic data capture systems, file transfers or central web based systems.

The ICH guidelines on Good Clinical Practice (GCP) use the term 'case report form' or 'CRF' to refer to these systems[1]. No matter what CRF is utilized, the quality and integrity of the data is of primary importance. The following recommendations are meant to assist in the design, development and quality assurance of the CRF such that the data collected will meet the highest standards.

For an extensive discussion regarding creation of CRFs and examples of actual data collection forms, see Data Collection Forms for Clinical Trials by Spilker[2]. The following is meant to highlight some of the most important points to consider during the design process. Specific design instruction for electronic data capture (EDC) systems may be found in the EDC section of the GCDMP document.

Minimum Standards

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

Best Practices

- Design the CRF along with protocol to assure collection of only these data the protocol specifies.

- Keep questions, prompts and instructions clear and concise.
- Design the CRF to follow the data flow from the perspective of the person completing it, taking into account the flow of study procedures and typical organization of data in a medical record.
- Avoid referential and redundant data points within the CRF whenever possible. If redundant data collection is used to assess data validity, the measurements should be obtained through independent means.
- Design the CRF with the primary safety and efficacy endpoints in mind as the main goal of data collection.
- Establish and maintain a library of standard forms.
- Make the CRF available for review at the clinical site prior to approval.
- Use NCR paper or other means to assure exact replicas of paper collection tools.

Design and Development

CRF design should begin as the protocol is being developed to assure that the protocol-specifications regarding data collection are reasonable and achievable[3]. Unfortunately, CRFs are often developed hastily after the protocol has been approved. At that point any data points that are found to be undesirable or unattainable may require an amendment to correct. When the protocol and CRFs are designed concurrently, such collaboration will provide the responsible parties with important feedback.[3] Consideration can be given to what data should be collected and how the data will be used to meet the objectives of the study. If a statistical analysis plan exists, it can be used as a guide to what data points are essential. Often, however, the plan is not final. In this situation, the statistical methods section of the protocol can provide guidance. Although this arrangement may cause the project to incur more time initially, the savings in time and cost brought by avoiding the confusion that amendments can cause should make up that investment.

Protocol-specification should be sufficient to uniquely identify the data to be collected in the study. Collection of extraneous data may adversely affect data quality by distracting site personnel efforts from the critical variables[4]. It is especially important to assure that key variables are defined prior to or during CRF development and that they are appropriately captured on the CRF.

The flow of the CRF should closely follow the data flow from the perspective of the person completing the tool. For example, if the CRF will be completed from a medical record, a log type collection device should facilitate easy transcription of the information and assure that updates to the form are made to the database. However, if the information will be solicited from the subject at each visit, a collection device should be provided for each visit. Additionally, logically related data should be grouped together whenever possible taking into account any limitations or constraints related to the data management system that will be used.

Consider the ISS/ISE or other pooled analyses that will require common data structures to facilitate integration. Collecting the data on forms that promote a common structure will avoid the need for mapping or converting at a later time. To facilitate this continuity,

some organizations have standardized protocol templates, CRFs, database structures, validation procedures and reporting tables.

Documented procedures should exist to insure that the tool accurately collects only the data specified by the study protocol, thus minimizing errors, ambiguity, and redundancy and decreasing completion time. Additionally, procedures should be in place to outline methods for peer and project team approval of CRF design and for control of the original and all subsequent versions of the document.

When the content and format of the CRF is based on a rating instrument created by an independent source, it is important that the validity of that tool be maintained. One example is the use of a psychiatric rating scale or quality of life questionnaire where the subject will be completing the instrument directly. If any changes in content or format are necessary, the independent source should be consulted to insure that the validity of the tool is not compromised by the change. Documentation of the continued validity of the tool should be maintained.

All subject data should be attributable to a subject. Each section that can be separated or viewed separately must contain sufficient identifiers to uniquely identify the data contained in the section.

A provision for investigator signature should be included for timely documentation of the investigator's review of the data as represented.

Clarity and Ease of Use

Questions and prompts should be made specific and clear enough to assure that complete and comparable data are obtained across the various populations using the CRFs.

Provide form completion instructions and definitions for items not directly measurable. For example, “Did the subject have hypertension?” should be clarified by the necessary blood pressure range, length of time sustained or necessity of specific intervention for the condition. Additionally, instructions for the acceptable method of correcting entries should be specified.

Phrase questions in the positive. Multiple negatives in some uses are still negative or, at the least, confusing. For example, use “Did the subject follow the instructions?” rather than “Did the subject fail to follow the instructions?” Leading questions should always be avoided.

Use consistent formatting and take into account the intended use of the tool while designing an aesthetic CRF layout. For example, CRFs completed by site personnel might look quite different from those completed by subjects.

At some point most data must be coded prior to analysis or display. Whenever possible, therefore, data should be collected in coded form. Examples of coded formats include multiple choice pick lists and yes/no check boxes. Careful use of coded formats can both

provide for multiple responses where needed and track the total number of responses while, at the same time, encourage the individual completing the form to select at least one response. In cases where the possible responses are known, they can be conveniently structured into a pick list, and the responses can be coded without biasing the distribution of responses. It is best to design the CRF so that the site does the coding because they are in the best situation to pick the correct assignment considering the availability of source documents and their familiarity with the case. Having them do so minimizes errors and reduces data processing time. With the possible exception of providing information about safety issues, free text is rarely useful without an extensive coding resource.

Maintain consistency in the order of similar answer choices throughout the CRF. For example, yes/no and the placement of a "none", "Not Applicable" or "Other" choice within a series of choices should not change throughout the CRF. The question or prompt should indicate if multiple choices are mutually exclusive or not. If not, there should be an indication if more than one choice will be accepted or if only one selection should be checked.

Careful consideration should be given to the flow of the CRF if the forms will be collected from the site at specified intervals during the study rather than at the end. An example of this might be a twelve-month study where the CRFs will be collected from the site every three months. If possible, a separate collection tool should be provided and labeled for each interval. This will help prevent receiving duplicate information, or worse, nearly duplicate information where a few items previously submitted have been altered. However, there is always the possibility that trial timelines will change or that there will be the need for an interim analysis. If this situation necessitates bringing in data that is not final, the probability exists that the sites will make updates to their originals. Then, when the CRFs are eventually collected at the planned interval, the site's changes can be difficult to see and easily missed. Additionally, if data can span across the specified intervals, such as a continuing adverse event or concomitant medication, mechanisms must be adopted to insure the accuracy of the continuing data.

Referential Questions

Referential questions are created when the CRF is designed so that the answer (or lack of answer) to one or more questions is dependent on the answer to some other question. An example of this is the question "Did the subject have open heart surgery? If no, skip to the next page." This type of question sets up a dependant relationship that demands that both levels be completed correctly or the whole series is in error. Because of the relationship between the levels, referential questions can lead to problems if, for example, while revising the CRF one level is deleted without the other. Additionally, a common error in this situation is for one or more of the dependent questions to be left blank or not taken into consideration when an addendum is written. Best practices require that, when possible, no question refer to another question contained in a remote section of the CRF. Referential questions should be used only after thoughtful consideration. When deemed necessary, these questions should be clearly grouped together and apart from other questions or prompts to minimize confusion.

Redundancy

It is recognized that redundancy in data collection can be used to assess data validity in cases where other means are not practical[5]. If redundant data collection is used to assess data validity, the measurements should be obtained through independent means. For example, two pregnancy tests conducted at the same visit but on different types of samples (i.e. serum and urine) and that produced the same results would suggest that the data are valid.

Data based on the same measurement should not be collected more than once or in more than one place. This situation creates unnecessary work for the investigational site and the need to check for consistencies between the two data points. Similarly, if data are to be combined to generate a composite endpoint, do not collect the initial data more than once from the same source. The raw data or the calculation based on raw data should be collected but not both; raw data are generally preferable[3]. For example, collecting five blood pressure measurements taken two minutes apart and also collecting the average of those five blood pressures should be avoided. Site variations in the arithmetic method used in calculating the average will inevitably cause inconsistencies. The redundant collection of data based on the same measurement can also cause analysis inconsistencies. The raw measurements should be used to calculate the average in one algorithm and a second algorithm should utilize the average as collected. Unless the data are completely consistent, the two algorithms, both reasonable, can produce different results. The raw values are also easier to verify *via* source document. For this reason, a practical approach is to collect the raw data or the calculated data but not both.

Recommended Standard Operating Procedures

- CRF design
- CRF development
- CRF quality assurance
- CRF approval process
- CRF version control process
- Applicable training of site personnel on CRF use

References

1. International Conference on Harmonisation. *Harmonized Tripartite Guideline for Good Clinical Practice*. 2nd ed. Brookwood Medical Publications; 1996.
2. Spilker B, Schoenfelder J. *Data Collection Forms for Clinical Trials*. New York, NY: Raven Press; 1991.
3. Spilker B. *Guide to Clinical Trials and Developing Protocols*. New York, NY: Raven Press; 1984.
4. Califf, RM, Karnash SL, Woodlief, LH. Developing Systems for Cost-Effective Auditing of Clinical Trials. *Controlled Clinical Trials.* 1997;18:651-660
5. Calvert WS, Ma MJ. *Concepts and Case studies in Data Management*. Cary, NC: SAS Institute Inc; 1996.

Other Related Reading

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

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

Proschka S. *Practical Guide to Clinical Data Management.* Interpharm Press.

Data Privacy

Introduction

As the rapidly changing pace of technology produces new breakthroughs in data storage, linkage and mining techniques, policies and procedures regarding data privacy must be re-examined. This section will define the minimum standards and best practices for ensuring data privacy throughout the data management process.

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

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

Privacy protection afforded to research subjects include:

- Protocol review and approval by an Institutional Review Board (IRB)
- Right to informed consent
- Right of the subject to withdraw consent
- Right to notice of disclosure
- Confidential collection and submission of data

Although the majority of issues with data privacy rest with site management or clinical monitoring functions, at a minimum, Data Management professionals should be familiar with basic data privacy issues and follow principles established by their company to ensure the privacy of research subjects and compliance with GCP.

Throughout the data collection and management process, Data Management professionals are often exposed to media that includes primary medical and hospital records, genetic data, economic data, adverse drug event reports and several other sources.

It is not practical to have complete anonymity. Due to concerns for the prevention of scientific fraud and tracking possible consequences of an intervention or treatment while enrolled in a clinical trial, making it more critical that the information is safeguarded to the greatest extent possible.

Scope

This document will focus on the considerations for building and maintaining a platform to maintain a high degree of privacy protection (or security) for research subjects during the data collection and management procedures. With the complexity of clinical trial strategies, data can be transferred between sites, departments, subsidiaries and countries.

Since significant regulatory guidance exists on data privacy, all applicable regulations should be considered for the creation of company policy or standard operating procedures

to ensure full compliance with regulations governing the jurisdictions in which you are conducting business. References for various regulatory documents can be found in the Other related reading section of this document.

Minimum Standards

- Educate all personnel who directly or indirectly handle personally identifiable data on company procedures and data privacy concepts. Training sessions should include company policy, regulatory agency policy and applicable local, state, federal and international law.
- Design data collection instruments with the minimum subject identifiers needed including the design of case report forms, clinical and laboratory databases, data transfer specifications and any other area of data collection that may contain personal information.
- If identified, blind or otherwise address documentation (CRFs, lab reports, etc.) submitted to data management that contains any additional subject identifiers other than those used to link the documentation to a database record.
- Ensure ongoing review and updating of data management processes to ensure consistency with company privacy policies.

Best Practices

- Develop and maintain an environment that respects the privacy of research subjects. Consider long-term employee education programs that highlight the potential impact of lapses in data privacy, benefits of applying strict criteria when handling personal information and ensuring that procedures are in compliance with regulations.
- Implement procedures that occur prior to transfer of data to between sites, departments, subsidiaries and countries that ensure that all considerations about privacy have been considered, addressed and documented.
- Promote internal and external accountability through company policy and regulations governing the use of personal information.
- Collect or use personally identifiable data only when required for specific scientific reasons. Ensure those reasons for use are documented and justified.
- Implement procedures for using data for an alternate or new purpose other than what was originally intended by the informed consent. Ensure that all considerations about privacy have been considered, addressed and documented.
- Enforce a policy of "NO" access to personal data as a baseline. Evaluate any request for this information, and if it is determined that it is required for specific scientific reasons, and ensure that all considerations about privacy have been considered, addressed and documented.
- Special procedures are put in place to store, access and report on extremely sensitive data, such as any type of genetic information.
- Make compliance with data privacy regulations a central focus of audits and a contract contingency when using external service providers.
- Maintain proper physical and electronic security measures. Data should be stored in protective environments relevant to the type of media being stored. Paper case report forms should be stored in an environment with regulated access. Proper precautions

should be taken to prevent external access to data such as password and firewall security.
- If identified, address any data that is submitted to data management that appears to be collected without consent or authorization being secured.

Data Collection

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

Although it is the responsibility of the investigator to ensure that subjects have been given a proper informed consent, it may be beneficial to create a case report form module that contains the question, "Did the subject read, understand and sign the informed consent?" This allows data management to process data into the clinical database with confidence that proper consent was acquired.

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

Recent scientific advances in the area of genetics have now made it possible to store the ultimate identifier, subject DNA. Utmost care should be taken to isolate and protect this data. Strict standards that include completely independent data servers and storage facilities, separate groups to manage genomic data and specific standard operating procedures dedicated to the processing and use of this data should be adopted.

Ensure that any external vendors subscribe to standards that meet or surpass internal standards. For example, lab reports generated from central labs should not contain any subject specific information. This information should be built into data transfer and reporting specifications. As an overall strategy, ensure your company is performing external audits of vendors that include investigations into their compliance with regulations on protection of personal data.

Physical and Electronic Data Security

All data, paper or electronic, should be safeguarded with high standards and those policies and procedures should be reviewed on a regular basis to ensure the latest data protection standards are being implemented. Please refer to the section titled Data Storage in this document to find minimum standards and best practices regarding the physical and electronic security of clinical trials data.

Policy Definition and Training
Corporate policy definition and training should be based on relevant company policy, regulatory agency policy and applicable local, state, federal and international la w. The policy training sessions should address the implementation and maintenance of standards and potential harm to subjects when basic principals are not followed.

Recommended Standard Operating Procedures
- Maintenance of Data Privacy in the Data Management Process

References
N/A

Other Related Reading
Love, Cynthia B; Thomson, Elizabeth J.; Royal, Carmine D., compilers. Ethical issues in research involving human participants [bibliography on the Internet]. Bethesda (MD): National Library of Medicine; 1999 April. (Current bibliographies in medicine; no. 99-3). 4621 citations from January 1989 through November 1998. Available from: http://www.nlm.nih.gov/pubs/resources.html Bioethics Education Materials and Resources Subcommittee

International Conference on Harmonization. *Harmonized Tripartite Guideline for Good Clinical Practice*. 2nd ed. Brook wood Medical Publications; 1996.

United States of America
1. *Code of Federal Regulations*, Title 45, Volume 1, Parts 160 and 164. Department of Health and Human Services. US Government Printing Office, 1998.
2. *Health Insurance Portability and Accountability Act of 1996*, (HIPAA) Public Law 104-191, as amended, 42 United States Code 1320-d. Online: http://aspe.hhs.gov/admnsimp/pvcrec0.htm
3. *Federal Privacy Act*, 1974. PL 93-579,5 USC 552a, as amended. Online: http://www.epic.org/privacy/laws/privacy_act.html
4. US Department of Commerce. *Safe Harbor Framework.* Washington, July 2000. Online: http://ohsr.od.nih.gov/mpa/45cfr46.php3
5. The United States Department of Heath and Human Services Data Council. Online: http://aspe.hhs.gov/datacncl/

European Union
1. European Group on Ethics in Science and New Technologies to the European Commission. *Ethical Issues of Healthcare in the Information Society*. Opinion No. 13, 30 July 1999. Online: http://europa.eu.int/comm/european_group_ethics/docs/avis13_en.pdf
2. European Parliament and Council of Europe. *Directive 95/46/EC of the European Parliament and of the Council 24 October 1995 on the protection of individuals with regard to the processing of personal data and the free movement of such da*ta.

Official Journal L 281, 23/11/1995 p. 0031-0050Online: http://europa.eu.int/eur-lex/en/lif/dat/1995/en_395L0045.html

Australia

1. National Statement on Ethical Conduct in Research Involving Humans. Canberra, 1999. Joint National Health and Medical Research Council/Australian Vice-Chancellors' Committee Statement and Guidelines on Research Practice (1997).
2. *Privacy Act 198*8. Act No. 119 of 1988, as amended. Online: http://www.austlii.edu.au

Canada

1. Canadian Institutes of Health Research. *A Compendium of Canadian Legislation Respecting the Protection of Personal Information in Health Research.* Public Works and Government Services Canada: Ottawa, 2000. Online: http://www.cihr.ca
2. Personal Information Protection and Electronic Documents Act. S.C. 2000. c.5. Online: http://www.privcom.gc.ca

Electronic Data Capture Principles

Introduction

Electronic Research Data Management emerged in the 1970s and has developed over the last 30 years as a suite of processes and tools to enhance the management, quality control quality assurance, and archiving of clinical trial research data. The return-on-investment has been proven for the automation of Clinical Trials Information Management (CTIM) processes from the keying through the summarization and archival processes. Remote Data Entry (RDE) processes also emerged in the 1970s[1], but languished for 20 years without significantly impacting the Clinical Trials arena. However, in the 1990s the development of Electronic Data Capture (EDC) tools for clinical trials research became more focused. Today, the decision to use EDC is up to each company and regulatory agencies are prepared to accept submission for which EDC tools were used. EDC systems of the future should be more than just a means to an end. Quality EDC systems can be the drivers of the whole clinical trials information management process. In addition data managers can provide significant value, based upon their training and strengths, in terms of designing processes to make the transition from paper to EDC efficient while maintaining the data integrity.

Scope

This section provides information on Electronic Data Capture for those companies who have chosen or are in the process of choosing to transfer some or all of their processes from the traditional paper data collection methods to EDC. It concentrates on establishing an environment conducive to incorporating EDC technologies into the clinical trial process from the viewpoint of data management. Practices, procedures, and recommendations are proposed for data managers to operate within the project team to align electronic data capture technology to support statistical and medical research needs. The focus in this version is on general procedures to support EDC for the e-CRF and data integration with non-CRF data.

Recommendations for patient diaries and IVRS will be addressed in a later version of the GCDMP.

Minimum Standards

- Assure compliance with 21CFR11[8] and consistency with FDA Guidance – Computerized Systems Used in Clinical Trials[9.]
- Ensure that User Acceptance Testing is completed prior to implementation and deployment to sites.
- Ensure that software systems validation is scheduled and completed prior to EDC study implementation.
- Ensure that your organizations quality standards support the utilization of automated data capture, management and archival.
- Ensure requirements for data transfers and integration with other systems are defined.
- Ensure sites have access and control of data up to database lock.
- Ensure that access of data is limited to authorized individuals.

- Ensure the availability of technical support for users.
- Ensure training is provided for all users of the EDC system.
- Ensure reinstallation qualification is completed for all repairs and updates to the EDC systems and applications.

Best Practices

- Plan studies to avoid "last minute" system modifications that introduce errors and complexity to the EDC systems.
- Develop e-CRFs or data collection tools with teams of individuals from monitoring, data management, statistics, regulatory affairs, and medical, ensuring adequate attention to the collection of safety data.
- Ensure that systems are user-friendly and flexible for data entry.
- Ensure that the EDC systems do not restrict the answers that site personnel can provide in a way that introduces bias into the clinical study.
- Ensure that adequate data validation procedures and query management tools are built into the EDC study software
- Ensure that data can be traced from acquisition to report and analysis files through easily accessible audit trials.
- Ensure ease and quality of all data transfers and integration with other databases by testing data transfers prior to deployment of EDC systems.
- Ensure processes are defined to integrate laboratory and other non-CFR data with the data from the eCRF.
- Ensure all User Acceptance Tests are documented.
- Define change control procedures for all "user-configurable procedures" such as edit specifications. Ensure these include careful documentation.
- Automate generation of reports on metrics and project status to facilitate project/site/patient management.
- Ensure all documentation for use by site staff is adequately reviewed.
- Ensure the availability of appropriate technical support, especially during early stages of the study. Adequate technical support should be available at any time.
- Develop and follow standard operating procedures (SOPs) for electronic data capture, data validation, and archiving of data.
- Create training materials program to address how work processes will change.
- Integrate metrics on process and cost/benefit into the EDC process to enable better EDC versus non-EDC comparisons and comparisons across EDC technologies.

Regulation and Guidance

The publication of 21CFR 11 in March 1997 and the FDA Guidance for Computerized Systems Used in Clinical Trials in April 1999 provided useful information for evaluating the emerging EDC systems. Then the biopharmaceutical industry began incorporating EDC into their regulated data collection procedures. Thus, implementation of automation from the beginning to the end of the CTIM process was possible. In 1997 the Clinical Data Interchange Standards Consortium (CDISC) issued the following comprehensive definition of Electronic Data Capture (EDC): "The process of collection of data into a

persistent electronic form. This includes data entry (e.g. keyboard EDC, pen-based systems, voice recognition) and automated (or direct) data acquisition (e.g. bar code scanners, blood pressure cuff devices)" [2]. A 2001 updated version of this definition follows:

Electronic Data Capture (EDC): collecting or acquiring data as a permanent electronic record with or without human interface (e.g., using data collection systems or applications that are modem-based, web-based, optical mark/character recognition, or involve audio text, interactive voice response, graphical interfaces, clinical laboratory interfaces, or touch screens). Note: "permanent" in the context of these definitions implies that any changes made to the electronic data are recorded via an audit trail.] [CDISC]

EDC Adoption

A 1998 Association of Clinical Research Professionals (ACRP) survey[3] reported that most clinical trials for biopharmaceutical product development were conducted using paper case report forms for recording the data. The same article reported that "although there are valid concerns that must be addressed, EDC experiences have still been viewed by the overwhelming majority of those surveyed (93%) in a positive or neutral way, i.e. not negatively. When compared to the existing paper process, collecting data electronically provides clear advantages, which will only increase as we improve the EDC solutions." [3]

Changing Roles for Data Managers

The role of data managers will be changed by EDC and the new roles are not yet clearly defined. Therefore professionals in this field should work quickly to define the roles of the CDM in the EDC environment.

Recommended Standard Operating Procedures

- System Setup/Installation/Training
- Data Collection and Handling
- System Maintenance
- Data Backup, Recovery, and Contingency Plans
- Security of the Server
- Security of the EDC software
- Change Control
- Maintenance of Coding Dictionaries
- Vendor Audits

References

1. Ma, JM. (1986) *A Modeling Approach to System Evaluation in Research Data Management.* Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.
2. CDISC (Clinical Data Interchange Standards Committee) www.cdisc.org/

3. Kush, R. D. (1999). Electronic Data Capture: A Survey. *The Monitor*. Fall 1999; 37-40.
4. DIA Workshop on "Electronic Data Capture: Technologies, Implications and Future Requirements", November 1998, Baltimore, MD.
5. Data Basics Summer 2001, p 8 and 9.
6. IIR Symposia on "Automated Data Capture", February 1998, Orlando, FL and September1998, Washington, D.C.
7. Kubick, W. R. (1998). The Elegant Machine: Applying Technology to Optimize Clinical Trials. *Drug Information Journal.* 32; 861-869.
8. Code of Federal Regulation, 1998, Title 21, Volume 1, Part 11: Electronic Records, Electronic Signatures.
9. Food and Drug Administration, April 1999, Guidance for Computerized Systems Used in Clinical Trials.
10. U.S. Department of Health and Human Services, Food and Drug Administration, 1998, Guidance for Industry: electronic Submissions of Case Report Forms (CRFs), Case Report Tabulations (CRTs) and Data to the Center for Biologics Evaluation and Research.

Other Related Reading

Electronic records; electronic signatures (1997). *Federal Register*. 62; 13429-13466.

Hopp, D. I. (1998). Three Topics Integral To The Use of The Internet For Clinical Trials: Connectivity, Communication & Security. *Drug Information Journal.* 32; 933-939.

Hyde, A. W. (1998). The Changing Face Of Electronic Data Capture: From Remote Data Entry To Direct Data Capture. *Drug Information Journal.* 32; 1089-1092.

Hyland, M. E., Kenyon, C. A. P., Allen, R. *et al* (1993). Diary Keeping in Asthma; Comparison of electronic and written methods. *British Medical Journal*. 306; 487-489.
Lampe, A. J. & Weiler, J. M. (1998). Data Capture From Investigators' & Sponsors' Perspectives: Balancing Quality, Speed & Cost. *Drug Information Journal.* 32; 871-886.

Latham, D., "EDC: Establishing standards for electronic data capture", interview of K. Vaillant, Pharmaceutical Visions, Autumn 1998.

Palm, Ulo, MD, MBA, Controlling the Increased Complexity of Electronic Clinical Data Collection and Management Technology by Data Modeling and System Engineering. D*rug Information Journal*. 36; 683-692.

Regalado, A. (1998). Re-engineering drug development, II: Clinical data collection and management. *Start-up*. 27.

Stokes, T. Computer systems validation (6-part series). *Applied Clinical Trials*. September 1996, January 1997, February 1997, April 1997, June 1997, August 1997.

CRF Printing and Vendor Selection

Introduction

The Case Report Form (CRF) is a critical document for capturing relevant data in a clinical trial. While the development of the CRF has been addressed previously, the production aspects of this document must also be addressed. Once the CRF is print-ready, the next step is to have the CRF printed.

Scope

This section will review considerations for outsourcing the printing of CRFs. Use of the following guidelines will help ensure the same quality and service from the contracted print vendor that the Clinical Data Manager expects to receive.

Minimum Standards

- Establish specifications outlining CRF printing and distribution requirements. These should include a complete list of items to be included in the CRF binder, the total number of each item to be printed, type of paper, type of binding, collation order, type and number of tab banks, number of tabs per bank, images to be printed and instructions for printing.
- Provide packaging instructions to the printer.
- Submit new printing specifications (including printing and shipping timetables) to the printers each time significant modifications are made to the CRF or any item outlined in the specifications.
- Obtain approval by appropriate team members of the final print-ready CRF, CRF printing specifications, and the shipping/distribution timetable prior to the submission of the final printing specifications to the printer.

Best Practices

- A vendor qualification program should be in place and utilized in order to select a vendor.
- Other study materials such as pocket cards, study schedule posters, pre-printed return envelopes and study contact information should be printed to compliment the CRF and associated materials and should be distributed at the same time.
- The printer should provide a prototype of the CRF book including all pages, tabs, spine label and cover for review and approval before the final print run.
- A vendor evaluation program should be in place and utilized throughout the vendor relationship.

CRF Binder Specifications

The final print-ready CRF, printing specifications, and shipping/distribution timetable should be approved by appropriate project team members prior to submitting the final printing specifications to the printer. CRF Binder Specifications should include all of the information the vendor needs to produce the CRF binder and associated materials.

To determine the total number of CRFs, diaries or other required pages to be printed, consider the number of evaluable patients required per the protocol, the expected drop-out/replacement rate, as well as a back-up supply. The back-up supply should 10-15% of the total number of patients enrolled. If materials are distributed in packages, overage estimates should take into account the extra items that are in the pack. For example, if SAE forms are printed on a pad of 100 forms, they will be distributed in allotments of 100. Be aware that a site that requires 101 pages actually uses 200 printed forms.

Also estimate the number of CRF papers with a breakdown of the number of NCR pages, non-NCR pages, and other pages (e.g., diary or Quality of Life pages).

Paper
Specify the paper. Include information on the type of paper, color, page weight, hole punch, perforation and gum for each page or section. For example, conventional three-part, no carbon required (NCR) paper comes in many colors and weights. Many organizations use a white, yellow, pink combination or a white, yellow, heavy card stock combination. The type and number of NCR pages required depend on the workflow and system used. Traditionally, white is the original copy, yellow the working Data Management copy and pink the site copy. Scanning or fax-based systems may require only two copies (original white copy for scanning and the site copy).

There are other special considerations with the use of NCR paper. Printer specifications should include a piece of cardboard or other provision for the site to protect unused pages while completing a CRF page. When using a new vendor or a new paper supplier, it is advisable to test the NCR paper. The copy quality on the second or third ply is dependent on the quality of NCR paper. The weight of the paper should also be specified depending on your workflow. Certain weight paper works more efficiently when scanned or faxed. If trying a new vendor or type of paper, test the fax and scan quality.

Consideration for collection of adverse events and concomitant medications must be taken. If AEs and medications are collected at each visit and these pages are harvested at every monitor visit, a pull page system may be used. For example, use four-part NCR paper where the fourth page is harvested first (a pull page) to collect the data earlier. Another option is to have the fourth copy be non-NCR so the next copy reflects only the changes to the data.

Tabs
Tabs are very helpful to the sites in navigating the CRF during the clinical trial. Specify the number of tab banks and number of tabs per bank. Organizing the printing specifications by tabs is also a good way to communicate the collation order to the printer. Also specify the paperweight of the tabs (usually card stock), type and color of mylar dip or other laminate on the tabs, and the text to be printed on each tab or tab page.

Binding, Packaging and Shipments
Specify the type of binding, binder color, width, number of inside pockets, cover text or art and spine label.

Specify the packaging instructions and include a packing list of the items that each site should receive. For example, special forms such as drug accountability logs, screening logs, SAE forms, subject diaries and questionnaires may be bound separately in books or pads. Special forms may also be conveniently shrink wrapped in appropriate numbers for each site.

If the printer is shipping materials to sites, provide shipping instructions. Specify the number of sites and the number of items per site, the shipping company and how shipped (e.g., ground or air). The location of sites should be considered when finalizing timelines. Shipping to international sites may take longer. Include timetable instructions for shipping that include a tracking and inventory process for notification to the sponsor.

Information Commonly Provided With Printing Specifications
In addition to the specifications, the following should be provided to the printer if applicable:

- The final camera-ready artwork of the CRF, diary and other pages in electronic files; the format of the electronic file should be discussed with the printer in advance.
- Specifications for CRF layout (e.g., layout of the CRF identifying location of tabs, tab back instructions, collation order of pages, etc.)
- A list of tabs, including the breakdown by bank and color
- Camera-ready artwork of instructions to be printed on the tab backs
- Company logo and text for spine label
- If the printer is shipping to the sites, a list of sites, addresses and shipping and details on how the printer will know when the site is approved to receive study materials.
- Priorities and specifications for printing the barcode if applicable
- Tentative timetable for final master copy to the printer, review prior to final printing run, date by which shipment must arrive.

The printer should provide a complete prototype of the CRF book, including all pages, tabs, spine label and cover for review and approval before the final print run.

New printing specifications (including printing and shipping timetables) should be submitted to the printers each time significant modifications are made to the CRF or any item outlined in the specifications. An example of a CRF Printing Specifications Checklist can be found in Appendix A.

Selection & Evaluation of CRF Printing Vendors
Print vendors should be qualified. Select print vendors who specialize in CRF printing and have an understanding of the clinical trial process and CRF design. The print vendor should understand the importance of maintaining timelines in the printing and shipping of CRFs before the first patient is enrolled at each site. The printer should be knowledgeable regarding time-to-ship internationally and customs regulations.

Evaluation criteria should include the following: accuracy of printing, quality of service, turnaround time (turnaround time on initial print job and additional requests for extra pages), pricing, CRF design experience, digital or rotary printing, bar-coding capabilities, changes for re-setup, and storage charges. Other criteria to consider is whether the printer outsources parts of each job such as printing mylar tabs, separate charges for printing on tab backs, volume discounts, international shipping capabilities and turnaround times.

Recommended Standard Operating Procedures

- CRF Design
- CRF Production Guidelines
- CRF Printing Specifications
- Vendor Selection

References

N/A

Other Related Reading

N/A

Preparation and Preservation of CRF Completion Guidelines

Introduction

When completion of the Case Report Forms is performed at the investigator sites, it is imperative that the forms are completed as accurately as possible. Training the sites to complete the forms and the site monitors to review the data on the forms according to concise, complete and logical guidelines will result in far fewer queries to the sites. Accurate CRF completion and review results in a more meaningful data analysis[1], quicker validation of data and ensures a timelier database lock.

CRF completion guidelines should ensure that all required fields are completed, and that the data provided within these forms are logical within the scope of the specific protocol. Data on paper forms should be recorded so that they are legible. CRF completion guidelines are tools that provide instruction and guidance on how the sponsors are expecting the case report forms to be completed.[2]

CRF completion guidelines apply to all parties participating in the clinical trial. These include data managers, study monitors, programmers, and statisticians. They also apply to the investigators, site coordinators who enter data, and to CRO personnel who are performing data management, programming, monitoring, or analysis. They should be a specific reference for study monitors, particularly when a study has multiple monitors on a multi-center study, to ensure consistent entry and interpretation of data.

The CRF completion guidelines may exist in various formats and can be used for different types of CRFs or other data collection tools and methods. For traditional paper CRFs, the CRF completion guidelines are printed as part of the CRF or as a separate document. For electronic CRFs or EDC systems, they may be provided as instructions on the forms, an online help system, or system prompts or dialogs generated relative to the data being entered.[3]

Scope

The scope of this section is to describe how CRF Completion guidelines are to be created and used to aid in the precise and logical capture of clinical study data.

CRF completion is not necessarily a paper-based activity. Therefore, CRF completion guidelines should be modified and applied to paper Case Report Forms, as well as remote data entry systems (both local electronic data capture systems and central web-based systems) and interactive voice response systems.

Minimum Standards

- Create CRF completion guidelines for at least every multiple site protocol and document the process by which this is completed and distributed.
- Provide site coordinators and monitors with CRF completion guidelines and training on the use of these guidelines prior to first patient enrollment; document the training.

- Provide Data Management, Biostatistics, and research team personnel with CRF completion guidelines so that these groups are aware of how the sites are instructed to complete the CRFs.

Best Practices

- Develop guidelines in collaboration with representatives from clinical research, programming, and biostatistics.
- Design the CRF completion guidelines from the perspective of the site coordinators and monitors who will be using these guidelines, taking into account the clinical treatment procedures at the site, such as the organization of medical records and methods being used to obtain measurements.
- Establish a formal sign-off procedure for CRF completion guidelines. Include this with the actual CRF sign-off process. Document any changes and maintain version control of the document.
- Present at an investigator's meeting (or similar forum) with Data Management representation to review the CRF and review the CRF completion guidelines. Provide the site staff and the monitors with a correctly completed sample CRF and the CRF completion guidelines at the time of training.
- Stress the importance of not leaving blanks – if a data item is unavailable or unknown, have sites enter N/A or unknown in the particular field. Electronic Data Capture software should ensure that data cannot be left unaddressed by site personnel or ensure that a process is in place to document this (such as a query audit trail)
- Include a general instructions section and a section with page-by-page instructions.
- Include a list of acceptable abbreviations (if any) that can be used within the CRF.
- For paper studies, print CRF Completion guidelines on facing pages of a blank CRF for a given protocol, so that when one views each case report form page, one sees that page's CRF completion guidelines at the same time. Guidelines should be located as close as possible to the CRF fields being completed. Every page should include instructions on proper completion.
- Review data quality periodically, re-educate site personnel, and revise CRF completion guidelines as necessary, particularly for long-term studies. Guidelines may require revision or updating as the protocol or CRF are revised or operational decisions are made.

Format and Content of CRF Completion Guidelines:

CRF completion guidelines can be instructions within a specific section of a given CRF page (e.g., check only one box, Record all medications taken within the last 7 days), additional instructions included within the CRF (e.g., instructions printed on facing pages, such as Adverse Event Guidelines), or a separate document that provides detailed instructions (e.g., CRF completion manual, operational decision memos).

Included below is a suggested format for CRF Completion Guidelines that are created as a separate document, such as a CRF Completion Manual or Clinical Data Monitoring

Guidelines. Those responsible for the CRF Design should determine the formatting of CRF Completion guidelines that are actually part of the CRF.

General Instructions
This section should include information that applies to filling out the entire CRF.

General Instructions should include (at a minimum) the following:
- Ensure the use of permanent media (blue or black ink).
- Ensure that all items in a CRF are completed
- Ensure that all items captured in the CRF are legible
- Ensure that all items in a CRF are captured in a logical sequence
- Provide a list of acceptable abbreviations (may vary between studies or indications)
- Provide contact information if question arises while completing the CRF
- Specify procedure for making corrections to the data. For example, "Corrections to the data should be made by drawing a line through the incorrect entry and then writing the correct response above the original entry. All changes should be initialed and dated."
- Describe study convention for visits or assessments that are not done.

Page-specific instructions
Every page should have specific instructions per the protocol on how the data should be recorded. Keep these instructions brief and focus on critical fields and those that may be interpreted differently.

Page-specific instructions should include (at a minimum) the following:
- List/Explain procedure for clearly reporting:
  - any visits that a subject fails to make, tests that are not conducted, and examinations that are not performed.;
  - all withdrawals and dropouts of enrolled subjects from the trial.
- Provide a description of the "stopping rules" or "discontinuation criteria" for individual subjects, parts of trial, and entire trial. Include the subject withdrawal criteria and procedures specifying:
  - when and how to withdraw subjects from the trial/investigational product treatment;
  - the type and timing of the data to be collected for withdrawn subjects, and how it is to be recorded.
- Provide instructions for determining Adverse Events and Serious Adverse Events

Recommended Standard Operating Procedures
- Preparation, Review, Revision and Distribution of CRF Completion Guidelines
- Study Start and Other Investigators' Meetings
- Training Records
- Study Initiation Process

References

1. *ICH Harmonized Tripartite Guideline for Good Clinical Practice*. 1996. Brookwood Medical Publications, 2nd ed.
2. Spilker B., *Guide to Clinical Trials and Developing Protocols*. 1984. New York: Raven Press.
3. Spilker B, Schoenfelder J., *Data Collection Forms for Clinical Trials*. 1991 New York: Raven Press.

Other Related Reading

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

CDM Presentation at Investigator Meetings

Introduction

The investigator meeting may provide an early opportunity for the data manager to be in contact with site personnel on a clinical trial. It is often a joint effort among the Project Manager, the Clinical Team Leader, the Statistician, and the CDM Lead to describe procedures for preparing, conducting, and managing multi-center investigational trials. A CDM presence at this meeting can provide a well-rounded overview of the entire study. The purpose of a CDM presentation is to familiarize the site investigators and their staff with the protocol and case report forms (CRFs), the study procedures, the completion of casebook procedures or schedules, data collection and cleaning processes, and the electronic data capture (EDC) system and equipment if applicable. At minimum, the data manager should present the CRF completion guidelines at the investigator meeting.

Scope

This procedure commences when the group takes on the responsibility for preparing and conducting the meeting and ends when the required meeting documentation has been distributed.

Minimum Standards

- The data manager should prepare material for the meeting binders.
- Provide sample CRFs and annotated data clarification forms (DCFs.)
- Prepare a flowchart of the overall data collection process.
- Define team members and site responsibilities for data cleaning.

Best Practices

- Conduct a practice session prior to the meeting. Be conscious of content complexity and timing as well as voice projection and inflection.
- Check that a sign-in log is available to provide documentation of meeting attendance, including sponsor representatives.
- Prepare meeting minutes, including all questions and answers in accordance with the scope of work.
- Tape the session for use in subsequent training.
- Allow sufficient time to finalize the CRF after the investigator meeting.

Procedures

- The presentation materials should concentrate on bullet points with the presenter providing an expanded discussion around the brief words on the screen
- Whenever possible, flowcharts or other visuals should be used to indicate a process (AE reporting.)
- CRF presentations should always use completed CRFs as examples. It is often valuable to complete an entire CRF packet as if it were actual patient data. This would allow the attendees to see the proper recording of data for the various treatment scenarios.

- The presenter should demonstrate consistency checks between pages and should point out potential spots for errors. Following are some of the cross checks that can be discussed:
    - Compare current medical history to any concomitant medications. (E.g., if subject has hypertension and is taking medication, it should be appropriate to show that they are taking an anti-hypertensive.)
    - Compare medical history to physical examination. (E.g., if subject has a history of bronchitis, the physical exam may show rhonchi.)
    - Compare termination page to AE page. (E.g., if subject withdrew from the study due to an AE, an AE should be indicated on the AE page and the action taken should be discontinuation from study.)
    - Provide an example where efficacy and safety data show a logical progression. (E.g., compare baseline vital signs with subsequent vital signs.)
    - Make certain investigational product accountability corresponds with dosing regimens outlined in the protocol and drug return logs in the patient packets.
    - Check that visit dates match the visit windows specified in the protocol.
    - Compare subject history and physical examination to basic eligibility criteria.

- The CDM presenter should use this opportunity to explain the data manager's role in the overall scheme of the trial conduction, including but not limiting discussion to the following:
    - The CDM does not simply verify that housekeeping rules are in place, such as consistent initials on all pages.
    - Establish project standards for handling missing data, partial dates, and illegible text entries.
    - Ensure that GCP guidelines are followed by providing examples that indicate the proper mechanism for making corrections on the CRF or fill in the missing zeros.
    - CDM plays a significant part in establishing milestones and meeting timelines as well as providing performance feedback to sites, perhaps in the form of trend reports for DCFs.

- Use this opportunity to walk the audience through the DCF process, including but not limiting discussion to the following:
    - Familiarize participants with various reports that organize DCFs by data item or by file names. Educate on problem areas of the CRF or common mistakes on DCF completion.
    - Show site personnel how to fix problems before the monitor gets there and you will show them how to save time and money.
    - Be sure the sites understand that the Clinical Data Management drives database lock and that the cleaner the data are to start with, the quicker database lock will occur.
    - Explain any status reports the site may receive that are based on data in the database. For example, outstanding CRF reports, outstanding DCF reports, outstanding lab samples. These are often used, especially in studies in which

only a percent of the CRFs are monitored, or where monitoring visits are infrequent.

- Explain the relevance to the sites of the reports and any related workflow information. For example, If sites are paid for CRFs entered as of the 25th of the month, provide the site with the send by or submit date that will assure that that data will be entered and included in the report run on the 25th.
- Describe the procedure for requesting modification of the CRF.

Recommended Standard Operating Procedures

- Data Collection and Handling Procedures

References

N/A

Other Related Reading

N/A

Data Storage

Introduction

The secure, efficient and accessible storage of clinical trial data is central to the success of clinical trials research. Whether data are collected using validated electronic tools or traditional paper forms, data are often transferred many times during a clinical trial. These transfers occur between functional groups within an organization as well as between companies, CROs and regulatory agencies. The potential for data corruption and version control errors during data storage and transfer is significant and must be minimized to assure consistency of results and data quality.

Scope

The scope of this section is to provide key considerations for the data storage and archival of clinical trial data.

Minimum Standards

- During the conduct of a clinical trial, store all original data collected (e.g., Case Report Forms and electronic laboratory data) in secured areas such as rooms or file cabinets with controlled access (e.g., locks). These original documents are to be considered part of the audit trail for tracing back to the source data and should be protected and controlled as rigorously as the electronic audit trail of database modifications or backup procedures.
- Document the procedures for granting access to database servers, establishing system controls and assigning passwords. This process is especially important in a trial where the original data collection is done electronically and no paper backup exists.

Best Practices

- Store clinical data in such a way that backup copies can be made easily and frequently. For example, paper documents should be scanned and electronically archived.
- Utilize open formats whenever possible for archival, storage and transport of data (e.g., ASCII, SAS Transport, Portable Document Format (pdf), CDISC ODM Model). Adherence to this practice enables access to the data by multiple systems or reviewers, currently and in the future.

Physical Storage

The physical security of original data sources (e.g., case report forms, electronic data files, and other original data documents) should be maintained carefully. Original paper and electronic documents should be warehoused in secured rooms or file cabinets with controlled access. Whenever possible, paper documents should be scanned soon after

receipt and archived electronically so that they are included with the backup of other electronic files.

Database servers can be the primary warehouse of clinical data and should be physically secured with appropriate SOPs in place to regulate access. Direct access to database servers should be restricted to those individuals with the responsibility for monitoring and backing up the system. All other access to database servers should be controlled by logical security and occur across a secure network using appropriate system controls and password access controls.

Special considerations must be given to the physical security of computers used in an electronic data collection trial. In cases where data are entered over a live connection to a central database, the physical security of the central server, most likely warehoused with a sponsor or vendor, is a primary consideration. If any data are stored locally at the study site (as in the case of a hybrid or “offline” system) before being sent to a central server, the physical security of the system at the source of data entry is more critical. In these cases, care must be taken to assure the physical and logical security of any computer that stores clinical data for any period of time.

Access permission controls and passwords are vital to assure that only authorized personnel have the ability to access study data. An appropriate administrator designated by company policy should assign permissions on an as-needed basis. Mechanisms should be in place to capture and prevent unauthorized attempts to access a system. Notification should be made to administrator if such attempts take place. A procedure should be established describing the selection of passwords and the frequency with which they should be changed. Passwords should never be shared among individuals or study teams. These operating procedures are designed to minimize the opportunity for data corruption via accidental or intentional manipulation of the electronic raw data.

Trials that are utilizing electronic data collection and management will necessarily regard a user’s authentication (login name and password) as the user’s electronic signature to maintain compliance with Code of Federal Regulations Title 21 Part 11. All data entry and modification should be captured and stored in an audit trail (username, date and time stamps) that regards the electronic signature as evidence of the user’s authority to alter information in the clinical database.

Electronic Storage and Warehousing

In addition to access controls, database design and organization are important considerations for a thorough data storage system. Database validation and associated documentation is the cornerstone of a secure and reliable system. All database validation and user acceptance documentation should be readily available to the study personnel to assure that all database functions being performed on a study have been validated for quality and reliability. Additionally, consideration should be given to assure that project team access to clinical data is sufficient to expedite efficient and high quality interim reporting, data metrics evaluation, and safety reporting requirements (see chapters on "Safety Data Management and Reporting" and "Measuring Data Quality")

The need for thorough validation and trial database design is even more critical for trials utilizing electronic data collection. Data collected on paper CRFs provide the opportunity to re-enter information if necessary due to a software malfunction or unintended loss of information. Because electronic data collection eliminates the paper document as a step between the study trial observations and the database, it is critical that database validation and reliability issues are resolved on the validated system prior to the entry of any actual study information. As electronic data entry moves closer to the point of patient care, electronic information more often will be the source data and, as such, require protection and secure warehousing.

Data Archival

Several archival procedures should be followed to assure that the data are preserved in their raw format. Most importantly, the database itself should be locked upon completion of a study. This means that permissions to further modify the data are removed from all except the most critical study personnel. A thorough study archive includes all of the following:

- Database design specifications - Documentation of the table definitions used to build the study database and file structure.
- Raw data - The final raw data files preserved within the study database format and all original data transfers in their raw format.
- Audit trail - A complete electronic audit trail documenting all modifications by date, time and user identification.
- Final data - It is critical to preserve the final data in a standard file format (e.g., ASCII, SAS transport) so that it can be easily accessed, reviewed or migrated to another system.
- Original study documents – The original and/or scanned images of all original documents. These may be archived separately in a central records facility if necessary.
- Procedural Variation Documentation - If any variation from standard operating procedures or working practices occurred during the conduct of the trial, memos and relevant information about those variations should be included.
- Database Closure - Documentation of each database lock and unlock describing the time and conditions surrounding those procedures (see GCDMP guidance on Database Closure for additional information).
- Site copies of data may be required for audit purposes. If needed, these copies should be locked, read-only datasets delivered on CD-ROM or similar storage medium.

Recommended Standard Operating Procedures

In addition to SOPs, please also reference sections of the GCDMP document "Database Validation", "Data Entry and Data Processing" and "Database Closure". The following SOPs are recommended as a standard for controlling database storage and archival:

- Database validation

- Database design
- Database closure (including procedures for unlocking a locked database)
- Storage of original documents both during and after the trial
- Forms management and *e*-data management (this procedure should cover shipping and handling of original and/or working copies of relevant study documents. If electronic documents and files are used, the SOP should specifically address file transfer specifications and storage for those files.)
- Version/change control for revisions to software
- Version/change control for revisions to hardware
- Version/change control for revisions to data
- Version/change control for revisions to documentation
- Disaster recovery
- System controls and security

It is also advisable for investigational sites to maintain their own SOPs related to physical and logical computer security.

References

1. *Code of Federal Regulations,* Title 21, Volume 1, Part 11. Food and Drug Administration. US Government Printing Office, 1998.
2. Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials.* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; April 1999.
3. Food and Drug Administration . *Guidance for Industry: Electronic Submissions of Case Report Forms (CRFs), Case Report Tabulations (CRTs) and Data to the Center for Biologics Evaluation and Research.* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; 1998.

Other Related Reading

N/A

Database Validation, Programming and Standards

Introduction

The clinical data management system used to conduct study research performs many important functions. 21CFR Part 11 and the predicate rule applicable to the drug, device or biologic in development require that thorough documentation exist at all levels of clinical trials data collection and management. Given the multi-faceted responsibilities of most data management systems, the validation process is often ongoing and can become quite complicated. This section focuses on two primary areas of responsibility:

- Validation of the clinical data management application itself meaning the responsibility of a data management organization to prospectively validate a clinical database that was purchased and installed for the purpose of performing data management tasks.
- Validation of programming related to the development of a study or protocol-specific database.

The distinction between these two cases is often vendor-specific. For example, one database management system may allow the user to map data entry fields to the variable in the database where the data are stored. In this case, the mapping should be addressed in the study-specific programming validation. Other products may perform the mapping internally based on the entry screen design. In this case, the mapping function would be validated once, during the initial application installation or upgrade and not require validation on every study. A practical suggestion is that only the new code, interface to new code or new functions need to be validated provided that it can be ensured that corruption to other code or functions did not occur.

Scope

This section addresses the validation activities that should accompany the installation of a database management system and its patches or upgrades, as well as the testing and validation necessary when designing specific trial databases on that system. This section does not address the validation associated with software applications development during the software development life cycle or development and validation of statistical analysis tools. The validation measures necessary for software development and proprietary systems design are very different and more complex than the process of validating an ‘off the shelf’ data management system. Organizations developing database management systems are referred to 21CRF Part 11 and the associated Guidance Document.

Although the software development life cycle validation approach advocated in the device and GLP regulations is appropriate for application development, it is not appropriate or practical for the set-up of individual protocols within a validated database management system.

Minimum Standards

- Validate database management systems in their local environment prior to use.
- Test the set-up and programming of each study managed within the validated database management system to assure that any new code generated or new system functionality used for the study, works according to the user's specifications.
- Define the testing methodology, scope, problem reporting and resolution, test data and acceptance criterion within the Test Plan.

Best Practices

- Identify all intended uses of study-specific programming.
- Use organization standards to document programs.
- Make use of code libraries.

Clinical Database Validation

The primary concern that must be addressed during database validation is the dependability of the clinical database. Study data will be maintained in the database management system and, ultimately, the outcome of the clinical trial rests on the integrity of those data and the reliability of the data handling process.

When implementing and validating a clinical data management system, the sponsor must ensure that all tests are designed to conform to established requirements for completeness, accuracy, reliability and consistent intended performance. Data management software purchased off the shelf should have been heavily tested by the vendor who originally wrote the software. These development validation measures are typically referred to as 'design level validation' and do not require repeating by the end user. Ideally, however, documentation of the design level validation specifications and testing will be available to the sponsor, at least for audit purposes and contractually through escrow.

Sponsors should assure the completion and documentation of functional level testing. This includes documenting the effect of any known limitations, problems and defects on functions used for the Sponsor's study. Some examples of additional validation that might be performed and validated by sponsors using test data include the following:

- Data entry screen testing to assure that data are mapped to intended database structures
- Validation of any generic integrity constraints or data checking routines that execute during data entry (e.g. range, date, format, coding, field discrepancies)
- Testing of data verification functions such as second entry verification, file comparison and batch verification
- Batch data transfer to the clinical trial database from separate data entry systems (e.g. electronically transferred data or remote data entry systems)

Some important documentation to be considered and addressed during database validation includes the following (see also Recommended SOPs):

- Prospectively written design specifications that describe what the software is intended to do and how it is intended to do it (A version controlled system manual can serve as design specifications.)
- A test plan based on the design specification and the intended use of the database management system that outlines both how the testing will be conducted and the criteria for acceptance or rejection of the software based on testing outcomes.
- Results documentation that describes why the software will be accepted or rejected based on specific test results.
- Appropriate review and approval documents.

Trial-Specific Programming and Validation

After a clinical data management system has been validated and approved for use within a sponsor organization, validation must focus instead on specific study or protocol database design and implementation. Validation at this phase can be addressed in three major categories: database design, data entry or capture and other study-specific programming. Following are some specific considerations to be given at both levels.

Database design should be based on standard data architectures within a sponsor organization. Utilizing standard ways of designing study databases and maintaining study data allow validation efforts and results to be easily documented, maintained and leveraged across many projects. If there are data structure libraries available, the templates should be accessible and adaptable where necessary to accommodate specific and unique project requirements. When standards are not available, efforts should be made to keep database design and data structures as consistent as possible within projects and, wherever possible, across projects. For example, data structures developed for Phase I trials should be used throughout Phase II and III trials wherever appropriate. If use of standards is not possible, as in the case of a contract organization designing a database according to sponsor specifications, the specifications are sufficient.

Database Specifications should, at a minimum, provide the following for every variable:

1. Name and label
2. Dataset label, panel or other logical group in which the data belongs
3. Type (numeric, character, integer, decimal, date)
4. Length (including number of characters before and after the decimal point where applicable)
5. Definitions for all coded values
6. Indication of the origin of the data
7. Algorithms for derived or calculated variables

The use of standards can greatly simplify the specification process by providing the specifications for standard items, or a shorthand way of indicating standard items so that the relevant information can be obtained from the standards.

There are several considerations to address when testing a study data capture system. Most important is the need to assure that data being entered through a data entry screen or

captured via some other transfer process (e.g. electronic lab data transfers) map to the correct variables in the clinical trial database. Useful validation measures to consider when testing this functionality may include entering test or 'dummy' data into the screens or loading test data transfer files so that output data listings from the database can be reviewed to assure that the variables were correctly added and saved within the database structure. It is also critical to validate the data field definitions in terms of length and type. Will all study data be accepted by the database? Are variable lengths sufficient to prevent truncating or rounding? Do character and numeric formats provide the necessary output for analysis files, query management software and other modules within the sponsor's overall clinical data management system? If the database is programmed to flag out of range data, are those flags being appropriately triggered during data entry or import?

Database entry or capture validation should also verify that key records management issues are addressed. It should be assured that the database will not accept entry of duplicate records and that those primary key variables are appropriately assigned and managed by the database structure definition. When discrepancies between first and second pass data entry are resolved, validation should assure that one record with the correct data is permanently and correctly inserted into the study database and can be extracted. Most importantly, the audit trail for the trial should be validated and protected so that all manipulation of the study database or external files is recorded by date, time and user stamps in an audit trail that can be accessed as read-only for the life of the data.

Other examples of study-specific programming are data loading or transfer programming, and programming done to validate the data (edit checks, query rules, procedures). This includes any code written to check the data and can occur at the time of entry or later as a batch job. This programming must be validated if action is taken regarding clinical data intended for submission as a result of the programming. Examples include programming that identifies data discrepancies such that queries are sent to clinical investigators or in-house data editing conventions followed for items identified by the programming.

Best practices include identifying all intended uses of study-specific programming, and testing each logic condition in the programming based on a test plan. Algorithms for derived variables must be validated.

Practical suggestions include utilizing organization standards to document as much of the programming specification and test plans as possible, and code libraries to reduce the amount of new code generated for a protocol. The entire test plan can be a Standard Operating Procedure that contains the testing methodology, scope, purpose, acceptance criterion and approvals and specifies the format for test data and problem reporting.

Recommended Standard Operating Procedures

- System installation
- System validation
- Change control - outlining most importantly the validation procedure when a major system component is upgraded or changed.

- Data backup/recovery - outlining most importantly the frequency and archival strategy of database backups and the specific plan for recovery in the event of a system crash or other disaster resulting in loss of data.
- Trial-specific database design
- Validation of trial-specific database design
- Trial-specific database programming
- Validation of trial-specific database programming
- Data collection and handling
- Receipt of electronic data
- Database and systems security
- Site qualification

References

1. Food and Drug Administration. *Guidance for Industry: Computerized Systems Used in Clinical Trials.* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; 1997.
2. Association for Clinical Data Management and Statisticians in the Pharmaceutical Industry. *Computer Systems Validation in Clinical Research*, 1997.
3. International Conference on Harmonisation. Good Clinical Practice: Consolidated Guideline. *Federal Register.* 1997; 62(90): 2571.
4. *Code of Federal Regulations,* Title 21, Part 11, Electronic Records; Electronic Signatures. Food and Drug Administration. US Government Printing Office.

Other Related Reading

N/A

Data Entry and Data Processing

Introduction

The purpose of data processing is to assure that data are reliable, complete and accurate. Data processing includes receiving, entering, cleaning, coding, reconciliation and transferring of data. The goal of data processing is to efficiently produce quality data for analysis. The Institute of Medicine defines quality data as data that supports conclusion and interpretations equivalent to those derived from theoretical error-free data.[1] Clinical trials vary widely in study design and operational plan and address the unique information required of each trial. Throughout the trial, each component and step in data processing should assure that an appropriate level of quality is maintained. In fact, ICH E6, Good Clinical Practice states, "Quality Control should be applied to each stage in the data handling process to assure that data are reliable and processed correctly.[2]

With electronic data capture systems, traditional data management roles change. In some cases, site personnel will now do data entry. Sites will also have the capability to run edit checks and make data edits. The data manager may no longer actually make edits to the data but must guide the sites through the data clean-up process remotely. This can be done via automated checks that are built into the computer system as well as through queries typed into the computer system. The benefits of electronic data capture include a more stream-lined process of data entry and data cleaning, and the ability to identify safety issues earlier.

Scope

The scope of this section focuses on the CDM functions of data processing, including receipt of data, review of data and data cleaning (discrepancy identification and resolution) in an electronic data capture (EDC) or a paper system. For EDC, this also covers the process of making sure that sites are entering their data in a timely fashion.

Minimum Standards

- Utilize written procedures that describe the data processing steps and required quality level. Ensure enough specificity to reproduce the analysis database from the source documentation.
- Ensure that all procedures, guidelines, working practices, or references are current and available to employees for the tasks they perform.
- Apply "Quality Control . . . to each stage in the data handling process to assure that the data are reliable and processed correctly."[1] An example of a quality control measure is to periodically inspect samples of data then take corrective action if errors are found. In such cases, data of poor quality (errors) are classified as being the result of poor processes. (Refer to the Assuring Data Quality section in this document for more details.)
- Provide descriptive guidelines for staff who write queries for investigator sites.

Best Practices

- Address the purpose, characteristics and complexity of each study in the data cleaning procedures.
- Define the list of critical variables (usually those of primary and secondary safety and efficacy and subject identifiers) before the data cleaning activities are defined.
- Monitor data processing production functions adequately to assure stable and desirable quality levels consistent with the needs of the trial.
- Provide easy access for all employees to documentation for all current procedures, guidelines, working practices or references for the tasks they will perform.
- New CDM production processes should be tested in a test environment by employees at the same level as those who will perform the processing steps, using the proposed working practices, guidelines or references prior to production work being done.
- Provide CRF completion/data entry instructions and helpful tips for both paper and electronic systems. For EDC, such a document might contain information to clarify how to enter particular forms or information about data cleaning. For example, if the patient withdrew from the study while on study medication, for the purpose of data cleaning there should be instructions to look for an Adverse Event where the action taken on study medication is drug stopped. In a paper-based system, there also should be data review guidelines that include issues such as described in the previous example.
- Inform sites of timelines for data entry, running data checks and replicating the data.
- Establish database quality criteria. Establish a quality control plan with the study team that appropriately addresses the primary efficacy data.

Data Receiving

There is great diversity across the industry regarding data receipt. Data are acquired through fax transmission, regular mail, express delivery companies with tracking ability, private couriers, hand delivered by monitors, entered over the web or transferred using other electronic means. Regardless of the data acquisition mechanism, the process by which data are received, confirmed as received, and made available for data processing should be documented in sufficient detail so that the origin of the data and the data received is documented.

It is not always necessary or efficient to track every page of data as it is received. Alternative approaches include the following:

1. Tracking pages as received when they are data entered. This can eliminate an extensive and expensive manual process and replaces it with an electronic process where the tracking is a cost-free by-product of the data entry step. The trade-off is that the steps between receipt and entry should be minimal and for reliable received dates, there should not be a backlog of entry, but a steady and predictable flow.
2. Tracking whole submissions. For example, tracking Visit 2 as received and noting as exceptions missing pages.

Computer-aided page checking can be of higher integrity and efficiency than manual processes. Regardless of the method used, data receiving procedures should facilitate timely, high quality data processing.

Procedures should be in place to ensure blinding of information submitted to the data center as to subject identifying information (i.e., name, address, social security number), unless collection of these data is provided for in the informed consent, protocol and local regulations.

Expected visit date reports can be programmed into most reporting and tracking systems to follow a patient's progression through the study as well as to predict the last patient's last visit dates. This also helps to monitor the timeliness of the site's handling of study data.

Data Entry

The data entry process should address the data quality needs of the trial. Commonly used data entry processes include the following:

1. Independent double data entry with a third person compare where two people enter the same data independently and a third person resolves any discrepancies between first and second entry;
2. Double data entry with blind verification where two people enter the data independently and any discrepancies are resolved during second entry;
3. Double entry with interactive verification where the second entry operator resolves discrepancies between first and second entry and is aware of the first entered values;
4. Single entry with a manual review; and
5. Single entry with no manual review.

In the transition from heads down data entry to single entry EDC, there is a difference in the type of training and the skills required. For heads down data entry, the skill emphasis is on the number of keystrokes made and specific training on the database to be utilized. With electronic data capture systems that utilize single entry, the skill emphasis is on the understanding of the trial. Site personnel entering data in an electronic data capture system have received training on the specific system being used in the study as well as on study specific issues.

FDA and ICH regulations do not require double entry or any other specific data entry process. The skill level of the resource and the time available should be contributing facts in choosing a data entry process. In general, double data entry is helpful where data are subject to frequent random keystroke errors or where a random error would be likely to impact the analyses. A single entry process with good manual review can be better than a sloppy double entry one.

More companies are now conducting trials by electronic data capture. These personnel are trained on the specific data capture system being utilized in the study as well as the protocol and key data issues. CRAs then verify the data using source documents. Often

tick boxes or certain fields in the eCRF are used by the CRAs to indicate which fields and visits were verified. In many systems, this source document verification box becomes unchecked when data in the panel are changed.

The site should have clear guidelines regarding the sponsor's expectation of timing from a patient's visit until data are entered into the eCRF or onto the paper CRF. The data management team is often responsible for monitoring this via reports. For electronic data capture, the sites should be contacted if they are falling behind in this task. Even though the sites are entering and cleaning the data, resources are still necessary to manage the data. This includes training sites on the system, measuring site progress on data entry and cleaning, working through late forms and data discrepancies with the sites, looking at aggregate data for outlying sites, identification of data trends, coding and data transfers.

Systems that store automatic default values should not be used.[4] Automatic default values are those that are written to the database without action by the entry operator. Values that are derived, converted, calculated or hard-coded based on the value of an entered field do not constitute automatic default values.

Where applicable, system parameters should be set so that entry operators can exit the entry screen without saving the data they have entered. This enables the operators to correct, upon discovery, situations where data may have been erroneously entered. Requiring a consensus decision to save the data can contribute to a higher level of integrity. If the system does not allow for this technique of correction, a documented method to correct erroneously keyed information should exist.

Entry screens should be designed to minimize data entry errors. For paper studies, the database should follow the pages in the CRF or may even have screens that look like the paper CRF. With electronic data capture, this may include screen facilities such as radio buttons and pick lists. Suggestions for minimizing entry errors include printing coded values and conventions that can be easily displayed on the entry screen, clearly labeling entry fields, and providing sufficient space and convenience to enter expected data. The eCRF can be designed to have dependencies for fields that should only have data when other criteria are met. An example of this is allowing the entry of text for race only when the choice from the pick list is "other."

Electronic data capture tools usually can be designed to check variable type, allowing character or numeric variables only where appropriate. In addition, the system can be built to provide for numeric values to be checked against predetermined ranges upon entry.

User Manual for EDC

A user manual can be either a paper document or an on-line help manual. On-line manuals often contain print options to produce a hard copy version. Topics to consider for the User Manual include the following:

- Contact information of the individuals available to troubleshoot computer problems and the hours that such help is available.

- Instructions on how to enter data, delete data, and respond to queries.
- Instructions on how to enter data for single and multiple record panels if there is a difference in the system.
- Instructions regarding what to do if the system's date or time clock is incorrect. (The users themselves should not be able to change the internal system time stamp, as it is part of the audit trail.)
- Reminders to the users that there is a date/time stamp and a user name attached to every record, which is recorded as part of the audit trail. The audit trail may be visible or not, depending on the computer system. This audit trail must be readable by inspectors and auditors even if it is not visible during data entry.
- Information on computer system security.
- Instructions for proper computer shut down procedures to prevent loss of data.

Data Cleaning

Data cleaning or validation is a collection of activities used to assure the validity and accuracy of the data. Activities for data cleaning may include manual review of the data, or computer checks designed to identify inaccurate or invalid data using ranges, completeness, protocol violations and consistency checks or aggregate descriptive statistics to detect strange patterns in the data. Cody defines data cleaning to include the following:

- Making sure that the raw data were accurately entered into a computer-readable file.
- Checking that character variables contain only valid values.
- Checking that numeric values are within predetermined ranges.
- Checking for and eliminating duplicate data entries.
- Checking if there are missing values for variables where complete data are necessary.
- Checking for uniqueness of certain values, such as subject ID's.
- Checking for invalid date values and invalid date sequences.
- Verifying that complex multi-file [or cross panel] rules have been followed. For example, if an adverse event of type X occurs, other data such as concomitant medications or procedures might be expected.[4]

Note that many of the above listed data cleaning goals can be accomplished within a well-designed electronic data capture system and do not require post-entry effort as in a traditional paper-based system. In the electronic data capture environment, the site does much of the data cleaning at the point-of-entry. The site is in control of the data and must either make the data edit, or clarify the reason the data is acceptable.

It is important that the data manager know the number of characters the EDC system will allow in queries. Many data managers are accustomed to paper queries where space may be unlimited. In most data capture systems, long queries require a scroll bar for viewing the query. It becomes more important to write succinct queries that instruct the site to either correct the data or explain the reason for the discrepancy or "abnormal" data value. It is also important to be certain that the sites are not receiving the same data queries manually that are built into the computer system.

Range checks should be designed to identify statistical outliers, values that are physiologically impossible or outside the normal variation of the population under study. Consistency checks should be designed to identify potential data errors by checking sequential order of dates, corresponding events, and missing data indicated as existing elsewhere. Checks designed to identify protocol violations should be monitored so that timely action can be taken. Aggregate statistics or other checks indicating that a site is substantially different from the others should be monitored and investigated. Although there are some cases where manual review for data cleaning and validation is sufficient, programmatic validation provides high consistency and lower error rates.

Primary and secondary endpoints, key safety parameters and fields that uniquely identify subject data should be validated sufficiently to assure that the data are possible, complete (according to the normal variation of the population under study) and reasonable. Data cleaning and validation procedures should be neutral and should not suggest bias, or lead responses. Leading questions or forced responses can bias the results of the trial. Values identified through data cleaning and confirmed by the sites should be accepted as confirmed and handled by appropriate statistical methods.

In electronic data capture, there may be checks built into the system that initiate either at the time of data entry or when running the edit checks on batches of data. Other data management edit checks may be run and reviewed prior to queries being issued.

It is important for the data manager to understand the system and how the data checks are attached to the data fields to ease review and possible correction by the sites. When checks are not issued against the correct panel, the site may not take the appropriate action due to confusion. It is important for the data checks to be written to check a data field only once if these are to fire automatically. Data coordinators should be able to readily review previously issued data checks to prevent duplication of effort.

Since the site must respond to data queries prior to any in-house review, it is critical that the checks are properly tested prior to deployment. If inadequately tested checks are installed, unnecessary rework for the sites and for the data management team results.

Early in the study, a review of data for several patients from several centers can help detect problems regarding either data entry screens not functioning as expected or a site's lack of understanding of or compliance with the protocol.

New users of electronic data systems often do not realize that the data that is clean today may not be clean tomorrow if the site goes back and changes the data for any reason. These data changes may not be the result of data queries but rather a re-review of the source data. Some systems have the capability of locking data once it is clean. If this is the case, there must be a mechanism to allow data changes when the site finds discrepancies.

Documenting Data Changes
Data may be changed as a result of data cleaning procedures. Both the site and the data center must retain a record of all changes to the data. Data changes should be recorded and documented by a fax or original site signature acknowledging the new data. This is usually accomplished using a query or data clarification form. In these cases, the site is expected to keep a record of the change with their study records.

In an electronic data capture environment, the sites should do changes to the data. If non-site personnel make data changes, there should be a clear IOP/SOP documenting the circumstances where data can be changed and a record of any data change should be provided to the site. All documentation of data changes is considered to be essential study documentation and is subject to audit.

If any telephone conversations with the site are utilized to authorize data changes, these changes should be documented both by the site representative authorizing the change and by the data center representative talking with the site. In this way, a record of the conversation and authorization exists at both locations.

Data cleaning conventions may, under some circumstances, specify data that can be modified without a site's acknowledgement. Examples include appropriately qualified personnel correcting obvious spelling errors, converting values when the units are provided or filling missing identifiers when the true values are obvious. Because the site must have a record of all data changes, the site should receive and maintain a copy of each version of such data conventions.

An audit trail should reflect all deletions, updates or additions to data after initial entry. The audit trail should reflect the date and time of the change and the user making the change.[3] There should be clear documentation of the change as well as the reason for the change. In order to obtain consistent, accurate reasons for change, some electronic data capture systems offer a list of reasons for data changes well as an option for free text. Since these reasons cannot be changed, there should not be a default entry.

Medical Coding
If textual or free text data are collected and reported, they usually must be coded before they can be aggregated and used in summary analyses. Medical history, adverse events, procedures and medications are usually coded with standard dictionaries. The coding process consists of matching text collected on the CRF to terms in a standard dictionary. There are often items that cannot be matched, or coded without clarification from the site. Ulcers, for example, require a location (gastric, duodenal, mouth, foot, etc.) to be coded.

Protocol Amendments
Protocol amendments are a fact of life in pharmaceutical trials. Changes to the protocol may be made when new information becomes available or is requested by the sponsor or regulatory agencies. While not all protocol amendments require CRF changes, procedures should be in place to handle them when they do. With paper processes, CRF changes may take a few weeks to be received by the sites. By that time, the site has usually received

IRB approval. However, with electronic data capture, eCRF changes can be done remotely and sent to sites for receipt the next time they open the study database. In the case of a protocol amendment, IRB approval must be received prior to deployment of the eCRF change (s).

Recommended Standard Operating Procedures

- Data receiving
- Data security
- Data entry
- Data validation
- Computer security
- Computer system validation
- Specification and testing of data validation checks
- Data change procedure for electronic data capture studies
- Deployment of data validation checks for electronic data capture studies
- Medical coding

References

1. Davis JR, Nolan VP, Woodcock , Estabrook RW, eds. Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making, Workshop Report, Roundtable on Research and Development of Drugs, Biologics, and Medical Devices, Division of Health Sciences Policy, Institute of Medicine; 1999. Available from National Academy Press, 2101 Constitution Avenue, N.W. Box 285, Washington, D.C. 20055
2. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice*, E6, Section 5.1.3; 1996. Available from US Food and Drug Administration.
3. *Code of Federal Regulations*, Title 21, Part 11, Electronic Records; Electronic Signatures. Food and Drug Administration. US Government Printing Office.
4. Cody, R. *Cody's Data Cleaning Techniques using SAS Software*. Cary, NC: SAS Institute Inc; 1999.

Other Related Reading

N/A

Laboratory and Other External Data

Introduction

Often during the conduct of a clinical trial, much data external to the Case Report Forms (CRFs) will be collected. If not included in the primary safety or efficacy parameters, these data can be used for subject screening, routine safety and quality of life monitoring or trend analysis. In order to speed up this process and minimize the use of different analyzing methodologies and equipment, it is common that sponsors refer to the use of centralized vendors. Such vendors provide electronic transfer of computerized data into the sponsor's database, thereby offering quick results, standardized testing and reference and calibration values applied to data collected across study sites, and the potential to eliminate transcription errors and key entry of data. This section focuses on the external data most often required in clinical trials, their structure and handling.

Scope

What follows is the Data Management prospective of the challenges involved in incorporating any external data into a clinical database while assuring that the quality, integrity, confidentiality and plausibility of the clinical information is maintained. Further, processing steps that affect the data quality are identified and a solution framework proposed.

Since regulatory guidance exists or data interchange standards are already proposed, this section will reference on a smaller scale, but not attempt to fully cover, the areas of providing data for regulatory submissions, clinical data interchange standards (FDA[1], CDISC[2]) and validation of computer programs (FDA[3], ACDM/PSI[4]).

Minimum Standards

- Establish the procedures for collecting, transferring, loading, validating and editing external data through sponsor and vendor collaboration.
- Identify and involve vendors as early in the process as possible.
- Identify key individuals for communication and follow through.
- Provide written specifications, identify and agree upon, in advance, mandatory fields or critical variables for loading external data into the sponsor's database.
- Maintain a documentation trail.
- Ensure that parties involved have written Standard Operating Procedures and documentation to support that they are followed.
- Have written procedure for safeguarding the blind when primary efficacy data are collected externally.
- Apply quality control procedures to each stage of data handling to ensure that all data are reliable and have been processed correctly.

Best Practices

- Audit external data providers on regular basis as part of your vendor audit practice (also refer to GCDMP Vendor Management).

- Enforce a formal data clarification process for handling data discrepancies and data updates.
- Validate all programs and systems used for processing clinical trial data used in a clinical research environment (also refer to GCDMP Database Validation, Programming and Standards).
- Provide vendor-specific training. A clear understanding of what is expected by both sides is critical for quality and efficient conduct of the clinical research.

Types of External Data

External data can originate from different sources, but it is a common practice for a centralized vendor to specialize and produce one or more major data types. The following are examples:

- Laboratory and PK/PD Data
- Device Data (ECG, Flowmetry, Vital Signs, Images and other)
- Electronic Patient Diaries

It is of a significant importance to identify and describe the variables that must be included in any data transfer, regardless of where the data originate or the information contained within. The purpose of these variables is to merge the external data to the sponsor's clinical database, safeguard the blind, and assure that data belonging to a particular protocol, investigator and subject cannot be loaded to a subject enrolled into a different protocol or to an incorrect visit. Working with the end goal in mind, one can observe that these data may constitute an integral part of the dataset domains proposed by FDA/CDISC[1,2] for submission:

Dataset	**Description**
DEMO	Demographics and subject characteristics
DISPOSIT	Disposition
EXPOSURE	Drug exposure
AE	Adverse events
CONMEDS	Concomitant medications
CHEM	Labs – chemistry
HEMAT	Labs – hematology
URINE	Labs – urinalysis
ECG	Electrocardiogram
VITAL	Vital signs
PE	Physical examination
MEDHIST	Past medical history

Refer to CDISC for additional information.

External Data Processing Steps Affecting the Data Quality

The following areas may adversely affect the integration of external data and should be accounted for during database setup:

- Definition of key variable and mandatory fields
- Data editing and verification procedures

- Record formatting and file formats (e.g. SAS®, ASCII)
- Data transmission
- Database updates
- Data storage and archiving

Key Variables
To assure that sufficient information is available to identify and process data at the sponsor's site, it is imperative that key variables (those data that uniquely describe each sample record) be carefully selected. Without such variables it proves difficult, if not impossible, to accurately match patient, sample, visit, with the result records.

While these variables are intended to uniquely identify and clarify subject visit records, incomplete data collection or presentation of errors in either primary or secondary key variables can result in inadequate information. Therefore, completeness in the choice of variables collected and transferred offers means of increasing the accuracy and thus the overall quality of the process. Primary (Protocol Subject identifiers) and secondary (additional Subject and unique vendor identifiers) key variables can include the following:

Primary Key Variables (Protocol subject identifiers)	**Secondary Key Variables (Additional subject and vendor identifiers)**
Sponsor Name / ID	Subject's Gender
Study / Protocol ID (any combination of project and protocol)	Subject's Date of Birth
Site / Investigator ID	Subject's Initials
Subject Identifier (Subject Number, Screening Number or number assigned by the CRF used)	Transmission Date / Time
Clinical Event ID (Visit Number)	Date associated with the Subject visit
Sample ID (vendor or device specific sample identifier or a subject visit)	Sequence Number (when more than one observation per record exists)

Data acquisition forms, whether conventional or electronic (CRF, e-CRF), should be designed to facilitate the full and accurate reporting of key information at the study site.

Parties involved in the process should identify in writing and agree upon in advance key variables or fields for loading external data into the sponsor's database. Avoid duplication of information. For example, if Subject Initials and Date of Birth are already in the database from the CRF and are not selected as primary keys, do not transfer these variables on the external file. Specify the format of the key variables and value ranges in advance in order to incorporate them in range-checking programs.

When any of the efficacy parameters are collected in the external data, particular attention should be paid to safeguard the blind. For example, bone density indicators in an osteoporosis trial may be collected with a study's lab data and could be blinded to the

physicians and clinical personnel at the sponsor's site. In case of full double or full triple blind trial, these data must only be disclosed to parties not directly involved in the trial or data safety monitoring committee. A written procedure must exist describing how this data will be handled and to whom it can be disclosed before the clinical database lock. In a similar scenario, subjects may be excluded from the efficacy analysis for loss of baseline data if any of the pre-treatment blind results are incidentally revealed to personnel directly involved in handling the subject.

Data Editing and Verification Procedures

For quality and timely processing of data, errors must be eliminated at the source or as close to the source as possible. To facilitate this, sponsors and vendors must work together to develop editing and verification procedures. These should include:

- Provisions for treatment of partial data;
- Checking for duplicate demographic details and results (real or near real time where possible);
- Range of subject numbers allocated for the study or investigator or both;
- Range of or treatment codes allocated per study or investigator or both;

Sponsor and vendor should identify key individuals for communication and follow-though. A representative from clinical data management should be included. It is recommended that the sponsor provide a range of Subject and Treatment codes for each protocol before external data are received for integration. The allocated ranges should be included in data validation routines and any discrepancies handled as part of a formal discrepancy management mechanism. Very often, a centralized vendor (ECG, Laboratory) with quick results turnaround time will be able to identify and resolve data discrepancies before any other clinical information is entered into the database or even reviewed.

The vendor should perform duplicate record checks as each subject visit data is received. Duplicates should be resolved following a formal data clarification process with the investigative site.

Whenever possible, the sponsor should provide the vendor with a complete listing of subjects' demographic details or IVRS demographic data for an independent reconciliation of sponsor and remote database during the study conduct or before database lock.

Vendor and sponsor should agree upon procedures for assuring that the sponsor receives complete data. If partial records are included in a data delivery, they should be indicated as such. Vendor should provide procedural verification and assurance that a hard copy of the results is identical to the electronically transferred results. Any changes to the system or the programs used to create either of the reports must be tested and documented accordingly. If data are transformed during processing, it should always be possible to compare the original data and observations with the processed data.

If applicable, the vendor should provide a complete list of reference values and their effective dates at the onset of the study. Procedures to minimize the chance of changes during the course of the study must be implemented.

Definition and details of the process for resolution of discrepancies between external and CRF data should be established as part of the study set up. The process should address the issues of both Sponsor and Vendor or third party participant.

Record Formatting and File Formats

Quality and efficient integration of data demands up-front consensus between the sponsor and vendor with respect to record and file format. Areas for initial discussion include the size of data fields, clarification of numeric versus character fields, decimal granularity, use of characters such as ">" and "<", quotation marks, commas and other special characters. Special consideration should be paid to handling of null or missing data.

Depending upon the characteristics of the Database Management Systems and expertise at the sponsor and vendor sites, there may be a wide variety of acceptable record, field and file formats. Thus, both parties must negotiate in writing a mutually acceptable and detailed record format structure.

Areas to consider include the following:

- The sponsor should provide in writing a complete list of reportable variables in the order required. If data is requested in a SAS® dataset, provide the output of the CONTENTS procedure as part of the specification. For ASCII files, specify the column positions or delimiter, record heading as well as field justification.
- Character and numeric fields should be differentiated. Field formats should be specified, in advance, as numeric or character. It is recommended to minimize the reporting of results that can be either character or numeric.
- Sponsor requirements on date and time reporting should be negotiated and specified in writing; examples include DATE9., YYYYMMMDD or TIME5., HH:MM (24 hr).
- Procedures should explicitly describe the handling of greater than (>) or less than (<) signs. Absolute values should be used where possible or to separate the numeric and character portion of the observation into two fields.
- If applicable, comments regarding the condition of the sample or its non-availability should be reported in a field separate from the results.
- The test data in the agreed upon format should be available in a file to be used during database set-up and validation at the receiving Sponsor or designee. Successful generation, transmittal, receipt, loading and screening of the test data validate the data transmittal process.

Data Management professionals should evaluate and leverage on the experience of some of the existing and emerging vendor independent standards for data interchange between clinical systems, namely HL7[5], ACDM's Standards for Electronic Transfer of Laboratory Data[6] and CDISC[2].

Data Transmission
Problems encountered with transmission of data from vendor to sponsor will result in data being lost or loaded incorrectly. In all cases, complete naming conventions and labeling information must be established to facilitate the transmission process. Any data transferred between the vendor and sponsor must contain sufficient information to be uniquely linked to the source of the data and corresponding project and protocol. Origin, date created, date sent, number of records and a version-controlled file naming convention should be followed.

Public encryption mechanisms such as PGP® (Pretty Good Privacy®) are recommended for use when transferring data via the Internet. Thus, the data transfer process will assure compliance with the regulatory guidelines and provide authenticity and confidentiality protection. Not all countries allow the use of strong encryption software. In such cases, consider the use of password-protected files such as ZIP archives or dial up FTP transfer. Both processes will verify the integrity of the file being transferred and provide feedback in case of file corruption.

Procedures for Database Updates (also refer to GCDMP Data Processing)
The processes by which updates to subjects' records are made are among the most vulnerable for generation of errors. Special consideration should be paid if the edit affects any of the primary key variables, and thus propagates multiple records.

Errors generated by the data cleaning process in the sponsor's database should be communicated back to the vendor for follow up and resolution through a formal data clarification process. To update a record in those instances where the original records were either incomplete or contained erroneous data, the vendor frequently will send a second transmission. Updates can be sent either as a full or partial transmission depending upon the capabilities of the systems in place. It is essential that the vendor and sponsor establish procedures that define how retransmissions are identified and handled throughout the study.

Areas to consider include the following:
- During study set up, provide the vendor with a list of in-house data checks and supporting documentation and sample subject number allocations.
- Use correction flags. Where possible, two separate types of flags should be used to distinguish an initial record from a correction or addition.
- Corrections to key variables should be identified and flagged. Updates to key variables should be sent as full records (key variables as well as result variables) and should be flagged at a record level.
- Only current results should be reported.
- The source systems should be designed to permit data changes in such a way that data changes are documented and that there is no deletion of entered data; i.e., maintain an audit trail[7].

- If applicable, vendors should provide the investigator site and sponsor with updated hard copy information, in addition to electronic updates.

File Storage and Archiving (also refer to GCDMP Data Storage and GCDMP Database Closure)

Ultimate responsibility for the quality and integrity of the trial data always resides with the sponsor[8]. Thus, the sponsor should specify in the contract a definitive time period beyond the initial transmission of information during which the records will be maintained by the vendor and be available and accessible to the sponsor and regulatory agencies. It is desirable that vendors maintain active copies of data files during the study stages that require unconstrained accessibility, after which time the vendor should maintain an archived version for the remainder of the retention period. When all reports have been finalized, and sponsor's database locked, a study should no longer require access to the records except for auditing purposes during the record retention period.

For additional information refer to FDA Guidance for Industry, Computer Systems Used in Clinical Trials[3].

Recommended Related Standard Operating Procedures (SOPs) Titles

SOPs should be established for, but not limited to the following:

Sponsor (CRO)	**External Data Provider (Vendor)**
External Data Loading and Validation	Data Extraction and Validation
Query Generation and Vendor (remote) Database updates	Data Transfer and Discrepancy Handling
Vendor Auditing	Database updates
Database lock procedures	Database Archiving and Security
Study-specific procedures (including the handling of extra/unscheduled data)	Study-specific procedures (including the handling of extra/unscheduled data)

References

1. Food and Drug Administration. *Guidance for Industry: Providing Regulatory Submissions in Electronic Format, General Considerations (IT2) and NDAs (IT3).* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration.
2. *Submission Metadata Model*. Version 1.1. Clinical Data Interchange Standards Consortium. April 30, 2000.
3. Food and Drug Administration. *Guidance for Industry, Computerized Systems Used in Clinical Trials,* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; April 1999: Chapter VI, C and D.
4. Association for Clinical Data Management / Statisticians in the Pharmaceutical Industry. *Computer System Validation in Clinical Research, A Practical Guide*.
5. HL7 Working Group. *An Application Protocol for Electronic Data Exchange in Healthcare Environments.* Health Level Seven, Inc.

6. Association for Clinical Data Management (ACDM). *Standards for Electronic Transfer of Laboratory Data.*
7. *Code of Federal Regulations,* Title 21, Volume 1, Part 11, *Electronic Records; Electronic Signatures*. Food and Drug Administration. US Government Printing Office, 1998.
8. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6: Consolidated Guidance,* Section 5.2; April 1996.

Other Related Reading

Bornstein, S. CDM Task List. *Data Basics.* Fall 1999

Bloom JC. Drug Information Association (DIA). Clinical Trial Symposium, January 12-13, 1998.

1. Pharmaceutical Manufacturers Association, Clinical Data Management Group Task Force. Electronic Exchange of Laboratory Data: Recommendations for Improving Quality and Reducing the Hidden Costs. *Drug Information Journal*. 1993; 27:933-942.

Peairs D, Ginsberg A, Gillibrand D, Malcolm R. New Technologies Enhance Central Laboratory Data Collection: Transfer, and Analysis. *Applied Clinical Trials*. 1996; 5

Goodman DB. Standard for ECGs in Clinical Trials. *Drug Information Journal* 1996; 30:237-242.

Dictionary Management

Introduction

A variety of dictionaries may be used in the processing, analyzing and reporting of data collected in clinical trials. The dictionaries range in size and complexity, from simple code lists with a few entries, to large and complex dictionary systems such as MedDRA (Medical Dictionary for Regulatory Activities) and WHODRUG (World Health Organization Drug Dictionary) with thousands of entries and related tables. Some dictionaries are relatively stable and have been used for years, while others are more recent and may be updated regularly. Establishing and maintaining coding dictionaries is an important role that the data manager takes in the tasks involved in coding for clinical trials.

The transition to a different or complex coding dictionary presents multiple challenges to data management. Challenges such as a lack of familiarity with the content, organization, use and capabilities of the new dictionary must be addressed. Processes for managing the release of multiple versions of the same dictionary over a short period of time, handling different dictionaries or versions that have been used and integrating data coded with different dictionaries or versions must be established.

Scope

This section focuses on the management of the dictionaries used for the coding of adverse events, medications, medical history, etc. It does not cover the actual process of coding, which is specifically addressed in other sections of this document (Safety Data Management and Reporting), as well as in the ICH-Endorsed Guide for MedDRA Users[1].

Minimum Standards

- Select appropriate dictionaries that meet project requirements.
- Install and securely maintain dictionaries.
- Implement an audit trail for all changes to the coding dictionary.
- Identify dictionary and version in clinical study reports and integrated summaries.
- Store all versions of dictionaries for future reference.

Best Practices

- Select an auto-encoder to facilitate the consistent use of dictionaries.
- Include version of dictionary in metadata.
- Store data from all coded dictionary levels whether or not all are initially analyzed or reported.
- Establish process for evaluating a change in a dictionary or version.
- Allow search for coded terms to evaluate possible effect of a version change.
- Use the same version of a dictionary for coding in combined (integrated) studies.
- Educate other individuals involved in the analyzing and reporting coded data on the functionality and capabilities of the coding dictionaries used.

- Modify published coding dictionaries only through the appropriate process of submitting changes to the organizations maintaining the dictionaries.

Dictionary Selection

One of the most critical factors in the successful use of a coding dictionary is its selection. The needs of all users that process, analyze and report the coded data should be carefully evaluated. In addition, the regulatory agencies that will eventually review the information must be considered in this decision as well.

One common area for the use of a dictionary is in the coding of adverse events. Recognizing that pharmaceutical and biotech companies are increasingly conducting global trials and submitting marketing applications to multiple regulatory agencies, an initiative to establish a global dictionary was undertaken by the International Conference on Harmonisation (ICH), resulting in MedDRA (Medical Dictionary for Regulatory Activities[2].) It is especially useful in the electronic transmission of adverse event reporting and coding of clinical trial data. The Food and Drug Administration is currently using MedDRA in its Adverse Events Reporting Systems (AERS)[3]. It is planned that MedDRA will eventually replace the multiple coding dictionaries in use, such as COSTART and WHO-ART, as well as proprietary variations of those dictionaries that were developed in-house by many of the sponsor companies.

Medications used by study participants prior to or concurrently with study materials are also commonly coded in order to facilitate the reporting and analyzing of possible drug interactions. A variety of dictionaries or medication references are available that provide information about prescription, generic, OTC (over-the-counter) medications or herbal supplements. The references used for coding in a particular study should be selected based upon the scope of medications included, how current the reference is, the frequency of updates to include the release of new medications, and the coding information available in the dictionary. Such coding information may include generic terms, active ingredients, indications, or ATC (Anatomical Therapeutic Chemical) classification. WHODRUG[4] is one of the dictionaries that is commonly used for coding medications.

Other CRF data may be coded using established medical or clinical references (e.g., medical history body classifications, medication route of administration, adverse event severity and relationships) or in-house determined code lists (e.g., gender, action taken). These same code lists should be utilized across similar studies and projects to ensure the consistent analysis and reporting of the coded data.

Auto-encoder

An auto-encoder is useful when coding adverse events and medications utilizing dictionaries with a large number of entries. The auto-encoder can facilitate consistent coding, without having to rely on the manual reevaluation of previously coded terms. The coding algorithms should be evaluated, such as its ability to handle synonyms, misspellings, capitalization, word variations, word order, exclusion of irrelevant verbatim text, etc. Some clinical database systems include an auto-encoder and the ability to load

electronic versions of coding dictionaries. Other auto-encoders may be available as separate systems. Both integrated and stand-alone auto-encoding systems must be fully validated according to current standards. Other features to be considered when selecting an auto-encoder include ease of use, security features and, ideally, the ability to load multiple dictionaries or versions.

Dictionary Change Control

The selected dictionary and version should be documented, so that it can be referenced in clinical study reports or integrated summaries that report on the coded terms. The coding dictionaries should be securely maintained. If modifications are made in-house to a published reference, it should be clearly stated, so as not to mislead reviewers, who are familiar with the published reference. Any changes made to dictionary entries should be included in an audit trail. This practice of modifying published dictionaries is clearly discouraged by the ICH[1]. The organizations that maintain dictionaries usually have an established process for submitting change requests if for, example an adverse event term or medication is reported that is not included the dictionary. This process allows for a review of the requested change and the dissemination of the request to others using the dictionary.

Dictionary Version Control

Coding dictionaries may be available in an electronic or printed format and multiple versions may be released or published. The dictionary and version used for a given project, time period or sets of data should be clearly documented. When there are multiple ongoing studies, a determination should be made of which dictionary and version will be used for the coding for each study and a systematic process and instructions should be in place to ensure the use of the appropriate dictionary and version. The evaluation of the extent of the changes between versions, their impact on previously coded terms and the criteria for implementing the latest version and recoding should be established[5].

The need for version control and documentation is increased by the practice of using different dictionaries or versions over time. For example, various versions may be used for the coding of post-market safety data versus the clinical data, between different studies for the same drug or even within long-term studies. The impact is greater for adverse events, since a term may be deactivated or reassigned to a more appropriate term, rendering the earlier term assignment no longer current. Most of the changes to medication dictionaries are simply to include new medications.

The dictionary and version information can be maintained within the clinical database, within the auto-encoder as the dictionary files are loaded, or within the metadata of data sets containing coded data. As version information is incorporated into the electronic files by the organizations maintaining the published dictionaries[6], that information may be available without the need for additional in-house action.

Reporting of Coded Terms

An important factor in managing the dictionaries and the versions is the reporting of the coded data. All levels, codes or group assignments for the data should be stored. The risk of not storing all coding information is the necessity to recode or retrieve the additional coding information for reports at other levels or groups that are needed at a latter date.

Changes to the dictionary and version may occur during the course of a project across individual studies. Ideally, the integrated data that is combined should be coded with a single dictionary version. However, an impact analysis between versions may help to identify the necessity for recoding data for integrated studies. If deemed necessary to recode data, it would only be required for those studies that were coded with a different version. Recoding data minimizes the resources and time needed to prepare the integrated data, while allowing for the utilization of the latest version for analysis. In most cases, the evaluation of safety is performed at the integrated level, rather than at the individual study report level. The use of pooled estimates across multiple studies aids in identifying rare events and determining incidence rates[3]. If uniform coding is not employed, the use of different dictionaries or versions among the individual study reports and integrated summaries should be documented.

MedDRA

Since MedDRA was initially published, updates have been released on a quarterly basis. Following the release of version 4.0 in June 2001, the frequency of the updates was reduced to semi-annually. Since the initial release, significant reviews and modifications have taken place. These reviews have addressed assignments to System Organ Class, consistent use of terminology, retirement of terms from current status and the addition of new terms identified during the implementation and actual use of the dictionary in clinical studies. The organization responsible for publishing and maintaining MedDRA is MSSO (Maintenance and Support Services Organization). MSSO recognizes the need to stabilize the MedDRA terminology in order to facilitate the adoption of a new dictionary. Concern that overwhelming amounts of resources are needed to maintain frequent version updates, required frequent recoding or reanalysis of adverse events[7] is also a recognized.

The ability to search for adverse events that have been coded and the dictionary entries at each level within the various versions facilitates the review of the impact of the change and whether an improved coded term is available in the new version. It is possible that in an update to a given version, none of the coded terms were modified or no new terms were added under its assigned HLT.

The general practice has been to code an adverse event to a single body system. In evaluating the safety of a product, the regulatory agencies seek to characterize the safety profile and determine if there is adequate evidence of the effects on the body. If an adverse event is symptomatic in multiple systems (e.g., migraine), the question arises as to what effect it has on each system. MedDRA offers a solution to this problem with its multi-axiality, but the coding practices must be adapted to take advantage of the potential

benefits and the coding process should be clearly defined and documented. This new functionality should be reviewed with other individuals involved in analyzing and reporting of adverse events.

Coding and reporting practices using MedDRA will be addressed in future versions of the GCDMP, at a time when newer practices and techniques are more widespread.

Recommended Standard Operating Procedures

- Selection and Maintenance of Coding Dictionaries
- Security, Change and Version Control

References

1. International Conference on Harmonisation, April, 2002, *MedDRA® Term Selection: Points to Consider, Release 3.0*
2. Available from the MedDRA maintenance and Support Services Organization (MSSO), http://www.meddramsso.com/
3. Food and Drug Administration, November 1996, *Reviewer Guidance Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review*
4. Available from Uppsala Monitoring Centre, http://www.who-umc.org/
5. Maintenance and Support Services Organization (MSSO), July 2000, *Recommendations for MedDRA™ Versioning for Summary Reporting*
6. Maintenance and Support Services Organization (MSSO), January 2001, *Messenger Quarterly Newsletter*
7. Maintenance and Support Services Organization (MSSO), May 2002, *Messenger Quarterly Newsletter*

Other Related Reading

N/A

Safety Data Management and Reporting

Introduction

Safety data often house the most challenging aspects of the management and reporting of clinical trial data. Consideration for return-on-investment often results in curtailment of efforts for the query process for scrubbing safety data and limits the methods for reporting the data.

Estimating resource requirements and balancing business value against scientific theory are critical in the planning of work effort. Scientific principles also motivate clinical trial scientists to use judgment in determining the standards to be set for a given study or program, the quality markers to be used, the levels of precision, and the depths of analysis and reporting. When data are stored and scrubbed as if there is a high degree of precision and reliability, but when, in fact, the information has a soft basis, reports can reflect an over-reliance on questionable data, which can result in inferential errors. Soft information can still be quite useful, but to avoid misrepresentation, there needs to be clear identification of the nature of the data themselves.

The quality of data is really determined in the field. If the quality of the information as recorded in source documents is poor, there is little that data managers or statisticians can do to repair it. Instead, it behooves the data managers to ensure that the database accurately conveys to users the limitations of the quality. There is an imperative for the statistician to ensure that analyses and data displays acknowledge the limitations.

The processes of data capture, management, and reporting are highly integrated. Considerations for best practices would be deficient to consider reporting guidelines in the absence of guidelines for the earlier processes.

Scope

The safety data in a clinical study are simultaneously a rich source of information and an enormous challenge to the clinical trial scientist. The data manager and statistician who are a part of the product team must work closely with each other and with the other team members to ensure that safety data are captured in a sensible way to facilitate proper interpretation and meaningful analysis and summary. Ensuring quality requires that the team capture, process, and report the data in a way to facilitate drawing reliable conclusions. The business and scientific balance to be considered by the data managers and statisticians is that resources may be wasted and diversionary when expended on efforts that have no effect on conclusions.

Safety data may be displayed and reported in prohibitively many ways. Judgment and scientific selection are needed to identify the trends and salient features of the data to ensure adequate reporting of results that pertain to product effects. Producing voluminous pages that are incomprehensible and clinically meaningless can dilute the effects that are real, the discernment of which is the driving goal of the safety data processing and reporting.

This section proposes practices, procedures, and recommendations for data managers to operate within the project team working closely with statisticians, monitors, and clinical research to ensure that data management practices support statistical and medical purposes. Data managers are better equipped to function as fully-integrated team members when there is a basic understanding of the activities and needs of other team members, particularly statisticians.

Minimum Standards

When considering the capture, management, analysis, and reporting of safety data, the following minimum standards are recommended:

- Ensure compliance with regulations.
- Ensure that the standard of quality supports the utilization of the data.
- Ensure that conclusions about the safety profile of a compound can be reliably drawn from the database.
- Ensure that safety risks are identified and reported accurately.
- Ensure that normal ranges are properly linked to laboratory data. If normal ranges are unavailable, ensure that the reference ranges used are documented as such. This is especially crucial when normal ranges are updated frequently.

Best Practices

When considering the capture, management, analysis, and reporting of safety data, the following best practices are recommended:

- Develop CRFs with teams of individuals from monitoring, data management, statistics, regulatory affairs, and medical, ensuring adequate attention to the collection of safety data.
- Consider the level of precision that can be attained in the study and select the CRF format for collecting AEs appropriate for that level. Also consider the level of precision in the analysis.
- Define severity, understanding its uses and limitations.
- Examine laboratory data from the perspectives of categorical shifts, changes in magnitude for the group, individual significant values or changes, and listings. Consider related parameters for compounds with potential toxicity in specific body systems.
- Consider laboratory normalization techniques when combining data across studies or centers where varying normal ranges are used.
- Include data managers and statisticians working together when considering computerization, management, reporting, and analysis of safety data. These are highly integrated and require joint considerations of individual team constituents. Develop standard operating procedures (SOPs) for data capture, data validation, statistical analysis, and reporting of data that include guidelines for this team approach.
- Document the status and quality of safety data and include this documentation with the database.
- Include clear links for comparators, such as normal ranges for laboratory data, with the database.

- Consider levels of precision in the capture and the reporting of safety data, to reduce the likelihood of over-interpretation or misinterpretation.
- Time-to-event analyses are only meaningful when the timing of the event is reliably known.
- Consider both categorical shifts (from status of normal to abnormal) and magnitude changes for analysis and reporting of laboratory data. Examination of significant values may give different information from examination of significant changes.
- Apply standards commensurate with the utilization of the results residing in the databases when using databases for safety reporting (expedited reporting, ongoing review by monitoring boards, routine reporting, etc.) If important decisions rely on the information, know the level of quality and the appropriateness of the data.

Available Guidelines

One definition of "quality data" is "a collection of data from which reliable conclusions can be drawn." The goal of reporting safety data is to convey information that would facilitate reliable conclusions to be drawn. In investigative clinical research trials, one of the key objectives generally is to investigate, characterize, establish, or confirm the safety profile of an investigational product. The management and reporting of the safety data from the trial need to support that objective.

The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has issued several guidelines to provide guidance to industry for how to manage and report clinical trial safety data.

E1A describes expectations for extent of population exposure for drugs intended for long-term treatment of non-life-threatening conditions. The guideline acknowledges that safety evaluation during clinical drug development is not expected to characterize rare adverse events (AEs), for example, those occurring in less than 1 in 1000 subjects. Total short-term exposure is expected to be about 1500 subjects; exposure for six months by 300 to 600 subjects should be adequate; exposure for a minimum of one year by 100 subjects should be adequate. Exceptions are noted.

E2A, E2B, and E2C are Clinical Safety Data Management guidelines and provide guidance for definitions and standards for expedited reporting, data elements for transmission of individual case safety reports, and for periodic safety update reports for marketed drugs.

E3 is the guideline on "Structure and content of clinical Study Reports." This guideline provides detailed recommendations and specific suggestions for data displays of safety data. It is noted that the guideline shows "demography" as a subsection of "efficacy evaluation" and "extent of exposure" as a subsection of "safety evaluation." FDA regulations require (for studies for which it makes sense and for integrated summaries) that efficacy and safety data be analyzed with particular consideration for age, sex, and race. ICH guidance encourages that the analysis of both efficacy and safety data consider extent of exposure (including compliance). It is imperative to understand that both demography and dose exposure

relate to both efficacy and safety. It follows that the analysis and reporting of safety data should consider the characteristics of the presenting population and the extent of exposure to the investigational compound.

E5, "Ethnic Factors in the Acceptability of Foreign Clinical Data," advises that there are concerns "…that ethnic differences may affect the medication's safety, efficacy, dosage, and dose regimen…" This guideline also delineates between extrinsic ethnic factors (those associated with environment and culture, e.g., diet, use of tobacco, use of alcohol) and intrinsic ethnic factors (those that help define and identify a subpopulation, e.g., age, sex, weight, organ dysfunction).

E6 is the consolidated good clinical practice (GCP) guideline. This guideline contains principles of GCP that underscore the scientific basis of the clinical trial and specify qualifications for the personnel and systems involved in all aspects of the clinical trial, that adherence to good scientific principles is required, and that the documentation of the adherence is needed.

E9 is a guideline geared toward the statistician, which includes substantial advice for the analysis of safety data.

Other guidance documents are available from the ICH and from regulatory agencies that give advice for capturing, managing, and reporting safety data. A sponsor should refer to IND regulations (21CFR312) and NDA regulations (21CFR314) to ensure compliance with FDA regulations for investigational and marketed products.

Safety Reporting

Safety data are reported and examined at various stages of an investigation and by different assessors. IND regulations specify expedited reporting for serious or alarming adverse events; many studies have safety data monitoring boards (SDMB) that review data as they accumulate in a study; the sponsor's medical monitor reviews safety data, frequently masked to the treatment. Then, after market approval, there are NDA regulations that specify safety reporting. Data managers and statisticians need to ensure that the reports provided are supported by the quality appropriate for the purpose of the report.

FDA requires that sponsors provide them with what they need to meet their obligations to Congress and to the public. Complying with IND and NDA regulations is aided by understanding their mission and motivation. Before marketing, IND regulations apply.

One key purpose of the IND regulations is to facilitate FDA's monitoring of the investigation, including protection of the safety and rights of individuals enrolled into trials and the scientific quality of the investigation in terms of its ability to adequately demonstrate the efficacy of the compound. FDA requires annual reports, which are brief reports of the progress of the investigation, including any newly identified safety trends or risks that may impact the investigation. FDA also requires expedited reports of "any adverse experience associated with the use of the drug that is both serious and

unexpected"(21CFR312.32). Written notification of such events is required within 15 calendar days. For such events that are fatal or life threatening, a telephone or facsimile transmission is required within 7 calendar days. Additional details of IND safety reports, annual reports, and IND specifications are provided in 21CFR312.

After marketing, FDA has a different perspective and a different goal. Once the compound is in medicine cabinets, it is much more difficult (if not impossible) for FDA to get back, if a recall is necessary. The regulations provided in 21CFR314 describe reporting requirements after approval: periodic reports are required quarterly for three years after approval, and annually thereafter. In addition, under NDA regulations, each adverse experience that is both serious and unexpected, whether foreign or domestic, must be reported within 15 calendar days. Additional details of NDA safety reporting, periodic reports, annual reports, and NDA specifications are provided in 21CFR314.

In addition to FDA's monitoring of investigations and review of safety data, the FDA requires sponsors to have medical monitors who are expected to review safety data. Sponsors frequently have safety data monitoring boards, comprised of individuals separate from the conduct of the study that may conduct interim analyses and review accumulating data, blinded or unblinded. Data monitoring committees often make recommendations or decisions to halt an ongoing investigation, usually due to (1) overwhelming efficacy, (2) unacceptable safety risk, or (3) futility. These boards may also make recommendations for changes in the ongoing study, such as a dose reduction, or the elimination of an arm of the study with an unacceptable safety risk.

Any review of safety data that is based on reported information from a safety database (as opposed to CRFs) relies on that database; if the quality is poor, the decisions taken may be wrong. Review of accumulating data often implies there are a mixture of complete data and partial data, and a mixture of clean and dirty data. To provide the optimal information to the users of the dynamic database, it is recommended that the quality be known and reported to the reviewers with the safety data. It is generally not helpful to report to data reviewers that some data are dirty, but to not identify which data are dirty.

Capture, Management, And Reporting Of Adverse Events

Clinical adverse events frequently house the most important safety information in a clinical study. Ensuring that methods of collection, coding, analysis, and reporting facilitate the drawing of reliable conclusions requires understanding characteristics and limitations of adverse event data.

- **Precision**

  The precision with which AE data are captured relates directly to how the data can be analyzed and reported. There are three basic types of precision in a clinical trial:

  - **High Precision**

    Investigation in a Phase 1 sequestered environment (a phase 1 house) often incorporates high precision, continuous medical monitoring in a study. With a few subjects in a sequestered environment, a nurse or physician is by the bedside

continuously. In such an environment, clock time may be recorded, so that precise data can be collected for onset and offset of an AE. Elapsed time since initiation of treatment and duration of the AE, hence, can be calculated in a meaningful way. Clock time is meaningful in such an environment for some events, although, it may be difficult to assess the precise minute that sleepiness begins or a rash is cleared.

- **Moderate Precision**
  Investigation in a hospital often incorporates moderate precision, daily and frequent, but not continuous, medical monitoring in a study. Hospitalization offers a controlled and sequestered environment such that a nurse or physician can assess the subject daily. In such an environment, clock time may not make sense for all events, but date can be precisely recorded. Onset and offset of an AE can be recorded in terms of days, but not hours. Elapsed days since initiation of treatment and duration (in days) of the AE can be calculated.

- **Low Precision**
  Investigation in an outpatient study where subjects return to the facility after days, weeks, or months incorporates low precision. In such an environment, neither clock time nor date may be meaningful. Use of subject diaries may assist with elapsed time since treatment or duration, but subject diaries are notoriously inaccurate. In such studies, it is recommended to capture frequency (single episode, intermittent, continuous), maximal severity, most-harsh relationship, etc. rather than to attempt to record each event with time of onset and offset.

  When an investigation is of low precision, but attempts have been made to record data as if there were moderate or high precision, the result is generally a database with dates (or times) that are rough guesses that may be far from accurate.

The precision with which AE data were collected has an important impact on how the data can be analyzed in a meaningful way. Dates cannot be interpreted in an outpatient study with the same reliance as in a sequestered study. If dates are present in the database, it may be tempting for the statistician to employ survival-analysis techniques to analyze time-to-event. If these dates are not accurate, the resulting analysis can lead to incorrect or unreliable conclusions.

- **Severity**
  When considering the capture of severity of adverse events, it is tempting to make the assessment in terms of its impact on activities. This method of assessment may be meaningful for some events, such as “pain,” but not meaningful for others, such as “alopecia.” In some cases, severity is not assessable at all. “Mild suicide” is not meaningful. Some events are episodic, not graduated by severity, such as “hair-line fracture.” Assessment of diarrhea as “severe” is often made because of duration or frequency of episodes, which are different parameters; diarrhea is episodic.

The concept of severity is only meaningful within a particular event. When one considers severity of AEs for an organ class (e.g., CNS), ranking among mild, moderate, and severe is not meaningful. If one considers "mild stroke" and "severe flush" (both CNS events) the ranking is not sensible, as opposed to the ranking of "mild headache" and "severe headache," where the relative ranking does make sense.

A common data display (encouraged by the ICH and the FDA) is a breakdown by severity. It is easy, in this context, to confuse severity with seriousness, or to misinterpret severity altogether. A breakdown that ignores the particular events and counts AEs that are mild separately from AEs that are moderate will give a distorted assessment when there are reports of "mild stroke" or "mild MI" and also "severe rash" or "severe sleepiness" in the same study. A more meaningful display breaks down severity within a particular event.

- **Dictionaries**

AE dictionaries are needed in order to group data for meaningful analysis. MedDRA is the ICH-developed and recommended dictionary for all medical events captured in clinical trials, including, but not limited to, AEs.

Utilization of MedDRA requires an understanding of the levels of terms and an understanding of the multi-axial functionality. The levels of terms are the following:

- Lowest level term (LLT)
- Preferred term (PT)
- High level term (HLT)
- High level group term (HLGT)
- System Organ Class (SOC)

It is noted that the SOC level within MedDRA is really a dual level, in that, MedDRA permits a primary SOC and one or several secondary SOCs.

The multi-axiality of MedDRA permits a single AE to be simultaneously coded to many SOCs. For example, a migraine headache could be coded to the nervous system (because of the involvement in the brain), the vascular system (because it is a vascular disorder), the GI system (if there is associated nausea and vomiting), eye disorders (if there are visual disturbances), etc.

MedDRA is not just another dictionary; it is another way of thinking about medical information. It is imperative for managers of medical information to understand the flexibility of MedDRA as well the implications on reporting of its storage and implementation.

- **Version Control**
  Updated versions of dictionaries frequently change pathways to body systems or organ classes. Such changes in a dictionary can have a substantial effect on the conclusions of the effects of a product on the body. The version of a dictionary used for classification of AEs into body systems can, thus, impact the labeling of the product. The data manager involved with the implementation of a dictionary for a study (or product) must attend to ensuring consistency, where possible, and to ensuring the ability to replicate, as there must be a clear trail leading from the data to the labeling.

  Most standard dictionaries that have been widely used have been reasonably stable (COSTART, WHO, ICD-series, etc.) MedDRA is updated periodically. Management of the version requires more resources when the updates are more frequent.

  One suggested practice for ensuring consistency within a long-term study is to execute the dictionary against the AE data as the study progresses (for purposes of medical monitoring, interim analysis for safety review boards, or other important purposes), but prior to data lock, re-execute the dictionary using the current version of the standard dictionary, so that the entire study is executed against the same version of the dictionary, and that version is as current as reasonable. Additional queries may result at the time of re-execution (that is the reason this execution should be done prior to database lock.)

  It is recommended that the version of the dictionary used in any study be stored with the database, to ensure reproducibility.

- **Encoding**
  Auto-encoding is highly recommended to executing a dictionary against AEs. There are auto-encoding programs available to assist with the programming aspect. Training of the monitors and site personnel to cultivate an understanding of the coding process should facilitate capture of AE data in a format that can be auto-encoded. Training should include guidelines such as the following:
    - Avoid use of adjectives as initial words (e.g., “weeping wound” may be coded to “crying” or “faint rash” may be coded to “syncope”)
    - Avoid use of symbols and abbreviations in the AE text field (symbols may be interpreted by different individuals differently)
    - Avoid inclusion of severity (e.g., “severe headache” in the AE text box inhibits auto-encoding; severity should be recorded in the severity box, not the AE text box)
    - Ensure AE text has clinical meaning (e.g., “bouncing off the walls” and “feeling weird” are difficult to interpret)
    - Ensure AE text has clear meaning (e.g., “cold feeling” may be interpreted as “chills” or “flu symptoms”).

Encoding within the database may add unnecessary complexity to the management of the database in cases where final coding may require judgment. After database lock, if medical judgment indicates that the default pathway inaccurately captures the medical condition, a change would require unlocking the database if the auto-encoding is done within the database proper. Reflecting the auto-encoding in a separate file (e.g., an AE analysis file) offers the possibility of reflecting changes in medical judgment after database lock, if deemed essential. This practice does impose the need for an audit trail on analysis files.

- **Hard-coding**
  Hard-coding or coding outside the clinical database, is generally a dangerous practice. For coding AEs, hard-coding is sometimes used to introduce medical judgment that the standard dictionary does not offer. When events such as "strange feeling" are reported and no additional information from the site is available, the medical monitor for the study may have insight that assists with the codification of the event, which can be inserted into the AE analysis file through hard-coding. It is possible to use "pass-through" text for the AE preferred term (conventionally, many sponsors make use of quotation marks to indicate verbatim text that is passed through by a program to the preferred term field) and hard-code the body system.

  Any use of hard-coding requires careful documentation.

- **Lumping and Splitting**
  Among coders, there are lumpers and splitters, and there is no universally agreed method of handling AE text that includes more than one event. "Tingling in hands and arms" is regarded by some as a single event and by others as two events. There are consequences to decisions to lump or split.

  When two events are reported in the same text field (e.g., "indigestion and diarrhea"), if splitting is done by the data management staff rather than the site, inconsistencies within the database may result. Often, when the data manager splits the AE text into two (or more) events, the associated items are duplicated (or replicated). If, for example, there is medication that is given for treatment of the AE, and the concomitant medications page of the CRF shows as the reason for use one event, (e.g., indigestion), the splitting induces an AE with treatment given for which no treatment is recorded.

  Medical judgment may also be introduced into the database by the data manager, inadvertently. If the severity of the compound event is recorded as "severe," then the duplication of the attributes of the AE impute "severe" to the other event(s), which may not be the physician's judgment for that particularly component of the AE.

Coding of AEs has significant impact on the analysis and interpretation of the safety data for a product. The perspective that coding is a clerical function is naïve and risky. As the

world moves toward the full implementation of MedDRA, the role of coding will have an even greater impact on the interpretation of safety data.

Capture, Management, and Reporting of Laboratory Data

Laboratory data differ from most other data importantly in characteristics. Most clinical adverse events can be observed by either the subject or the physician. An elevation in bilirubin or cholesterol is not generally observable. Even in high precision studies, the time of an elevation of a clinical chemistry analyte (for example) is impossible to know. At the time of a blood draw, it can be known whether or not the value is elevated, but when it became elevated is unknown.

The peculiarities of laboratory data need to be respected in the management of the data. Ensuring that the units of the data as they were captured are reflected clearly in the storage of the values requires attention (in many databases, units are separated from the values). In combining data across studies, it becomes particularly challenging and particularly important to ensure proper linkage with the units, in order to protect against unreliable conclusions being drawn from the reported laboratory data.

One of the most challenging aspects of managing laboratory data is linking the data to the appropriate normal range. In the capture of data, if the data do not come to the data manger on electronic medium, attention should be given to ensure the link between each value and the appropriate normal range.

In cases where normal ranges are not available or not obtainable, reference ranges may be used, which may be derived from normal ranges that are available in the study or from a reference book. Documentation of the use of reference ranges in lieu of normal ranges must be clear for users of the database.

Normalization techniques for laboratory data are often employed for convenience of combining data across studies, for example. Normalization techniques generally include a transformation of the data into a unitless value between “0” and “1” when the value is normal, below “0” when the normal is below the lower limit of the normal range, and above “1” when the value is greater than the upper limit of the normal range.

Reporting laboratory data can be prohibitively resource-intensive, if judgment and selection are not a part of the planning for data displays. The ICH and FDA have given specific guidance for how to report laboratory data.

- **Treatment-emergent abnormal values (TEAVs)**
  Comparative data summaries (and supportive listings) are strongly encouraged that provide a summary (1-page) by treatment group (for parallel studies) for analytes included in the study for hematology, clinical chemistry, urinalysis, or other laboratory panel or group. Such a summary provides a valuable overview of movement from the normal state pre-dose to an abnormal state (at any time post-treatment, in either direction, for any analyte).

- **Clinically significant values or changes**
  Comparative data summaries (and supportive listings) are recommended. These provide summaries and details by treatment group of analytes with significant changes or values, such as analytes where the baseline value is doubled or tripled, or where a value is observed in an analyte that is twice the upper limit of the normal range, or where a change exceeds the width of the normal range.

- **Groups means and changes**
  Displays of means and mean changes from baseline levels are useful within a group (to indicate a trend in an analyte) or among groups (to examine treatment group differences or trends that may be dose-related).

- **Shift tables**
  Shift tables frequently are 3x3 tables that show the status before treatment (below normal, normal, above normal) versus the status after treatment (below normal, normal, above normal). These displays ignore magnitude of change. The display depicts the movement (or lack thereof) from one category before treatment to another category after treatment.

- **Individual data displays**
  Listings of individual data are needed for adequate reporting of most clinical trials. When the study is large, individual listings may be voluminous, and reporting needs to consider practical aspects of summarization.

- **Related groups of analytes**
  Summaries by related groups of analytes are useful for some studies or integrated summaries. Products that may be prone to cause liver damage may need careful examination of analytes that relate to hepatic function. It may be useful to prepare a summary on a single page that includes for the hepatic-function-related analytes proportions of subjects who double the baseline, triple the baseline, have a change of fixed magnitude, or exceed an alert or toxic threshold.

    - **Other Data**
      Safety data can have forms other than AEs and laboratory values. Capture of data from specialty tests (e.g., electrocardiograms, electroencephalographs) requires understanding the common data derived from the test, the format, precision, and special attributes.

      Physical examinations are customary in clinical trials. In a broad sense, the physical exam is a screening method; if an unexpected, significant abnormality is detected during a physical exam, there is generally a specialty test to confirm the event. The data with greater reliability, then, is the specialty test.

      In considering data capture, free-text commentary boxes are generally discouraged. If they are used for medical monitoring purposes, they can be shaded so that the reviewing medical monitor can have the prose handy, but the text need

not be computerized, necessarily. Making effective use of the investigator's comment log can ensure that essential text (which is generally minimal) is computerized, if warranted.

The management of "Other Data" depends on the form of that information. For physical examinations or specialty tests for which free-text commentary is permitted, there are methods of managing the commentary without compromising quality standards of the database.

Free-text commentary can be computerized using word-processing rather than a data entry system, and then it can be proofread, rather than double-keyed. Using this method permits the free-text commentary to be computerized and linked to the database without being a part of the database proper, so that quality standards can be maintained for the database proper, but reasonable standards may apply to free-text prose.

One method used by some sponsors that avoids computerization of verbose commentary is codification, in which a medically qualified individual reads the information and judges it to be relevant or not relevant, or perhaps critical. A code can be applied and keyed, where "0=no comment," "1=comment, not relevant," "3=comment, relevant," and "4=comment, critical."

- **SAE Data**

  Expedited reports are required by regulatory agencies for certain serious adverse events. In many companies, receiving reports of SAEs, computerizing these reports, and managing these reports is the responsibility of a dedicated group of individuals. Often, this group is separate from the data management group that is responsible for computerizing and managing data reported from clinical trials.

  The SAE database often includes safety data from various sources. Reports can be received from patients in clinical trials, from spouses who took trial medication (accidentally) and had AEs, or from patients who took marketed drugs and who are not participating in any trial. These reports can come from individuals who give reports over the telephone to a sponsor, from employees who report to the sponsor that they were told about adverse reactions to marketed products, from physicians, from the literature, or even from regulatory agencies. These reports are generally single-keyed, often by individuals other than professional data mangers. These reports generally are not queried. The data within these SAE databases may be dirty, incomplete, duplicate, fragmentary, etc. In contrast, the reports of SAEs from clinical trials that are reported on the AE page of the CRF are subjected to rigorous data management procedures, including scrubbing, querying, and verification to ensure accuracy. These two types of databases generally have important differences in their sources, their quality levels, their uses, and their customers.

Reconciliation of SAE data and the clinical trial database that houses the relevant SAE reports is not always straightforward. Different sponsors have vastly different methods of managing these two databases.

Good clinical data management practices include provision for reconciling important disparities between serious adverse events that are captured both in the SAE database and in the clinical trial database. The business balance perspective encourages users of these databases to recognize that SAE databases may not be queried or updated as clinical trial databases are, and, consequently, there may be some discrepancies because of the reporting of preliminary medical judgments that were later changed with updated information.

General Safety Data

The FDA draft "Reviewer Guidance: Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review" (November 1996) provides specific guidance to industry that reflects thinking within the Agency about safety data.

In the above-referenced document, the FDA described the concept of clinical domains for a review of systems:

- Cardiovascular
- Gastrointestinal
- Hemic and Lymphatic
- Metabolic and endocrine
- Musculoskeletal
- Nervous
- Respiratory
- Dermatological
- Special Senses
- Genitourinary
- Miscellaneous.

In this document, the FDA specifies that an NDA be reviewed against each clinical domain with two key questions as goals:

- Are the safety data adequate to assess the influence of the product on the clinical domain?
- What do the data indicate about the influence of the product on the clinical domain?

An imperative to statisticians who are involved with the reporting of safety data is the need to review safety data to ensure that the influence of the investigational product on each clinical domain is described clearly.

The design of the study must be considered in reporting clinical trial safety data. In a multi-center study, the ICH and the FDA urge examining the influence on the results of center effects, to ensure that the results are not carried by a single center or dominated by a small proportion of the total study.

In a multi-center study, center effects are typical and are a nuisance. There are three sources of contributions to the center effects:

- The investigator as an individual (the bedside manner, personal biases, peculiar methods of assessment)
- The environment (equipment, SOPs, staff, etc.)
- The subject population (those who frequent the hospital or clinic, VA hospital, university hospital, country clinic, etc.)

When the study employs one investigator who may be on the staff of several hospitals, or when a cluster of hospitals shares equipment and has common SOPs, or when a study makes heavy use of referrals, these attributes affect the interpretation of the center effects. Reporting data in a multi-center study requires understanding the source of variability among centers and the reasonableness of displaying data by center or by clusters of centers.

Recommended Standard Operating Procedures

- Coding of adverse events
- Maintenance of coding dictionaries
- Reconciliation of serious AEs in SAE database with clinical trial database
- Management of AE analysis file
- Management of laboratory data and normal ranges
- Preparing integrated summaries of safety data

References

1. *Code of Federal Regulations,* Title 21, Volume 1, Part 312*, IND Regulations*. Food and Drug Administration. US Government Printing Office.
2. *Code of Federal Regulations,* Title 21, Volume 1, Part 314*, NDA Regulations*. Food and Drug Administration. US Government Printing Office.
3. *Code of Federal Regulations,* Title 21, Volume 1, Part 11*, Electronic Signature Regulations*. Food and Drug Administration. US Government Printing Office.
4. International Conference on Harmonisation. E1A: *The Extent of Population Exposure to Assess Clinical Safety*.
5. International Conference on Harmonisation. E2A: *Clinical Safety Data Management: Definitions and Standards for Expedited Reporting* .
6. International Conference on Harmonisation. E2B: *Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports*.
7. International Conference on Harmonisation. E2C: *Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs*.
8. International Conference on Harmonisation. E3: *Structure and Content of clinical Study Reports*.
9. International Conference on Harmonisation. E4: *Dose-Response Information to Support Drug Registration*.

10. International Conference on Harmonisation. E5: *Ethnic Factors in the Acceptability of Foreign clinical Data*.
11. International Conference on Harmonisation. E6: *Good clinical Practice; Consolidated Guideline*.
12. International Conference on Harmonisation. E7: *Studies in Support of Special Populations: Geriatrics*.
13. International Conference on Harmonisation. E8: *General Considerations for Clinical Trials*.
14. International Conference on Harmonisation. E9: *Statistical Principles for Clinical Trials*.
15. International Conference on Harmonisation. E10: *Choice of Control Group in Clinical Trials*.
16. International Conference on Harmonisation. M1: *Standardization of Medical Terminology for Regulatory Purposes*.

Other Related Reading

N/A

Serious Adverse Event Data Reconciliation

Introduction

Serious Adverse Event (SAE) data reconciliation is the comparison of key safety data variables between two databases. Reconciliation is performed to ensure that events residing in any SAE Database and those residing in the Clinical Database are consistent. It is an iterative process that occurs several times during the study. When to reconcile is determined by the frequency of data receipt, the scheduling of safety updates, and the timing of interim and final reports.

Scope

This procedure applies to all projects where both a Data Management Clinical Database and a Drug Safety SAE Database are maintained as two separate databases.

Minimum Standards

- Create project-specific entry and edit instructions including deletion and change control procedures.
- Standardize the capture of SAE data elements in both the clinical database and the safety database.
- Conduct the reconciliation of event terms so that they are at least similar if not exactly the same.

Best Practices

- Establish the time intervals in the project where reconciliation will be performed and in particular the mechanisms to cover interim analyses or safety data reporting. Often SAEs continue to be reported after a clinical trial has concluded. Therefore, it is important to establish a cut-off point after which no SAEs will be added to the Clinical database, even if the Safety database is updated.
- Identify the data items to be reconciled. This may include, but not be limited to the following:
    - Protocol
    - Investigator
    - Subject identification
        - Randomization number
    - Initials
        - Date of Birth
        - Gender
        - Race
    - Event's case number
    - Diagnosis
    - Verbatim
    - Coded or preferred term
    - Onset date
    - Resolution date
    - Date of death
    - Outcome

- Severity
- Causality assessment
- Action taken with study drug

Sometimes other data items are used from other modules for further reconciliation or clarification.

From the Demography module, items used may include but not be limited to the following:

- Subject identification
- Age
- Weight
- Date of birth
- Gender
- Race
-

From the Discontinuation module, items used may include but not be limited to the following:

- Subject identification
- Primary reason for discontinuation being an event
- Cause of Hospitalization
- Cause of Death on the death certificate
- Autopsy Result

From the Concomitant Medications module, items used may include but not be limited to the following:

- Subject identification
- Medication Name
- Start date
- Stop date
- Indication

- Whenever possible, customize the database fields to be used in reconciliation to be programmatically compared without compromising the integrity of the software or the databases. Even programmatic reconciliation of fewer than 100 events can be cost-effective in both time and quality. The process can be validated once and run as frequently as data and time allows.

- When initiating the reconciliation process, the Clinical Data Management Lead should confirm that all data to be included in the reconciliation have been entered and validated; that any data clarifications have been returned and applied to the clinical database; and that the coding of AE verbatim terms against the common dictionary has been completed.

- The Clinical Data Management and Drug Safety Leads should establish a mutually agreeable turn-around time for researching, retrieving, and correcting any discrepancies found during or since the last reconciliation period.

- Read-write access to either database (but not both) is granted to personnel trained in data entry for the purpose of and whose responsibilities include data entry, data modification, or data validation. Read-only access is granted to personnel related to the reconciliation but who are not directly responsible for those tasks related to data modification. System administration rights are limited to personnel responsible for database configuration.

Procedures

- Obtain the SAE information to be reconciled from both the Drug Safety and the Clinical databases.
  - Hard-copy listings from the Safety Database are usually provided to Data Management and the two databases are reconciled through direct comparison of manual listings. However, in some instances the two databases can be compared programmatically and a listing of differences provided. Either way, the differences will require manual review by trained staff. Ancillary documents (e.g., hospitalization discharge summary, death certificate, autopsy report) can also be used for clarification or corroboration.

- Verify that all the SAEs from the Clinical database reside in the Drug Safety database. Note that some SAEs from the Safety database will not be in the Clinical database until all CRFs are collected and entered.

- Document all SAEs that are in the Clinical database but are not in the Safety database. These are potentially unreported events. Include copies of the appropriate CRFs to be forwarded to the Drug Safety contact person.

- Research all SAEs in the Safety Database that are not in the Clinical Database.
  - If the visit has been monitored, collected, and entered by CDM, the site should be queried to request the original missing event page. Do not add SAEs to the Clinical database without the data for that visit having been monitored against source documents according to the study's Clinical Monitoring Guidelines. Only those updates signed and dated by the site staff after the CRF page has been monitored and retrieved are acceptable from Drug Safety for updating the Clinical Database.

- Research and resolve all differences between the SAEs that are present in both databases.

- Depending on the nature of the discrepancies, it may be necessary to seek input from the company's medical staff before deciding on a course of action. Some discrepancies are acceptable. For example, slight variations in terminology used in

describing events are of no consequence; start dates may differ, as an event may start as non-serious before progressing to serious. Site-authorized updates to the CRF received by Clinical Data Management are copied to Drug Safety for assessment and, if appropriate, for inclusion in the Safety Database. Clinical Data Management generates queries to clarify discrepancies and forwards them to the sites for resolution. Resolved queries from the site are returned through Data Management to be used to update either or both databases by their respective trained staff. Communication of these updates can be facilitated by use of a standard template such as the Sample SAE Data Reconciliation Form provided in Appendix B.

- Prior to data lock, verify that all queries have been returned and integrated into the database correctly. A QC process should be in place to ensure that this is done accurately and consistently. Ensure that reconciliation has been performed on all events. Ensure that all expected SAE information has been received. Written notification should be made when reconciliation has been successfully completed to avoid confusion should the Safety Database be held open for updates after the study ends.

- Any final inconsistencies that cannot be resolved should be documented in the CDM Data Handling Report or the equivalent.

Recommended Standard Operating Procedures
- Safety Database Setup, Management, and Validation
- Serious Adverse Event Reconciliation Work Instruction
- Coding of Clinical Data

References
N/A

Other Related Reading
N/A

Measuring Data Quality

Introduction

The purpose of the data quality measurement section is to promote understanding and discussion on industry accepted standards and measures of data quality. Federal regulations and guidelines do not address minimum acceptable data quality levels for clinical trial data. In fact, there is limited published research investigating the distribution or characteristics of clinical trial data errors.

Even less published information exists on methods of quantifying data quality. Therefore, commonly used estimates of data quality often fail to account for likely sources of error, underestimating the true error rate. Due to differences in methodology, estimated error rates are often not comparable from trial to trial, vendor to vendor, auditor to auditor, or sponsor to sponsor. They are thus of little use in the regulatory review process.

It is important that Data Management professionals take a proactive role in setting appropriate standards for acceptable data quality levels, methods for quantifying data quality, and practices to assure data quality. Cleaning all data collected or attempting to create an error free database may not be necessary or practical. It is necessary, however to make reasonable attempts to identify and understand the errors that are present in clinical trial data and to correct errors that will have an impact on study results. The industry currently has all of the information, tools and methods necessary to assess data quality. It is critical and, in fact, responsible to quantify data quality and to evaluate the impact on the validity of the trial.[1]

Scope

The Institute of Medicine (IOM) defines quality data as data that support conclusions and interpretations equivalent to those derived from error-free data[2]. This section gives methods and practical suggestions for using the IOM definition of data quality to answer the question, "Is the data quality acceptable for this trial?" There are many steps to making this definition of data quality operational. These steps involve understanding sources of errors, identifying errors through inspections, using inspection results to measure data quality, and assessing the impact of the data quality on the conclusions drawn from the trial.

It is the Data Manager's responsibility to help define the necessary data quality level of a trial, to design data collection and processing so as to achieve the desired level of data quality, to measure data quality, to monitor it throughout the trial, and to communicate data quality information to the statistician. This measuring data quality section provides Minimum Standards, Best Practices and methods for the CDM tasks of measuring data quality. It is the statistician's responsibility to assess the impact of the data quality on the trial results.

Minimum Standards

- Use quantitative methods to measure data quality.
- Use standard processes for each trial. Familiar, controlled and understood processes prevent errors.
- Use statistically appropriate inspection samples for decision making.
- Do at least one quality inspection prior to or after locking a trial database that determines the error rate of the database if process control is not demonstrated. In the case where 100% manual review is used instead of quantitative measures, the method and timing of review should be stated in the clinical study report to support confidence in the trial data.
- Statistically assess and document data quality.

Best Practices

- Include compliance of procedures to regulations, compliance of practices to written documentation and conformance of data to source documentation and written procedures in the inspections assessing data quality.
- Provide error rates for variables used in primary and secondary safety and efficacy.
- "When a long series of data processing steps occurs between the source document and the final summaries (as when the source document is transcribed to a subject's chart, transcribed onto a case report form, entered into a database, and stored in data tables from which a narrative summary is produced)" compare the final summary directly against the source document, at least on a sample of cases.[3]
- Monitor aggregate data by site during the treatment phase of the trial to detect sites whose data differ significantly in terms of central tendency and variability so that corrective action can be taken.
- Use predefined criteria to trigger site audits based on the performance-monitoring reports.
- Use inspections (CRF-to-database) early, mid and late trial, or documentation of process control or a standardized and validated data collection process.
- Document anomalies discovered during the data handling process that required corrective action in a data handling report and available at the time of analysis and regulatory review.
- Streamline data collection and handling to limit the number of hand-offs and transfers.
- After the safety distribution is well understood, conduct large trials that collect only the minimum amount of data.
- Perform a data quality impact analysis.
- Evaluate the results of the impact analysis and propose system and process changes.

Data Errors

A data error occurs when a data point inaccurately represents a true value. The source or cause of a data error could be incorrect transcription at the clinical site, incorrect data processing, an ambiguous question that yields unintended responses or collecting data outside a required time window. This definition is intentionally broad and includes errors with root causes of misunderstandings, mistakes, mismanagement, negligence and fraud.

The clinical research process can be complex, involving many possible process steps. Each step at which data are transcribed, transferred or otherwise manipulated has an error rate associated with it. Each subsequent step can create or correct errors.

Common errors in clinical trial data are compiled from several references and are shown in Table 1.[1, 3, 4, 5] Most of these errors are identifiable with proper planning, validation and quality control. Table 1 also categorizes the sources of error by the most efficient way to detect the error.

Table 1: **Error Sources and Detection Opportunities**

Source Of Error	**Hard to Detect**	**Sometimes Possible to Detect with Redundant Data Collection**	**Most Easily Detected by Source Data Verification**	**Most Easily Detected in Data Validation**	**Most Easily Detected by Aggregate Statistics**	**Most Easily Detected by CRF-to-Database Inspection**
Subject completes questionnaire incorrectly or provides incorrect or incomplete answers to questions (lack of tool validation or bad form design)	**X**	**X**				
Subject does not follow trial conduct instructions	**X**		**X**			
Inadequate instructions given to the subject	**X**				**X**	
Site personnel trial conduct error (protocol violation)			**X**		**X**	
Data captured incorrectly on the source						
Site personnel transcription error			**X**			
Site equipment error					**X**	
Human error in reading equipment or print out or inter-rater-reliability			**X**			
Data Entry Error				**X**		**X**
Electronic Data Acquisition error (power glitch, back up that didn't run, lead not attached securely)				**X**		**X**
Data linked to the wrong subject				**X**		**X**
Database updated incorrectly from data clarification form or query						**X**
Programming error in user interface or database or data manipulations						**X**
Lost data				**X**		
Fraud	**X**				**X**	

Most data errors that are detectable are discrepancies between the source documents and the CRF. Some errors, such as missing data, outliers and local inconsistencies, are easily detectable by programmed checks that are run on the database routinely. On the other hand, an incorrect source, misunderstandings, non-compliance, protocol violations or fraud can sometimes cause errors that are difficult to detect. The most irresponsible source of error is an assumption that the data are automatically error free and correct.[6]

Discrepancies between the source and the CRF or Electronic Data Capture (EDC) database can be identified during source data verification (SDV*)* conducted as a part of routine monitoring visits to the site. These can also be identified through collection of redundant data[7]. The technique of using redundant data collection to identify data errors should be used when other error identification methods are not likely to catch the errors. Errors that are more difficult to identify include data recorded under the wrong subject number and data captured incorrectly on the source.

Lack of inter-rater reliability occurs in cases where there is subjectivity and more than one individual is assessing the data. For example, not all investigators use the same criterion for Myocardial Infarction. Some investigators use ECG information, others rely on enzymes or clinical presentation and many use combinations of these to judge whether or not a subject had a heart attack. Due to inter-rater reliability, the question "Did the subject have an MI?" could mean different things to different people. Inter-rater reliability can be assessed through comparisons of average ratings by rater or by site, or by a quality control sample of redundant reviews[1].

Fraud and protocol violations, although most definitely errors, can be difficult to detect without the use of special programming and the use of aggregate statistics.[1, 6, 5, 8, 9, 10]

This can take the form of informal PROC UNIVARIATE, PROC FREQ, or other SAS® procedures that provide quick aggregate statistics, or through Statistical Process Control (SPC) charts.[10] Throughout a trial, aggregate statistics or SPC charts should be available to the monitors to facilitate detection of misunderstandings, misconduct and fraud. When at the site performing SDV, the monitors often do not have access to aggregate data. Often the data have not been entered into a database yet. Clinical Data Management is the first point in many processes where the data are available for viewing in aggregate, across sites. It is at this, earliest point that aggregate statistics should be provided to monitors and other study personnel to quickly identify sites that are behaving differently from the rest. Reports of aggregate data should summarize the performance of individual centers in the areas of recruitment, extent of follow-up, compliance to treatment, completion of procedures, late visits, data queries and a few key data points to detect misunderstandings, misconduct and fraud.[3]

Discrepancies between the CRF and database can be caused by any step in the data handling process including programming errors. Data handling or processing errors are identified through a CRF-to-database comparison. Many errors can be identified and corrected during data processing, or as part of routine quality checks designed into the

data handling process. For example, data validation checks are usually run on an entire database and discrepancies are corrected as part of the data cleaning process.

The most efficient error detection method should be used for each source of error included in the inspection scope. Table 1 names four different types of error detection methods. Data validation or edit checks and programming utilizing aggregate statistics are applied consistently to all of the trial data. All errors that are identified in this manner can be corrected. However, not all errors can be detected using these methods. For example, unreported adverse events appearing as marginal comments on a CRF are difficult, if not impossible to identify with programming. Source Data Verification and CRF-to-database inspection will uncover errors that are difficult to catch with programming, but are resource intensive and are usually not applied to 100% of the trial data. They are also subject to human error and inconsistencies. Error detection methods that are applied to only a sample of the data can be used to obtain estimates of the distribution of undetected errors.

It is not practical to design a quality check for every possible error. Quality checks performed as part of data processing , such as data validation checks or edit checks, should target fields critical to the analysis, where errors are expected to be frequent and it is reasonable to expect a high percent of error resolution. There will always be errors that are not addressed by quality checks, and errors that slip through the quality check process undetected. The purpose of measuring data quality is to identify and quantify these errors so that quality checks can be added or deleted and the desired level of data quality maintained.

Data Quality Inspections

A 100% manual review of the data and the source is not practical, necessary or efficient in identifying errors or cleaning data. However, if 100% manual review is used instead of quantitative measures, the method and timing of review should be stated in the clinical study report to support confidence in the trial data.

Errors can be detected as a discrepancy between two representations of the data captured at different points in the data handling process. For example, at the investigator site a clinical trial monitor performs Source Data Verification (SDV) by comparing the medical record (a subject's chart) to the CRF. Any discrepancies between the two that are not explained by CRF completion instructions, the protocol or other approved site conventions are counted as errors. Another example is a CRF-to-Database inspection performed by comparing the CRF to the data in a database. The database could be the clinical database immediately following data entry, or it could be the analysis-ready datasets at database lock. In the case of the latter, an error is a discrepancy between the datasets and the CRF that is not explained by data handling conventions, site signed data clarification forms, or programming conventions defined in the trial analysis plan. The time required to obtain the print outs, obtain the CRFs, locate the correct versions of the data handling conventions, perform the comparison, and document the findings for an average-density, 100-page CRF is about 4 hours per CRF inspected. Proper

documentation of the errors is crucial to assure that correct information is available for error classification and error rate calculation.

Two different definitions of inspection are given in the glossary. ICH E6 defines inspection as the act by a regulatory authorities of conducting an official review of documents, facilities, records, and any other resources that are deemed by the authorities to be related to the clinical trial and that may be located at the site of the trial, at the sponsor's or CRO's facilities or both, or at other establishments deemed appropriate by the regulatory authorities. The American Society for Quality (ASQ) defines inspection as measuring, examining, testing, and gauging one or more characteristics of a product or service and comparing the results with specified requirements to determine whether conformity is achieved for each characteristic. Here the term inspection is used to indicate a scope narrower that an audit, and a process where the measuring can be done as a step in the work process with less independence than an audit. For example, a CRF-to-database inspection may be carried out by individuals in the same department, or on the same project team as those who did the work being inspected, as long as they are not the individuals who did the work that is being inspected. This definition of inspection is used throughout this section because the scope and level of independence are consistent with the ASQ definition.

It is very important that the scope of an inspection is planned to cover all sources of error if it is to be used to estimate the database error rate. The trial leadership should not be fooled into thinking that a CRF-to-database inspection comprehensively quantifies data quality. The ideal inspection scope addresses all of the sources of error by including an assessment of each error source in the inspection. Addressing all of the sources of error is most easily accomplished by comparing the final product (a clean or analysis database) to the source.[3] The inspection should incorporate programmatic checks for outliers and logic inconsistencies by assuring that they were applied reliably or by performing them at the time of inspection if other cleaning methods were used. If this is done on a representative sample of data points, the percentage of errors attributable to each source will be represented in the database error rate. The ideal can be approximated by a CRF-to-database inspection on a sample of CRFs, and a separate assessment of the errors detected during monitoring visits. The CRFs for the monitoring sample do not have to be the same as those in the CRF-to-database inspection, although that would be ideal. The error rate from this inspection includes contributions from the CRF-to-source errors not corrected plus the total CRF-to-source errors for any unmonitored CRFs, plus the errors from the CRF-to-database inspection.

If a standard data collection process is not used, inspections should be designed and conducted on every trial to sample the database and identify errors. These inspections provide a way to estimate the type, location and extent of errors that occur during the conduct of a clinical trial. If a standardized data handling process is used, the type, location and extent of errors should be known, measured and monitored. Therefore, an inspection on every trial will not be necessary if process control is demonstrated.

Minimum acceptable practices include a quality inspection prior to or after locking a trial database. A lock inspection provides the necessary assessment of the quality of the database, but does not provide information about type, location and extent of errors at a time when future errors can be prevented.[10] If serious errors exist, the risk, cost and time invested in the trial can be lost. Any data that are uncorrectable are labeled as such and are often excluded from analysis, or extra analyses are conducted to determine the effect of the flawed data. Conducting one "database closure" inspection will, however, assure that the quality of the data is known prior to conclusions being drawn from the analysis.

Sample Size for Audits, Inspections or Reviews

The number, type, extent and location of errors in the database should be estimated based on the qualified inspection of a representative sample of data. The information obtained in inspections can be used to represent the quality of the database only if an appropriate sample size was included in the inspection. A representative sample is a subset that contains similar attributes of interest, is usually randomly selected with respect to the attributes of interest, and is calculated through applicable statistical sampling techniques. There are many acceptable statistical sampling methods.

When using a "fixed percent" sample size provide ranges that define the size of databases and the quality levels for which the fixed percent is comparable to appropriate statistical sampling methods. For a small trial, a "fixed percent" sample size may not adequately support estimation of the true random error rate. For a large trial, a "fixed percent" may use hundreds of person hours while adding no additional information to the error rate estimation.

The best and only acceptable practice for sample size selection is using a statistically appropriate sample size for each inspection. This assures that the information obtained from the inspection can be considered representative of the entire database, and can be used in decision-making. This best practice can be achieved in many ways. Acceptable alternative practices to statistical sampling do not exist. The information obtained from an inspection is only useful in representing the quality of the database if the sample size was documented and appropriate. It is important that the sampling methodology be planned and documented by statistical personnel.

Error Rates: The Safe and Effective Way to Interpret Data Quality

In the absence of industry-wide standards for acceptable error rates, the phrase "quality data" is essentially undefined. Some organizations have arbitrarily set acceptable quality levels. Popular approaches include rates of 50 errors per 10,000 fields overall and different standards for critical and non-critical variables ranging from 0 - 10 errors per 10,000 fields for critical variables and 20 - 100 errors per 10,000 fields for non-critical variables. As an extreme case, unfounded rumors suggest that a single error detected in regulatory review could result in the rejection of an entire database or loss of credibility of a submission[2]. Another extreme example is the expectation that the error rate for "critical variables" or in "critical panels" should be zero and can be obtained by 100% manual review. Although the published data on error rates is sparse, study data have shown less than 1% of the data discrepant with the source, less than 2% unverifiable in

some cases and up to 3 or 5% in others[1]. The 1 & 2% error rates did not affect the trial conclusions.[1] The published error rate for human inspection is 15%, meaning that a human looking for errors can miss 15% of them on the first inspection[11]. These data, when interpreted in isolation, suggest that the arbitrarily set standards ranging from a 0 to 0.5% error rate for data processing may be unnecessary and masked by other, less quantified errors. When accounting for all error sources from the subject to the analysis database, an error rate of .1% at any data processing step may be too large depending on the error rates at the step in question and the number of prior and subsequent process steps.

Barring subjective assessments mapped to a numeric scale (such as a visual-analog scale), the available methods to quantify data quality are raw counts of numbers of errors, and error rates. Caution should be used when interpreting raw error count data. These data can be misinterpreted if used to compare the quality of different database tables in the same database, or data quality from inspections on two trials. Calculating an error rate guards against misinterpretation of error counts, can facilitate comparison of data quality across database tables and trials and, therefore, is the preferable method.

The error rate is the number of errors detected divided by the total number of fields inspected.

$$Error\,Rate = \frac{Number\ of\ Errors\ Found}{Number\ of\ Fields\ Inspected}$$

Error rates are usually expressed as the number of errors per 10,000 fields. Scaling the error counts in this way gives a distinct advantage over raw error counts. For example, say two database tables or datasets, DEMOG and VITALS were inspected at a sample size of 20 subjects. There are 100 fields in DEMOG and 400 fields in VITALS in the inspected sample. There are 10 errors found in DEMOG and 20 errors found in VITALS. The error rate is 1000 errors per 10,000 fields in DEMOG and 500 errors per 10,000 fields in VITALS. The DEMOG panel error rate is twice the VITALS panel error rate even though half as many errors were detected in DEMOG as in VITALS. By presenting error counts as errors per 10,000 fields, the data quality can be compared across not only database panels or datasets, but also across trials. The error rate gives a common scale of measurement for data quality. This is why we recommend error rates as a minimum standard. Error rates should always be presented along with a description of how they were calculated.

There are many ways to quantify data quality and calculate an error rate. While the differences among the methods can be subtle, the differences among the results can be a factor of two or more. Therefore, standard methods of identifying errors and quantifying data quality must accompany standards for acceptable data quality.

For the error rate to be used as a common measure of data quality across vendors and trial sponsors, a common algorithm should be followed to calculate an error rate. The units in

the numerator and denominator must be the same. For example, consider the hypothetical situation of two lab data vendors calculating error rates on the same database with three panels. The protocol number, site number, and sponsor number are hard coded in all of three database panels. Vendor 1 counts each of these fields in the field count as fields inspected for a total of 100,000 fields inspected, Vendor 2 does not include them in the field count since they are hard coded, for a total of 50,000 fields inspected. Both vendors do a data quality inspection. Both vendors find 10 errors. When they calculate the error rates, vendor 1 has an error rate half that of vendor 2 only because they did not follow the same algorithm for field counts.

This is why, in trials employing multiple vendors, the error rates must be calculated consistently. Likewise, in all trials being submitted for regulatory review, the error rates should be consistently counted and the algorithm followed should be disclosed.

There are several important concepts about error rates.
Error rates for parts of a data handling process can be combined to give an estimate of the error rate in the data handled using the process.
Correction rates, for example those achieved through data cleaning, should also be taken into account.
The error rate is only a part of process evaluation. It is important to know if the errors are in critical or non-critical fields. If the high error rates are in non-critical fields, they are of little impact.
Knowledge of the published error rates or your organization's error rates can help you choose the process paths and technology that will yield the highest quality for your organization.
Due to the compounding nature of error rates, and the difficulty of obtaining error rates for some processes, the notion of large simple trials seems to offer the highest integrity data. The best practice would be large trials that collect only the minimum amount of data using the data handling process with the least steps.

Recommended Standard Operating Procedures
- Measuring Data Quality
- Monitoring Data Quality
- Data Quality Acceptability Criterion

References
N/A

Other Related Reading
Califf RM, Karnash SL, Woodlief LH. Developing Systems for Cost-Effective Auditing of Clinical Trials. *Controlled Clinical Trials*. 1997;18:651-660.

Davis JR, Nolan VP, Woodcock J, Estabrook RW. eds. Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making. Workshop Report. *Roundtable on Research and Development of Drugs, Biologics, and Medical Devices,*

Division of Health Sciences Policy, Institute of Medicine. Washington DC: National Academy Press; 1999.
Knatterud GL, Rockhold FW, George SL,et al. Guidelines for Quality Assurance in Multicenter Trials: A Position Paper. *Controlled Clinical Trials*. 1998;19:477-493.

DeMets D. Distinctions between Fraud, Bias, Errors, Misunderstanding, and Incompetence. *Controlled Clinical Trials*. 1997;18:637-650.

Williams GW. Discussion: Errors, Misunderstandings, and Fraud. *Controlled Clinical Trials*. 1997;18:661-666.

Ventura D. Issues in Clinical Trial Management, Part II Corrupt Data: Sources and Solutions. *Research Nurse*. 1999; March/April:12-17.

Calvert WS, Ma JM. *Concepts and Case studies in Data Management*. Cary, NC: SAS Institute Inc.; 1996.

Bauer P. Statistical Process Monitoring During Clinical Trials. In: *Abstracts from the DIA Thirty Fifth Annual Meeting; June 27 – July 1 1999, Baltimore, MD*: Drug Information Association; 1999.

Knatterude GL. Quality Assurance in Multicenter Clinical Trials: An Overview. *Abstracts from the DIA Thirty Fifth Annual Meeting; June 27 – July 1 1999, Baltimore, MD*: Drug Information Association; 2000.

Nahm ML. Optimizing Data Quality Through Statistical Process Control. *Proceedings from the 15 Th. Annual US Clinical Data Management Symposium, March 2000, Philadelphia, PA:* Drug Information Association.

Anderson P, Wortman B. *CQA Primer.* Indiana: Quality Council of Indiana; 1997.

Assuring Data Quality

Introduction

Data Management professionals have all heard the three steps to compliance:

1. Say what you are going to do.
2. Do what you said you were going to.
3. Document that you did it.

Assuring data quality, however, goes beyond compliance. Data quality is not the responsibility of an independent quality assurance group. Assuring data quality begins with top management's commitment to appropriate infrastructure and resources to conduct clinical trials, and is the responsibility of everyone involved in the trial.

Discussing quality assurance of clinical trials can be complicated. The quality industry terminology and methodology as used in ISO standards or popular quality literature does not always correspond to those used in the clinical trial industry. Even more complexity is added because organizations use diverse jargon for quality practices. For example, in some organizations a QA (Quality Assurance) group performs independent audits on trials; in others, the same activity is called quality control and is performed by each department. In some organizations, QA is responsible for regulatory compliance and serves as an internal consultant to departments; in others, QA performs routine system audits internally and of vendors. In yet other organizations, QA plays a large role in SOP development and review, process improvement and training.

With the development of the Quality System Manual, the FDA has started to bridge the gap between the drug development industry and the quality industry. The purpose of this section is to build on these efforts by applying standard Quality System terminology and methods to assuring clinical trial data quality.

Scope

The section on Measuring Data Quality concentrated on identifying errors in clinical trial data, counting them, interpreting them. Although data quality measurement methods are necessary and should be applied to clinical trial operations as part of an overall planned approach to achieving data quality, a larger emphasis should be placed on error prevention early in the protocol and data handling process design stages.

The emphasis in this section is on infrastructures and practices used to assure data quality.

Minimum Standards

- Design and maintain data handling processes according to a documented organizational quality system.
- Provide sufficient information in the data processing documentation to reproduce the final analyses from the source data.

- Assess data quality for every study to be submitted for regulatory review.
- Assure that data quality is appropriate for the trial analysis.

Best Practices

- Have an organizational Quality Policy that is strongly supported by upper management, understood by all staff and supported by operational procedures.
- Use standardized or validated data collection and handling processes.
- Include as few steps as possible in the data Collection and data handling processes to eliminate the chances for error.
- Collect only data essential for interpretation of trial results.
- Focus on error prevention in Quality Assurance. Focus on process monitoring in Quality Control. Neither should be focused on inspecting quality into the final product (database or programming).
- Include compliance of procedures to regulations, compliance of practices to written documentation and conformance of data to source documentation and written procedures in audits to assess data quality.

Quality System Approach

A quality system encompasses the organizational structure, responsibilities, procedures, processes and resources for implementing quality management[1]. A Quality System Approach has been taken by the FDA in the device regulations[2]. This approach is also used by ISO and, due to its flexibility, has been successfully implemented in most industries including the clinical trial industry. The flexibility is crucial to our industry because of the individuality of each trial. The quality level and approach to achieving it should be appropriate for and should address the purpose, characteristics and complexity of each study[3]. The infrastructure advocated in a quality system approach provides the flexibility to account for trial differences in a controlled and consistent manner. A key concept in the quality system approach is that the structure, format, content and method of presentation should be determined based upon the needs of the organization[4]. Many components of a quality system, policies and procedures for example, are already in place in most clinical research organizations.

Regardless of the documentation structure used, the Quality System should assure the following:

1. Written procedures exist that specify the process steps and decision points applied to handling and processing of clinical data intended for regulatory submission;
2. Written procedures are specific enough to allow reproducibility of the clinical database used for analysis from the source documentation at the site;
3. Written procedures are followed;
4. Data are of sufficient quality to "support conclusions identical to those drawn from "error free" data" [1].

Quality Policy
A quality policy is the highest level of a quality system. It communicates and documents the overall intentions and directions of an organization with respect to quality, as formed by top management. The quality policy should provide for intermediate and detail levels of the quality system including management review procedures, quality manual, and quality plan[5].

Each organization should have a written quality policy. In other words, an organization's top-level management should demonstrate commitment through support for an adequately resourced infrastructure that facilitates error prevention through off-line quality control activities such as quality engineering, quality planning and procedures applicable to each trial.

Quality Manual & Plans
The organization should have a written quality manual that defines the quality practices, resources and activities relevant to the data handling services that are being provided by the organization. Most organizations already implement a quality manual as Standard Operating Procedures. The organization must have written procedures that describe how they intend to meet their quality requirements described in the quality policy. These can be specified in Quality Manual or Quality Plans for each trial or process. The quality manual and plans must include flexibility to address the specific requirements of therapeutics having specific regulatory requirements.

For organizations with a high degree of standardization, information that otherwise would be part of a trial or project specific plan may be found in an organization's quality system documentation such as the Quality Manual, Audit procedures and Standard Operating Procedures. The plan in this case should include reference to applicable quality system documents and how those documents assure the data quality of each trial. Quality plans may be specific to one trial or be generic to all trials for which the organization has all or partial responsibility.

Standard Operating Procedures
As previously stated, many organizations in this industry implement the quality manual as Standard Operating Procedures. Organizations should have a documented process for creation, review, deviation from and version control of SOPs. Effective dates should easily identify time periods over which the SOPs were used. SOPs should be archived according to documented organizational procedures and do not have to be archived with each trial. Planned deviations from SOPs should have the same level of review and approval of the SOPs from which they are deviating.

The detail level in SOPs is determined by the level of standardization within an organization. For example, an organization with standard CRF modules, database structure, and monitoring procedures may have more detail in SOPs, and thus require less in trial-specific documentation such as a Data Management Plan.

Each section of the Good Clinical Data Management Practices lists recommended Standard Operating Procedures. The Association for Clinical Data Management (ACDM), the European sister organization to SCDM, has published Guidelines for Writing Standard Operating Procedures[6]. These guidelines provide suggested content for the following CDM areas and are a valuable reference:

- Data Management Plan
- CRF Design
- Database Preparation
- Data Entry and Verification
- Data Coding
- Data Validation
- Discrepancy Management
- Editing Data
- Database Locking / Unlocking
- Database Unblinding
- Project Document and Archiving Procedures
- Selection and Management of Contract Research Organizations
- Creation and Maintenance of Training Records

Trial-Specific Procedures

It is recognized that each trial may have unique data handling needs due to sample size, visit schedule, type of data collected, amount of data collected, and method of data collection. Organizations should clearly document trial-specific procedures to assure that the analysis database is reproducible from the source documents. This documentation should provide supporting details to SOPs and may have a lower level of review and approval within the organization. Trial-specific procedures are often known as Data Handling Plans, Data Management Plans, Data Handling Protocols, Data Quality Management Plans, and Data Management Binders.

The ACDM has also published Data Handling Protocol (DHP) Guidelines[7]. These guidelines provide an outline and contents list of items to be covered in an organization's trial-specific procedures. The intent is that organizations customize the DHP contents and detail level depending on the detail level in their SOPs. The DHP Guidelines are an excellent reference for defining and developing organizational trial-specific procedures.

Standardization

There is currently much interest in standardization across the industry. Efforts described in published papers and presentations range from a standardized CRF Library, to standardizing protocol sections complete with CRF modules, database tables, edit checks and analysis tables and listings, to implementing a consistent data handling process with or without special software solutions. Unfortunately, there is little evaluative data on the topic. The general consensus seems to be that standardization has the potential to shorten timelines, reduce cost and increase data quality.

The clinical research process can be complex, potentially involving many process steps. Each step at which data are transcribed, transferred or otherwise manipulated has an error rate associated with it. Each subsequent step can add to or subtract from that error rate. For this reason, data collection and handling should be as standard and streamlined as possible, limiting the number of hand-offs and transfers.

Standardization helps to decrease errors. Regardless of the complexity, a standard process is familiar and understood. Sources of error are known and more often quantified. In a standard process there is less "uncharted territory" where unexpected errors and complications can arise. However, opportunities for standardization can be very different depending on the organization. A large pharmaceutical company has more potential for standardization than does a CRO, for example.

Standardization discourages the adding of unnecessary extra or ad-hoc process steps. The discussion on error rates in the Measuring Data Quality section provides a conceptual framework in which to understand the expectation of shorter timelines, reduced cost and increased data quality. The fact that error rates are unmeasured and in some cases unknown for some major process steps in clinical trials causes concern. The common practice of designing a new data handling process, or adding complexity to data handling processes for each individual trial is overwhelming when evaluated in terms of error rates. Use of a standard process enables an organization to fully characterize the process performance and institute controlled and evaluated improvements. Successful standardization efforts can also allow for the flexibility for trial teams to address and document trial-specific processes where necessary.

Quality Assurance of the CDM Function

Quality Assurance is the set of activities that ensure that practices are in place and effective in producing a quality product. ICH E6 defines Quality Assurance as: "All those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with Good Clinical Practice (GCP) and the applicable regulatory requirement(s)". In the clinical research industry, quality assurance has included the administration and support of organizational standard procedures and documentation and, in many cases, assessing compliance of the policies, work process and products with regulatory standards. For this reason, existing Quality Assurance requirements can be instrumental in interpreting the existing organizational policies and procedures in the context of a Quality System Approach and, where needed, work with functional groups to extend organizational policies and procedures to meet the intent of a Quality System Approach.

It is recommended that quality assurance activities such as developing and maintaining an organization's quality system be in place prior to trial work. Trial work should take place within an organization's documented quality system.

Assuring that the trial data meets the required quality level at data manipulation points requires specific tools and quantitative techniques that should be part of an organization's quality documentation. These include process monitoring, auditing, sampling and error

rate calculation and are essential to quantify the quality of the data and assess the potential impact of the data quality on the conclusions reached from the study. An organization's written procedures should describe the approach to assuring that data are reliable and processed correctly at each stage of data handling for a trial[8].

The quality documentation should be sufficient to assure consistent application and methodology in the use of quantitative tools to measure data quality. The quality plan should result in a reproducible quantification of the data quality and documentation that the data quality does not affect the ability of the data to support conclusions and interpretations equivalent to those reached from "error free" data.

Auditing as a Quality Assurance Function

Auditing is defined by the American Society for Quality (ASQ) as a systematic and independent examination to determine whether quality activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve objectives. Similarly, audit is defined by ICH E6 as: "A systematic and independent examination of trial-related activities and documents to determine whether the evaluated trial-related activities were conducted, and the data were recorded, analyzed and accurately reported according to the protocol, sponsor's standard operating procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s)"[8]. For this reason, the word audit is used throughout this document to indicate an independent assessment. The word inspection is used to indicate in-process or quality control activities to measure data quality.

Auditing of Clinical Data Management (CDM) requires not only training in auditing methodology, but also specific CDM familiarity as well as knowledge of computer programming fundamentals and industry regulations. Individuals auditing CDM functions should be qualified by training and experience to assess compliance of CDM procedures. Audits of the CDM function should be performed often enough to assure that the CDM process and quality control procedures are effective in producing reliable and reproducible data for analysis and regulatory review.

A comprehensive CDM audit evaluates the entire CDM quality system. Three levels of a CDM audit are:

1. Existence of written procedures for CDM processes that are compliant with Federal Regulations.
2. Documented compliance of the CDM organization to its written policy.
3. Objective evidence that the CDM process produces reliable clinical data for analysis and FDA review and quantification of data quality.

Regardless of an organization's documentation structure, written procedures should exist that specify the process steps and decision points applied to handling and processing of clinical data including manual review instructions, data entry conventions and instructions for obtaining data clarifications from the site. Written procedures should be specific enough to allow reproducibility of the clinical database used for analysis from the source documentation at the site. Compliance of these written procedures is assessed

by a "Table Top Audit" where an auditor assesses the written procedures with respect to the Federal Regulations.

Documented compliance of the CDM organization to its written policy consists of objective evidence that the written data handling procedures were followed. This can include a database audit trail, signed and dated checklists, signed data clarification requests from a site or interviews with CDM staff during an audit.

Objective evidence that the CDM process produces reliable clinical data for analysis and regulatory review is a multi-step process. The first step is the quantification of the quality of the clinical data, usually represented by an error rate. Additional objective evidence includes data that demonstrate that an organization's data handling process is operating in a state of control. Other important evidence is the assessment of the potential impact of the error rate on the conclusions or interpretations based on the data.

CDM Quality Control

ICH E6 defines Quality Control as "The operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled"[8]. Data quality is the result of the planning, execution and analysis of a clinical trial. Each step in the clinical research process should be designed to assure that the necessary level of data quality is maintained throughout the trial. This process starts with protocol and CRF design, includes data processing, analysis, and regulatory review and is complete at the close of the required record retention period. Quality control should be applied to each stage of data handling to assure that data are reliable and processed correctly[8]. Each process step where data are transcribed, transferred, updated or otherwise saved to a new medium requires a corresponding quality control step. Appropriate corrective action should be initiated where data quality does not meet pre-defined acceptance criteria.

The Clinical Data Management (CDM) function designs and performs many of the data manipulations that occur in a clinical trial. The CDM design functions include but are not limited to input to the study protocol regarding data collection and management, data collection tool design, database design, and definition of the data integrity checks. The CDM production functions include data review, data entry, data editing, discrepancy resolution, coding, reporting, data transfers and reconciliation with external data. The required quality level of each function differs not only by trial, but also depends on preceding and subsequent processes.

Quality must be measured before it can be controlled. Measurement necessitates quantifying data quality. Historically in the clinical research industry, data quality has been quantified through the calculation of an error rate. In order for error rates to be a useful means of comparing the quality of databases produced by different vendors, or for use in regulatory review, they must use the same scale and precision. They must measure the same thing and be obtained using a standard measuring tool. This means all error rates must be created equally and represent the same sources of error and count the errors in the same way.

Quality Controls are the operational techniques and activities that are used to fulfill requirements for quality. These inspection activities can be performed as part of the work process or part of monitoring the work process. Examples of quality control procedures include double data entry, programmatic data range and consistency checks, regular evaluation of error rates to assess process control and process steps that assess the quality of upstream processes. Error prevention, detection and monitoring activities should be described in an organization's written procedure and documented as evidence of ongoing quality control.

Quality control activities should be performed at the earliest point in a process to provide the maximum return in terms of error prevention. They should provide quantitative measures of the quality of the data and assess process control. Documentation of Quality Control activities should be available as evidence that these activities were performed. The database quality should serve as evidence that the quality controls are effective.

Two types of quality control can be utilized for the CDM process to meet the guidelines set forth in ICH E6, section 5.1.3. In-process quality control consists of reliability and compliance checks as process steps. In-process QC is recommended for the CDM design functions. These include CRF design, site education materials or instructions, CRF annotation, and programming specifications. Each item should be inspected or reviewed prior to being released for use. For example, consider that one CRF design error can have a cascade effect on data quality, initiating hundreds or thousands of resulting data errors. This approach is also indicated prior to database lock. The in-process quality control activities usually result in "go-no-go" decisions, each at a very crucial point in a trial's operational design.

General on-line quality control, however, is applied to processes, not necessarily to each trial, or as a step in a process. For CDM production or data processing functions, an on-line quality control approach uses inspection error rates to monitor an organization's process to assure that the process is operating in a state of control. A process that is in control will consistently produce data that are reliable and processed according to trial-specific guidelines. Process control includes inspecting periodic samples of the data, usually at regular intervals, and taking corrective action on the process only where, collectively, the inspection results indicate a trend, an out of control process or consistently poor quality. This approach is recommended on a risk/benefit basis because an error in processing has only an additive effect on the downstream data quality versus the cascade effect from a design error. However, each manipulation point that the incorrect data point passes through will have to be reworked to correct the error. Therefore, a process that is operating in a state of control will not only meet ICH E6 section 5.1.3, but will also reduce rework, data cleaning and inspection costs.

Recommended Standard Operating Procedures

- Development and Maintenance of Standard Operating Procedures
- Development of Planned Deviations from Standard Operating Procedures
- Development and Maintenance of Trial-specific Procedures

- Quality Assurance Audits

References
N/A

Other Related Reading
Juran JM. *Juran's Quality Control Handbook*, 4th ed. New York, NY: McGraw-Hill; 1988.

Code of Federal Regulation, Title 21, Parts 800-1299, and specifically 21 CFR 820 *Quality System Regulations*. Food and Drug Association. US Government Printing Office.

1. Davis JR, Nolan VP, Woodcock J, Estabrook RW, eds. Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making, Workshop Report. *Roundtable on Research and Development of Drugs, Biologics, and Medical Devices, Division of Health Sciences Policy, Institute of Medicine*. Washington, DC: National Academy Press; 1999.

2. ANSI/ISO/ASQC/Q9004-1-1994, Quality Management and Quality System Elements Guidelines.

Food and Drug Administration. *Guidance for Industry: Medical Device Quality Systems Manual*. Rockville MD: U.S. Department of Health and Human Services, Food and Drug Administration; April 1999.

Association for Clinical Data Management. *ACDM Standard Operating Procedures Working Party, Guidelines for Writing Standard Operating Procedures*, Version 2.0. November 1997. Available from ACDM Administrator, PO Box 129, Macclesfield, Cheshire, SK118FG.

Association for Clinical Data Management. *ACDM Data Handling Protocol Working Party, Data Handling Protocol Guidelines*, Version 2.0. November 1996. Available from ACDM Administrator, PO Box 129, Macclesfield, Cheshire, SK118FG.

International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Section 5.1.3; 1996.

Database Closure

Introduction

Closing or locking the study database is fundamental in preventing inadvertent or unauthorized changes once the final analysis and reporting of the data have begun. Although important in open-label studies, it is even more critical in randomized trials in order to preserve the integrity of the randomization process once the blind has been broken. A well-defined process for closing a database and change control procedures in the event of the necessity of reopening the database are essential.

Scope

Different terms are used to indicate the time-point at which the data are considered ready for analysis. Database closure includes interim analyses or reports from which trial or regulatory decisions will be made. The requirements for interim analyses and final analysis may differ. However, these requirements and steps taken to achieve them should be documented. Organizations perform and document database closure activities differently. For this reason, the section provides a list of recommended closure documentation rather than specific methods.

Minimum Standards

- Ensure there is a procedure defining database closure methodology.
- Document completion of all defined tasks or criteria prior to database closure.
- Ensure that all team members are notified and edit access is removed and documented at final database closure.
- Have written procedures with clear criterion for unlocking a database after closure.

Best Practices

- Develop and use a database closure checklist.

Database Closure Process and Checklist

Database lock must be documented as a definitive point in time, where proof of the removal of edit access can be shown. In order to decrease the necessity to unlock the database after this point, a well defined and organized procedure must be followed to ensure that all data have been processed, the quality level assessed and relevant study personnel are notified or approve the database lock.

Items to consider in the database closure preparation include the following:

- All data have been received and processed.
- All queries have been resolved.
- External data (e.g. electronic laboratory data) are reconciled with the study database and are complete.
- If a separate, serious adverse event database exists, it is reconciled with the main study database.

- The coding list has been reviewed for completeness and consistency.
- Final review of logic and consistency check output has taken place.
- Final review for obvious anomalies has taken place.
- Quality audit of the data and documentation of the error rate have occurred.
- All documentation is updated and stored where required by Standard Operation Procedures.

Once all steps are complete, a documented approval process should take place, which includes sign-off by relevant study personnel (e.g. Data Management, Biostatistics, monitoring representative, clinical/scientific representative). Once approvals have been obtained, edit access to the database should be removed and the date documented.

Errors Found After Database Closure

If, after database lock, errors are found, careful consideration should be given as to how to handle and document these errors. Issues to consider are primarily the effect on the safety and efficacy analysis. It is important to remember that not all errors found must be corrected in the database itself. Errors may also be documented in the statistical or clinical report. Although some companies choose to change all errors found, others may only change those that have a major impact on the safety/efficacy analysis. What is of primary importance is that a company has in place a predefined process to determine how such errors will be processed and documented.

If the database is unlocked after initial lock, the process must be well controlled and, once again, documented. Procedures should include notification of the project team, a clear definition of the change(s) being made and the date of the change. Re-locking the database should follow the same process as the initial lock for notification/approval.

Recommended Standard Operating Procedures

- Database Closure, to clearly define the steps to be taken to close and lock a database. A task checklist is recommended for this purpose as well as sign-off forms for approval and lock details.
- Change Control/Errors after closure to define conditions under which the database will be re-opened and the necessary documentation required for re-lock.

References

N/A

Other Related Reading

N/A

Clinical Data Archiving

Introduction

Clinical data archiving includes the planning, implementing and maintaining of a repository of documents and records that contain clinical data together with any interpretive information from a clinical trial.

Scope

This section provides an outline to help clinical data managers develop an archiving strategy. It includes a discussion of the regulatory requirements surrounding clinical data archives, a description of the components of an archive and information about data formats that can be used to support that archive. This document focuses on the components of the study archive that are the responsibility of data management. There is no discussion about the archiving of study documents such as the study protocol and other regulatory documents, as these sections seldom, if ever, are a data management responsibility.

Minimum Standards

- The clinical data archive should include a centralized table of contents for all studies.
- The accessibility of the clinical data archive should be tested following every major upgrade of the active clinical data management system.

Best Practices

- All clinical data, metadata, administrative data and reference data should be maintained in an industry standard, open systems format, such as CDISC ODM.
- An electronic repository links all study components including the clinical data, CRF (Case Report Form) images in PDF form, program files, validation records and regulatory documentation.
- The audit trail should be stored in open format files in a secure file system location.
- Copies of all user and system documentation for any applications used to collect or manage clinical data are retained in the corporate library or archive facility.
- Reports describing the study metadata, including data structures, edit check descriptions, lab-loading specifications are printed and stored in a central document library.
- The study validation binder should be included in the document library.
- System security reports, including user listings, access rights and the dates of authorization, should be printed and filed or scanned.
- The edit check archive should include all program code for edit checks, functions and sub-procedures together with a copy of the version control information.
- Paper CRFs should be scanned and indexed. If an EDC, Electronic Data Capture, system is used, entry screens should be archived as PDF.

Background

Most clinical data is collected as part of an effort to submit a licensing application to the FDA – either to CDER or CBER. The ICH GCP requirements stipulate that data

collected in a clinical trial must be maintained for a period of two years either following the last regulatory submission or following a decision to discontinue development of a compound/biologic or medical device. To meet this requirement, as well as to ensure that the sponsor is able to answer questions relating to the clinical trial data that may emerge many years after the trial is conducted, it is important to archive clinical data.

Historically, the most common mechanism for long term clinical data storage has been to extract the final data from the clinical data management system into SAS datasets. The extracted SAS datasets are still an important component of the clinical data archive however, with the increasing importance of electronic regulatory submissions in recent years, requirements for clinical data archives are changing. As a result, clinical records that are part of an electronic submission must now comply with the 21 CFR Part 11 ruling, which was originally published in 1997. Part 11 enforces specific requirements with respect to authentication and auditing of electronic records. In addition, the FDA's Guidance for Computer Systems Used in Clinical Trials defines requirements for data archiving. This guidance was published in 1999 as a guide to the interpretation of the Part 11 policies and other related policies. To fully meet the requirements of these regulations and guidelines, a more comprehensive archiving strategy is needed.

Regulations and Guidance

The 21 CFR Part 11 ruling includes no specific requirements for data retention or data archiving capabilities. However, the FDA has made it clear that the intent of the ruling is to supplement the Predicate rules and ICH GCP requirements for those cases where electronic records are either directly or indirectly part of an electronic submission.

Guidance documents with specific mention of archive and record retention requirements include:

- Guidance for Industry: Computer Systems Used in Clinical Trials (CSUCT) published by the FDA in 1999. This document describes fairly stringent requirements surrounding the need to preserve the systems environment in which electronic records are captured and managed.
- *ICH Good Clinical Practice* (Section 5 Investigator requirements) provides information about record retention requirements.
- Draft Guidance for Industry: 21 CFR Part 11 Electronic Records; Electronic Signatures Maintenance of Records. This Draft Guidance, published in July 2002 addresses some of the concerns raised by industry representatives about the stringency of the CSUCT guidance with respect to archiving and describes an alternate strategy involving migration of systems.

Regulatory Guidance is being actively developed in the area of electronic records handling. Before finalizing your clinical data archive design, it is important to consult with the Regulatory Affairs specialists within your organization to ensure your design approach is consistent with the organizations' regulatory policies.

A well-designed clinical data archive can facilitate compliance with the long-term data access requirements of the regulations – for paper based or for electronic clinical trials.

Archive Contents

In order for an auditor to successfully reconstruct a clinical trial, an auditor must be able to view not only the clinical data, but also the manner in which the data is obtained and managed.

Many electronic data records are obtained using data entry screens from an Electronic Data Capture (EDC) or Clinical Data Management (CDM) system. In order to recreate the manner in which data are collected, it is necessary to be able to demonstrate the way that the data entry screens looked and behaved during the entry process. Fortunately, most data collection systems are capable of providing data entry screen printouts both with and without the clinical data. For systems that provide on-line edit checking during the entry process, metadata about the edit checks – including the field where the check is applied, the program code of the actual check, the dates when the check was active – should be part of the archive as well.

In many trials a large volume of data may come from external systems such as lab test results from a central lab or ECG data from an ECG core lab. External data of this type is typically batch loaded into an EDC or a CDM system. In order to re-create this data collection process, load program specifications, logs from the operation of the loading programs and all of the interim load files should be retained.

Once data has been entered into an in-house electronic record, it may be edited to correct transcription errors or transformed as part of a statistical calculation. Enough information must be retained in the archive to trace the data and any modifications to the data. Enough information must also be retained to demonstrate that all modifications to the data have been made in accordance with all applicable guidelines and regulations. This will include the system audit trail, discrepancy logs, queries, query replies and query resolutions.

For data that is managed externally, but which is loaded into an in-house system for reconciliation, reviews or other purposes, it is generally sufficient to limit the archive to the actual data and any information pertaining to how the data is managed internally. The vendor can do archiving of any records that reflect how the data is managed in the external vendor's system. The trial sponsor is ultimately responsible for ensuring that any vendor, who provides trial data, works in accordance with regulatory requirements. Therefore, the sponsor should ensure that any signed contract with a vendor includes a section on archiving. The information in this section should comply with both sponsor and regulatory requirements.

A summary of the types of information that should be included in a clinical data archive is provided in the table on the following page:

Archive Component	Requirement
Clinical data	All data collected in the trial. This includes both CRF data and data that is collected externally (i.e., labs, ECGs or electronic patient diaries).
External data	For data that is collected externally and loaded into a CDMS system, the archive should include all of the load files.
Structural metadata	Information about the structure of the clinical data. Typically this will be information about the tables, variable item names, forms, and visits and any other objects. It also includes codelists.
Coding dictionaries	If data has been autoencoded, using a company dictionary or synonym table, a copy of the dictionary should be included.
Lab ranges	Laboratory reference ranges. If more than one version of reference ranges were used in the course of the trial, each version should be retained in the archive.
Audit trail	Entire contents of the study audit trail. It is essential that the study audit trail be included in the archive in a tamper-proof format.
Listings of edit checks, derived data	Edit check definitions. These may be provided either as program listing files or as a report from the study definition application.
Discrepancy management logs	Listings of records that failed edit checks together with information on how the discrepancies were managed during the course of the study.
Queries	Copies of all queries, query correspondence and query resolutions. Paper queries may be scanned and indexed.
Program code	Program code from data quality checking programs, data derivations and statistical analyses performed with the clinical data. Program documentation should be stored. Ideally, the program documents should be done online and indexed or hyperlinked.
CRF images in PDF format	For paper-based trials, CRF images are typically obtained by scanning the forms and converting them to PDF format. For Electronic Clinical Trials, PDF images of the electronic forms may be created by the EDC/M Application.
Data management plan	PDF or paper version of MS Word and Power Point documents containing the study data management plan.
Study validation documentation	Contents are described in the GCDMP chapter on systems validation. This document may be in paper or electronic form.

Technical Requirements

Designing a Clinical Data Archive for long-term accessibility presents a challenge in the face of proprietary applications, tools and platforms. As technology evolves, vendors provide new versions of their systems; however, they are not always economically motivated to ensure backward compatibility. The older a file, the more likely the file format will not be readable using the current version of a system. For this reason, the ideal clinical data archive should be based on standards and open systems.

The open formats that are typically used for clinical study archives are described in the table below. No single format is ideal in all circumstances. Due to the fact that a study archive will usually include many different types of information, it will most likely include multiple formats. The format chosen for each type of information should be based on the likely future use of the information. For example, if clinical data will need to be re-analyzed, it should be archived in a format that facilitates loading into a database or analysis tool.

Format	Description	Pros and Cons
Comma Separated Values (CSV)	Plain ASCII text with commas used field delimiters. CSV files can be edited with text editors, word processors and spreadsheet programs such as Microsoft Excel.	Pros: Conceptually straightforward. Can be readily imported into almost any database. Cons: Requires separate handling of metadata, administrative data and audit trails.
XML	Extensible Markup Language. Vendor independent, ASCII based technology for transfer of structured information between dissimilar systems. Used as the basis for the CDISC Operational Data Model.	Pros: Open standard created specifically for clinical trial data. Can include structural metadata, administrative data and clinical data within a single file. Cons: Still unfamiliar to many data managers and IT staff.
SAS Version 5 transport files	Open source format provided by SAS corporation. Commonly used for submitting clinical data to the FDA. Can be read by the SAS Viewer that is distributed free of charge on the SAS web site.	Pros: Familiar to clinical data managers and regulators. Works well with SAS data analysis tools. Cons: Proprietary format.

SAS version 5 transport files (continued)		Variable naming restrictions. Requires separate handling of metadata, administrative data and audit trails.
Adobe PDF	Open source format provided by Adobe Systems. Widely used standard for transmission of text documents. Default format for transmission of information to the FDA. Can be read by the Acrobat Reader that is available free of charge from the Adobe web site.	

Long-term data access requirements suggest that the choice of data format is limited to ASCII based formats or formats based on an open standard such as SAS Transport files. The choice may be further influenced by the format used in the original data management or data collection system.

Archives for Clinical Sites

The CFR predicate rules and the ICH Good Clinical Practice (GCP) guidelines specify that a copy of the clinical data must be retained at the investigator site throughout the records retention period. For paper based studies, this can be achieved by keeping a copy of the paper records at the site. For EDC studies that are conducted using an ASP model, it is important to have a strategy in place for ensuring that these guidelines are met. Many EDC vendors will provide PDF files for all of the eCRFs, electronic Case Report Forms, collected from the site during the trial.

Recommended Standard Operating Procedures

- Study Archiving Procedures

References

1. *Code of Federal Regulations,* Title 21, Volume 1, Part 11. Food and Drug Administration. US Government Printing Office; 1998.
2. *Guidance for Industry: Computerized Systems Used in Clinical Trials.* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; April 1999.
3. *International Conference on Harmonisation. Good Clinical Practice.* Federal Register; 1997.

Other Related Reading

The Data Archiving Challenge. Clinical Data Interchange Standards Consortium; May 2000.

Bleicher, Paul, *Diamonds May Be Forever, But Data?* Applied Clinical Trials; February 2002.

Training

Introduction

An effective training program plays a key role in ensuring regulatory compliance, performance effectiveness, and job satisfaction of Clinical Data Management (CDM) employees. There are a number of compelling reasons for developing and implementing effective training programs. Good Clinical Practices (GCP) and other regulatory guidance documents state that all personnel involved in clinical trials must be qualified and properly trained to perform their respective tasks.[1,2,3] Changes in the technical, business, and regulatory environments are occurring more rapidly than ever. As a result, the demands placed on CDM personnel and the scrutiny under which they must perform are increasing. Trainers must be sensitive to these demands and ensure that they optimize the value of the training experience for participants. This section discusses the design, development and delivery of effective training programs for clinical data managers. Core topics for inclusion in CDM training are also discussed.

Scope

This section addresses issues relevant to CDM training. It includes a brief overview of classroom training, as well as computer-based and web-enabled training issues and techniques.

Minimum Standards

- Learning objectives are documented for each component of the curriculum
- Curriculum and individual course offerings, including applicable SOPs, are reviewed and updated regularly to ensure that content remains current and relevant
- All CDM staff members are trained to perform the current required job functions for their assigned roles
- Training documentation is maintained and includes, at minimum, name of course offering, course objectives, name of course instructor, date of course offering, and names of attendees. This documentation should include training that occurs outside the organization.

Best Practices

- A role-specific training curriculum is documented for each position within the CDM organization
- A master training plan, regularly reviewed and revised, documents and prioritizes training needs of the CDM function
- Job needs analyses and audience analyses are performed and guide development of the training plan
- Individual development plans customized to each employee's career objectives and personal development needs are developed and documented
- Training curriculum is evaluated to determine if the class is best suited for instructor-based training, online user training, or a combination of both techniques
- Training support is available for all online user training

- An evaluation of the effectiveness of training is conducted
- Training is offered using a variety of methods to enhance learning
- Training content is consistently represented across all training materials and is consistently conveyed by instructors, coaches, mentors, peers and others involved in assisting the learner to master targeted concepts and skills
- Instructors are qualified by obtaining and maintaining current knowledge of the topics they teach
- Technical training occurs in a separate training environment that simulates the actual CDM work environment
- Standard Operating Procedures (SOPs) document the organizational responsibility for training
- Managers actively allocate time for employee training and development (the amount of time allocated is dependent upon the employee and organizational needs)

Master Training Plan

Training should be considered at both a macro level (overall training needs) and micro level (specific training needs). Appropriate training topics such as computer systems usage, (SOPs) and working practices should be included in the master training plan. The master training plan should be reviewed and approved by all key individuals involved in its development and use.

Training plan design should include an audience and job needs analysis. The Society of Clinical Data Management (SCDM) task list and capabilities matrix provide a good starting point for this analysis. Both are available on the SCDM. This analysis should be customized to the organization's assignment of roles and responsibilities. Consider entry behavior skills, knowledge, and abilities (SKAs) when designing the proposed level of instruction. To determine the entry behavior, test a small sample of the target audience to establish if the starting points of the training program and threshold knowledge are accurately represented. The audience analysis should also consider various training delivery mechanisms to accommodate learning styles, gender and ethnic learning differences, as well as the learning stages.

Establish clear learning objectives. This is a critical step as learning objectives form the basis for what is to be taught, the performance level, and the conditions under which the task is to be performed. Include a peer review or beta-test of training materials to ensure that the stated objectives are valid and achievable.

Once learning objectives have been established, the training method should be evaluated. It should be determined whether the course is most suitable as an instructor-based class or an online course. A cost analysis of the preferred training method should also be performed to determine if online training is feasible. If considering an internal online training program, this cost analysis should include a review of the company's infrastructure and resources for maintenance of an internal training website. After implementation, the master training plan and materials should be assessed and updated based on feedback, changing audience and job requirements.

Data Management Training Topics

The following is a list of topics that affect the daily work environment of data managers. This is not intended to be an exhaustive list of training topics, but rather as a reference guide for the development of a master training plan and individual development plans.

- **Standard Operating Procedures and Departmental Policies**
  Data management departments are required to have SOPs that describe their processes and operations. All data management employees are required to understand and follow these SOPs. Having the employee sign stating an SOP has been read and understood often constitutes required SOP training. It is recommended that this practice, used in isolation, be avoided because it often falls short of meeting its intended purpose. An example of an approach that may be more effective is to have the trainer go over each required SOP with employees and explain how it affects their daily work flow. Such training sessions often encourages questions that reveal inconsistencies in the implementation of the SOP. Follow-up activities such as study audits may also reveal such inconsistencies. Issues identified may be addressed by revising the SOP, if applicable, and/or otherwise intervening to change the working practices of employees. As SOPs are revised, training must also be updated.

- **Computer Software and Technical Skills**
  Data Management utilizes various computer software applications to enter, clean and analyze data. These applications, including clinical databases, case report form (CRF) imaging software, edit specification development, discrepancy management, and others, require training for all employees who use them. Depending on time and budgetary restrictions, this training may be performed by the software vendor, a third party training vendor, or an in-house trainer.

- **Regulations and Industry Standards**
  Data Management is required to work within the constraints of Food and Drug Administration (FDA) codes and regulations. In addition, industry standards give employees guidance in their common work practices. Information regarding standards such as GCP, ICH Guidelines, FDA Regulations and Guidance Documents, as well as GCDMP can be found in various publications, educational seminars or Internet web sites. A trainer should make such references available to all employees.

- **Professional Growth**
  Individual development plans should include topics that focus on the employee's growth outside of the technical skills required. Skills, such as leadership training, effective team skills, time management, conflict resolution, project management, presentation skills, listening skills, cultural diversity, and medical terminology allow employees to cultivate their professional skills, directing them to be more productive in a group setting. Often, the Human Resources department can provide outside resources for conducting such classes. Online courses also offer various training opportunities in this area.

- **Interdepartmental Processes**
  To be fully effective, CDM employees must also understand the processes that occur before and after the data is handled in data management (such as site monitoring, safety monitoring, statistical analysis, FDA submissions, etc.). One effective approach is to observe other departmental processes firsthand during cross-training. Another is to invite personnel from other business units or departments to attend data management meetings as guest speakers.

Training Techniques and Environment

The following information describes different training techniques that may be used to optimize participant learning satisfaction. The importance of controlling the environment to enhance the learning process is also discussed. Additional information regarding these methods may be obtained through the reference citations at the end of this document.

Principles of Learning

The principles and techniques described in this section are based on the Interaction Associates, Inc. workshop[4]. A trainer should balance the three principles of service, respect and authenticity to establish an environment that is focused on the learner's needs. These three principles facilitate the development of a sense of trust between the trainer and participants. The trainer demonstrates service to the participants by being prepared when questions arise, even during break time. Service may also be exemplified by arriving prepared with innovative methods for teaching the topic. Mutual respect between the trainer and trainees must be established immediately. Creating a set of ground rules and expectations can facilitate an atmosphere of respect. Acknowledging and validating participant concerns and different learning styles also earn respect from the participants. Finally, being honest and genuine creates a presence of authenticity within the group.

Strategies in Learning

Different strategies may be employed to guide decisions and to steer the direction of a training session. Using the learning pathway enables trainees to pass through a logical sequence of steps to learn new skills.[4] The first of the five steps in the learning pathway is to provide the definition, or meaning, of the skill or task. Then, validate why it is important. Follow with assimilation or comprehension of how the skill or task works, and then integrate how it is used in the daily working environment. This will allow the trainees to transition or incorporate the task with relation to other skills or tasks that they perform in their job. A trainer can organize the teaching of any concept, skill or task through the learning pathway.

A trainer also needs to balance the importance of content, process, and relationship when presenting a topic.[4] To ensure participant satisfaction, the trainer must provide enough content to keep trainees interested, while covering the objectives and meeting the participants' expectations. However, if the training session is comprised of only content, the learning process will be compromised. The trainer needs to think about the process or flow of the training session as well. Try to include all participants in the session, monitor

the pace of delivery, and consider the timeliness of each step. The trainer also needs to establish a trusting relationship with participants. This promotes a comfort level for trainees and allows them to feel at ease to ask questions and participate in the class.

Presentation Delivery/Tools and Techniques

Learning is best achieved by receiving information through a variety of methods or techniques. This section describes several methods used to present classroom training materials. Methods often employed for on-the-job training are newsletters, fact sheets or quick tip reference guides. Special attention to mentor-based training should be given to ensure consistent delivery of information. The learner should be encouraged to seek clarification and validate information through multiple means rather than relying on a single source of information.

Lecture is the traditional method of transferring information from trainer to participant. However, research shows that using lecture alone for an extended period of time does not provide the optimum retention level of training materials. Lecture should be used in conjunction with other learning methods such as those described below. Lecture may be integrated with testing, allowing time for self-assessment, or with discussing surveys or training materials within the group.

Multi-sensory techniques (visual, auditory and kinesthetic) increase the acquisition of training material. Training that impacts as many of the human senses as possible accommodates different learning styles, speeds, and needs. Examples of visually stimulating training are the use of flip charts, colorful presentations, or other visualization techniques. Varying voice tone during presentations or playing music can stimulate the auditory senses during training. The kinesthetic sense of touch can be incorporated into training by using exercises with physical movement or objects.

Group discussion and interaction among participants is an effective way to present a common topic. Understanding and comprehension of the topic is enhanced when trainees discuss the topic with each other. Discussing a topic allows groups to establish a personal connection with the content and provides a common basis for shared ideas. Discussion triggers, such as short readings, role playing, videos, or open-ended questions, help to stimulate discussions by connecting the participants with each other and the topic.

The "Open, Narrow, Close" technique of questioning is one approach that allows the trainer to maintain control of group discussions.[4] First, open up the content of a discussion with a broad question. Then, focus the discussion on a specific area or subtopic that was mentioned. Follow by closing and transitioning the discussion to the next topic. Questions posed by trainees should be recognized by the trainer as a learning opportunity or "teachable moment." It is imperative for the trainer to understand the question being asked. This can be achieved by paraphrasing the question, providing parallel comments to the question, or asking for clarification or expansion of certain aspects of the question or comment.

Assignments, simulations, games or other activities are examples of active learning techniques. Research indicates that learning is enhanced by physically performing a related activity. Select the type of activity that best supports learning objectives. Activities might include, but are not limited to brainstorming, round-table discussions, role-playing, or practicing tasks in a test environment. Using a three-step learning cycle, known as a construction spiral[5], is another method to engage trainees in the learning activity. Providing a post-training activity that facilitates the utilization of the new skills in the daily work environment can also be an effective technique.

Storytelling allows trainees to relate the topic to their own daily environment. Stories may relate a similar experience, draw an analogy to the topic being discussed, or give an example of how the topic relates to the participants. However, it is important not to generalize or make assumptions about participants when sharing stories. The level of trainer/trainee trust must be kept intact.

Online Training

Due to time and logistical constraints, it is often necessary to provide online training materials for employees. Online training can consist of outside vendor courses performed via the Internet, as well as internally developed training. This type of training provides flexibility because the class may be taken at a convenient location, time and pace. Online training is also beneficial since travel time and expenses involved in bringing employees to a central location for training are avoided.

Online training from outside vendors should be evaluated for relevance, content, and cost, as well as the organization and accuracy of course materials. Products from different vendors should be compared to assess the most valuable and relative course.

Internal training may be performed online via a department website on the company intranet. Training materials, such as presentations, examples, case-studies, quizzes, glossaries, and tip sheets are easily referenced from the web. Links to other areas, such as forms, and other training reference materials may also be posted.

An internal training website should contain a main menu listing the courses available. An introduction for each subtopic and summary of the objectives for the course should also be provided. As with instructor-led training, it is important to measure the knowledge obtained from the course to ensure that the objectives are understood. It is also important to use the different training techniques discussed earlier to keep the student interested. Visual graphics with 'screen shots' are particularly helpful with online software applications training. With online training, it is imperative that an instructor or resource center be available for questions from the student. Online courses should be constructed to avoid boredom, which can lead to skipping sections or doing the minimum to advance to the next section. Worksheets, study guides, clear endpoints and rewards for course completion can assist with these issues. Providing an environment for users to practice is also beneficial.

When assessing online training, whether internal or external, certain navigational features should be considered. Forward, back and exit buttons should be available at each menu to facilitate when the student is moving from screen to screen. A help button should be provided at each step to assist the student in navigation, as well as course guidance. Bookmark and sound card options are also beneficial. The accessibility by those students with language barriers or disabilities should be evaluated when setting up online training. Internet services, such as Bobby Worldwide (www.cast.org/bobby/), are available to test for language, vision or hearing issues from an internal website. Once the website is constructed, it can be sent for accessibility testing regarding obstacles such as sound cards for the hearing impaired, pages that are hard to read or color-dependent for the visually impaired, etc.

Trainer Qualifications

A data management trainer should have experience in the topic of presentation, as well as experience in training techniques. The trainer must understand industry standards, as well as departmental policies. Training techniques and methods may be found in various publications, some of which are listed in this document. Also, many companies offer training courses, which can be found on the Internet.

A trainer must always be prepared to handle strategic or 'teachable moments'. These situations may include an upset participant, an irrelevant or long-winded statement that guides the participants in an unplanned direction, or a compelling comment. When the need for transition from the current situation to the next step in the training process is recognized, the trainer must choose the best route of reaching that next step. However, the principles of service, respect and authenticity, as previously discussed, must be maintained during this process so the trainer/trainee trust stays intact.

Training Environment

It is important to regulate the physical and mental climate during training. Start by ensuring that pre-work assignments are distributed well in advance of the training event and that expectations are clearly understood. During the session, temperature, room arrangement, lighting, and external noise should be kept at an optimal level. The climate and tone are frequently set at the beginning of a training session. Beginning with a fun activity, providing food and drinks, or playing music sets the tone for an optimistic training atmosphere. It is also important to close the training session on a positive note. Summarize the key contents covered during the class. Recognize the efforts of and the goals accomplished by each participant. Encourage participants to establish a plan to implement the topics discussed into their daily working environment.

Evaluation and Feedback Techniques

Implement a 360^{o} feedback process regarding all aspects of the training experience. Feedback should include comments about the trainer, the training materials and the training environment. It may be appropriate to perform testing at this time. Explain to the trainees how the feedback will be managed. Encourage trainees to contact the trainer after the training session if necessary.

Recommended Standard Operating Procedures

- Data Management Training Program and Documentation

References

1. International Conference on Harmonization (ICH) ICH Expert Working Group. *Guideline for Good Clinical Practice, E6,* Section 2.8. 1996.
2. Code of Federal Regulations (CFR), Title 21, Part 11, *Electronic Records; Electronic Signatures,* Section 11.10 (i). 1997.
3. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. U.S. Department of Health and Human Services, Food and Drug Administration; 1999.
4. Interaction Associates, Inc workshop. *The Masterful Trainer[sm]: Principles, Strategies and Tools*. Interaction Associates, Inc., San Francisco, CA and Cambridge, Ma is an organizational learning and consulting company; 1997.
5. Drummond, Tom. A Brief Summary of the Best Practices in College Teaching Intended to Challenge the Professional Development of All Teachers. *http://nsccux.sccd.ctc.edu/~eceprog/bstprac.html* 2001.
6. National Seminars workshop. *Trainer's Creativity Day Camp[sm]*. National Seminars Group is a division of Rockhurst University Continuing Education Center, Inc.

Other Related Reading

Clark, Donald. Instructional System Design - Design Phase. *http://www.nwlink.com/~donclark/hrd/sat3.html* 2000.

Kolb, D. *Experiential Learning*. Englewood Cliffs, NJ: Prentice-Hall; 1984.

Lawson, Karen. *The Trainer's Handbook* San Francisco, CA: Jossey-Bass/Pfeifer; 1998.

MacKeracher, D. *Making Sense of Adult Learning*. Toronto: Culture Concepts, Inc.; 1996.

Mager, Robert, F. *Preparing Instructional Objectives* (2nd ed.). Belmont, CA: Fearon; 1975.

Menges, Robert J. *Review of Higher Education.* Research on Teaching and Learning: The Relevant and the Redundant. 1975; 11: 259-268.

Wlodkowski, Raymond J. *Enhancing Adult Motivation To Learn – A Comprehensive Guide For Teaching All Adults.* San Francisco, CA: Jossey-Bass; 1999.

Trace R&D Center, University of Wisconsin - Madison *Web Content Accessibility Guidelines 1.0,* http://www.w3.org/TR/WCAG10/ ; 1999

Vendor Management

Introduction

Vendors are used in all aspects of clinical trials and have particular relevance in the Data Management process. Some examples of vendors employed in data management include CROs, pharmacies, printers, laboratories, software suppliers, device companies and off-site storage. In discussing CROs, ICH E6 states that the "Ultimate responsibility for the quality and integrity of the trial data always resides with the sponsor." This statement should be extrapolated to include all vendors that have an impact on the final quality of data. For this reason it is important to ensure that quality data is received from vendors and that a documented process exists to consistently evaluate vendor services.

Scope

Vendor relationships differ widely across the industry including everything from printing of CRFs to medical oversight, to regulatory submission. This section examines communication of clear expectations between vendor and sponsor and sponsor involvement, including vendor qualification and the level of oversight appropriate for the vendor's scope of work.

The sponsor is responsible for the quality and integrity of the trial data.[1] Therefore, the sponsor must manage the vendor in a fashion that ensures quality and integrity. Additional detail and discussion regarding relationship management is beyond the scope of this section.

Minimum Standards

- Determine impact of vendor services on the outcome of the trial.
- Document process and support or education necessary to utilize the vendor.
- Clearly define expectations, deliverables and responsibilities. Both the sponsor and the vendor must participate in establishing definitions.[1]
- Conduct ongoing management of vendor activities, assuring regular communication and auditing throughout the clinical trial.

Best Practices

- Pre-certify vendors evaluating the quality system.
- Establish vendor auditing program before the trial begins.
- Evaluate vendors prior to contracting services.
- Oversee vendor use and management using a centralized organizational team.

Vendor Evaluation

Before a study begins and before a vendor is selected, an evaluation of the vendor should take place. The sponsor must understand specifically what the services are that will be provided and whether the data in the clinical trial could be directly or indirectly affected by these services.

The primary goals of a vendor evaluation are to determine that there is adequate quality, capable and appropriate resources available to provide the defined services. It is not necessary to conclude that a vendor cannot be utilized if found to be deficient. However, the support and management that will be required of the vendor should be defined. This support may take the form of SOP provisions, data management support, QA advice, documentation or system validation guidance. In many instances, the vendors must be educated to become invested in the quality of the data or service they are supplying.

Evaluation of a vendor must be scaled according to the impact of the vendor service on the outcome of a trial. CROs providing full service would require extensive evaluation, whereas the printer who does query binders may require little evaluation. Since the qualification process can be time consuming, it would be advantageous from a business perspective to pre-qualify vendors.

Considerations when evaluating a vendor can include the following:

- Evaluation of the vendor's QC/QA process
- The vendor's SOPs and proof of adherence
- Qualifications of the vendor's staff and their knowledge of SOPs
- Documentation of the vendor's change control processes
- Evaluation of sub-contractor if applicable
- Review of required accreditation in the vendor's field of work (e.g. lab certifications)
- Financial stability of the vendor
- Availability of documentation to regulatory authorities.

Vendor Management

It is important to detail specifically the vendor's responsibilities and to clearly define all deliverables in the contract. The vendor should be audited frequently enough to ensure compliance with agreed-upon processes and deliverables. A sponsor's quality management plan should be expanded to include quality management information from the vendor. It should be comprehensive enough to include processes and product expectations.

Continuing management of the vendor must occur at regular intervals throughout the trial. Milestones should be developed in advance and specifically defined deliverables agreed upon throughout the study. These milestones can be monitored through regular communication and status updates that are documented and provided to the sponsor. Out-of-scope or contractual issues should be discussed and handled, as they occur, in order to ensure that misunderstandings or problems do not come as a surprise at a later time.

Recommended Standard Operating Procedures

- Vendor Management, to include vendor qualifications, specifications for outsourced deliverables, ongoing management.
- Vendor Auditing with specific audit procedures for significant vendors (e.g. CRO auditing, Laboratory auditing.)

References

1. International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Section 5.2.2 and 5.2.3; 1996

Other Related Reading

N/A

Metrics for Clinical Trials

Introduction

The term "metric" simply refers to a measurement. In clinical research, metrics can quantitatively and qualitatively assess whether or not a process is efficient and effective, as well as whether the product being measured has or will have the expected quality. Metrics can be used at intervals throughout a process to ascertain if the process is working (i.e., a 'living baseline'). When a process has been completed, well-designed metrics can help indicate whether a goal was achieved with the expected quality produced.

This section provides information regarding metrics that are particularly relevant to Clinical Data Management (CDM). Best practices are written with respect to metrics identification and definition. Sample metrics are provided and considerations as to when metrics should be used to evaluate new processes and new systems (e.g., implementation of electronic data capture (EDC)) are discussed.

Metrics are not requirements nor are they the basis for regulations; they are considered best practices in the sense of measuring quality and efficiency in many industries. This section is written to provide helpful suggestions and considerations for those establishing a metrics program within a department or company. The content of this section should, therefore, not be considered for inclusion in regulatory auditing procedures.

Scope

Although metrics may be dependent upon other functional groups or departments, the scope of this GCDMP section concentrates on metrics that are relevant to the Clinical Data Manager's role in the planning and conduct of clinical trials.

Minimum Standards

- Evaluate the company's need to implement a metrics program to measure the deliverables and performance of processes that are the primary responsibility of CDM.

Best Practices

- Identify the metrics to be used to measure performances against goals and objectives.
- Ensure alignment of metrics with goals and contractual agreements (where appropriate.)
- Consider the aspects of cost, quantity, quality and timeliness when deciding what metrics to implement.
- Standardize the definitions for metrics by using consistent terminology and parameters.

- Ensure all stakeholders (e.g. project managers, contractors, clinicians, data managers, management) understand and are in agreement with the definition of what is being measured and how it is being measured by a metric.
- Agree upon well-defined metrics at the onset of project and use those metrics to evaluate the performance of the project during all stages.
- Ensure that the effort needed to collect and report a metric is appropriately offset by the benefit.
- Select a set of key metrics that apply to all trials and projects. Use these metrics as the basis for comparison of process performance across all trials and projects.
- Determine if additional metrics are necessary based on changes in technology or processes from other trials and projects.
- Strive to use existing primary data (e.g., audit trails, tracking systems) to collect metrics.
- Minimize the re-entry of data to improve timeliness and accuracy of metrics data.
- Document the process for collecting, reporting and communicating study-specific metrics into the Data Management Plan.
- Document corrective action to be taken if metrics demonstrate that goals and objectives are not being achieved.

Identification of Metrics: Linking Metrics with Goals

Metrics can be very useful to assess processes used in clinical trials, especially within clinical data management. Clinical trials are often evaluated within the realm of strategic (i.e., corporate) and tactical (i.e., operational) objectives. Metric assessments should be based on at least two of four core criteria (quantity, cost, time and quality).

Quantity, cost, time and quality are linked together such that measuring each of them is important but focusing too much on one may adversely affect another. Therefore, overall project goals and objectives must be considered when selecting and evaluating metrics. Metrics that only address some of the four core criteria, but not all, will not provide a complete picture of performance. If one metric is met, it does not follow that the others are achieved. For instance, the overall calendar-based timeline may be attained but at the expense of excess hours having been spent because inefficient processes were used to meet the timeline.

There is a hierarchical relationship between the objectives of the company, department, and individual project or clinical trial. A company may have strategic objectives of achieving a certain level of quality in its product while achieving a particular profit margin at the end of the fiscal year. Each functional group, such as CDM, sets tactical goals and objectives for developing processes to ensure quality, while using resources efficiently in an effort to achieve profitability. A particular project manager or project team may have budget and time constraints, yet be expected to deliver a quality product. Each functional group must develop its objectives and its metrics within the context of

the corporate objectives. The existence of these hierarchical objectives and concurrent timelines drives the need for consistency in the definition and utilization of metrics. This consistency should apply not only across the functional groups but within the corporate hierarchy as well.

Regardless of what is being measured, or for what purpose, well-designed metrics[1] should be:

- relevant (answer critical business questions)
- enduring (of lasting relevance)
- robust (not subject to manipulation or variation due to process changes)
- valid (measure what it purports to measure)
- specific (clear and consistent)
- actionable (can drive decisions)
- practical (measurements can be made in a timely fashion without a significant drain on resources, and the necessary metric or metric parameters are available)

The effort needed to collect and report the metric should be offset by the benefit of the metric. Cost and quality metrics are often very challenging, whereas metrics dealing with cycle times are typically much easier to collect. The metrics that are collected and reported must be able to answer the questions that have been pre-defined to measure the success or failure of a project or process. The level of difficulty in the collection of metrics should be considered in respect to the significance of the value provided from the metric. Conversely, just because a metric is easy to collect and does not require an outlay of resources does not mean it should be collected if it has no benefit.

Standardizing Definitions for Metrics

Since metrics are shared between functional groups or partners, it is important that those metrics are based on standard definitions. The need for standardization of definitions is amplified if metrics are used for comparison purposes across trials or across companies (e.g., benchmarking projects). Communication between the various groups that are using the metric is enhanced by the use of standard definitions. For example, "time to database lock" is probably the most frequently cited metric for clinical trials. In the current culture, there is much emphasis on reducing this cycle time metric, either in order to expedite timely submission of the data to regulatory authorities or to facilitate strategic decisions on future studies within the drug development plan. This metric may have different meanings for different CDM groups and companies within the industry. For example, completion of database lock may be considered to occur:

- After a QA audit is acceptable and it is deemed permissible to break the blind
- When the data are 'frozen' and a sponsor accepts the data that have been transferred from their CRO (e.g., the database or transferred SAS datasets)
- Multiple times, depending upon SOPs and whether or not a company allows for database 'unlocking' to make changes to the database after it is originally locked

Likewise, the starting point for this cycle time metric can begin at any of the following points:

- When the last subject completes the last visit
- When the last CRF is received by the group performing data entry
- When the last data from the last visit are entered into a CRF or an EDC application
- When the last query or discrepancy is resolved
- When the last CRF data are deemed 'clean'

Due to the various interpretations of the term "database lock", all parties could potentially be working in different directions based on their presumption of when database lock occurs and what activities take place at that point. Without a standard definition of this metric, the goal may never be identified or achieved.

It is necessary for all functions that are affected by the goal being measured to be involved in the definition process for the metric. If the starting point for "time to database lock" is the date that the last patient completes the last visit, the monitoring group should work with the CDM group to develop and agree upon the definitions and the process used to achieve this cycle time. As for the end point, if it is defined as the point that the blind is broken, appropriate representatives should work together to understand their respective roles in this process. The Data Management Plan (or the applicable documentation including the metrics) should be kept current and reflect any decisions made regarding metrics to be collected and their definitions.

Key Metrics

Many companies and departments are striving to identify a set of metrics to use across all projects. Metrics should be based on goals and objectives set by a group or organization. It is often difficult to recommend specific metrics that would fit the needs of all parties involved. Most goals and objectives set by groups or organizations revolve around the interdependent areas of time, quality, cost, and quantity.

Agreement on a set of key and relevant metrics will facilitate achievement of pre-determined goals. Key metrics are collected on every clinical trial and used for comparison across trials. Although agreement on certain metrics are obtained by the overall company or department, there may still be the need for individual departments or project teams to maintain additional metrics (i.e., to assess their respective departmental or team efficiencies).

Metrics for Projects Employing New Processes or Technology

When considering a set of key metrics, it is important to design the metric to allow for tracking and comparisons across projects, regardless of the process or technology used in the clinical trial. Not only does this allow for an assessment of each project in comparison to similar projects, but it also allows for an evaluation of processes that may be redesigned to take advantage of new technology.

An example of a metric that is process and technologically independent is to measure the number of queries per the data fields for incoming data (as opposed to the number of queries per page or per patient.) Another useful metric is to measure the cycle time from

subject visit to when the data accessible in-house. These metrics are applicable to both paper-based clinical trials and to those using EDC or eClinical processes.

Collection and Reporting of Metrics

The context in which a metric is applied should be determined prior to the reporting of the metric. The metric's data source(s), data extraction date or reporting window should be included with each report. Additionally, each metric should be grouped according to its attributes. An attribute is a characteristic of a metric that helps the stakeholder understand the underlying causes for performance variances. The following list of attributes are suggested: drug program, therapeutic area, indication, study phase, data collection mode (e.g. EDC, paper, imaging), study design, size or complexity factor (e.g. number of sites, number of subjects, number of procedures), or resourcing model, (e.g. CRO, contractors, in-house staff, etc.) Invalid comparisons will be avoided by proper categorization and summarization of metrics according to their attributes.

Data Management Metrics

Quantity

In data management, the goal is to produce a database that is suitable for analysis. Common metrics reported by data management are "number of queries generated per patient", or "number of data entry errors found in 10,000 fields", etc. While these metrics may help diagnose problems, they do not measure whether or not the resulting database is analyzable. Instead, metrics should help to examine the amount of time and resources required to produce an analyzable database. This metric will assess if data managers are producing quality data in the timeframe given for the study. This metric can also be compared against goals and industry trends (benchmarking).

Cost

Cost can be measured by estimating the percent of full time employees (FTE) assigned to a project versus collecting all of the hours that were actually spent on that project. For this metric to be useful, it requires that one also be able to make the same measurements on all projects, including previous projects when possible, in order to calculate a baseline for comparisons. It may also be useful to track the hours for a certain timeframe for a project, as opposed to all project team members tracking all hours charged for a given process or function.

Time
The most relevant issue to consider when measuring cycle time is defining the exact start and stop points to be measured (as was discussed previously for 'database lock'.) It is also important to determine the value of timelines. For example, should timelines compromise quality? Alternatively, should more resources be provided to meet the timelines for that project?

Quality
In terms of quality, both process and product quality may be measured. Internal and external customer satisfaction, trends in data integrity measurements, types of queries and actions taken with queries are most commonly used. Quality of clinical data may also be assessed through measurements such as evaluable subjects or subject visits and protocol violations.

Appendix C provides a list of potential metrics and definitions to consider as a starting point for establishing a metrics program.

Timeliness and Accuracy of Metrics
Metrics should be available for review in a timely manner to provide maximum benefit. For example, it is valuable for project managers and department managers to gather status information for an ongoing trial. This information may include enrollment rates or the numbers or types of queries occurring on the CRF data. The greatest opportunity to take corrective action occurs when information is made available in 'real-time'. The earlier a problem is detected, the sooner it can be addressed.

Metrics are most timely if they are collected electronically with no re-entry required. In paper-based clinical trials, once the data are entered into the computer, data management metrics can be generated electronically and as frequently as necessary. Information regarding subject enrollment, visit completion, and other such indicators of trial status can be difficult to obtain in a timely fashion in paper-based trials. Teams often rely on the reporting of this information by each site and then subsequent entry of this information into a project management or project-tracking database. When more than one database contains the same information for a metric, it is important to determine the primary source to avoid to the need for reconciliation across databases. In clinical trials using EDC systems, the clinical trial data and queries are usually available in an electronic form earlier than in paper-based clinical trials. Subject enrollment, visit completion, and other useful metric information may also be available from the EDC tool.

Using Metrics from Various Sources
It can be difficult to obtain metrics when the parameters required for measurement are found in multiple databases. This is further compounded when certain complementary metrics, such as the project budget and the status of the data management processes, are rarely available for analogous time frames. As discussed in the previous paragraph, metrics can be synchronized with other relevant information if they are collected in a

timely manner. In addition, the more data that can be 'shared' electronically across systems, the lower the chance is of incurring error through re-entry of the data parameters that are needed for measurement. The use of new technologies, such as web portals, clinical dashboards, and visualization tools is now a viable option for reviewing metrics data. These tools can eliminate the need to actually integrate databases or re-enter data. They can also allow for views of complementary data within the same time frame.

Providing for the Users of Metrics

Metrics should be shared across all partners participating in a clinical trial, including CROs and vendors, when appropriate. When partners are involved, it should be determined early in the project planning stages which metrics will be collected, by whom, how and when they will be disseminated (e.g., with a common website or visualization tool one month after first patient signs consent). Communicating metrics results will allow all parties involved to be able to make corrective actions as soon as possible.

Action Plans: The Feedback Loop

The desired outcome or product from the use of metrics is obtained through a well-planned and executed process that includes interim assessments and feedback loops. Design the procedures needed to collect the pre-determined metrics that are used for assessing whether the goal has been reached. Determine specific intervals or time frames for when these interim and final assessments will be performed.

Metric reports can assist in the interim and final assessment of the overall project. Again, value is gained by categorizing metrics by their attributes to assist in comparison of the project's outcome to other projects in or outside the company. These reports should be run at agreed upon times throughout and at the end of the project. Reports should summarize the metrics that were collected. Such reports should also include an assessment of results against goals or objectives. Commentary around the results should also be provided. This commentary should include reasons for positive performance results and corrective action plans for improving results.

Interim indicators or flags can also be used for ongoing assessments of the process. Useful reports for the analysis of the metrics include trend analyses, statistical techniques, flagging of outliers, identifying unanticipated trends in the data, plots showing incoming data and analogous query rates, and listings of values such as changes from the baseline or summary tables[2].

Metrics and the Data Management Plan

The Data Management Plan is a tool that can be used to reflect the decisions that are made regarding the use of metrics on the project (e.g., the definitions of metrics, the means for collecting the metrics or communication related to the metrics). The metrics used for the project should be defined at the project planning and initiation stages. All project assumptions and assertions for establishing the metrics should also be documented. All key metrics reports or other documents relevant across projects should be referenced in the project documentation. Additionally, if there are new terms used or

new partners or contractors involved, it may be helpful to establish and maintain a "project dictionary or glossary".

Recommended Standard Operating Procedures

Typically, an SOP would not be written mandating a metric or the use of metrics. SOPs or guidelines that describe a particular standard or organizational objective should include all of the metric information (e.g., the metric(s) to be used, the data collected to calculate the metric, and the anticipated action plan resulting from the metrics analysis.)

References

1. Pitre, Armelde Definition of 'Metric', PhRMA Presentation, 2001

2. Walton, M. *The Deming Management Method*. New York: The Berkley Publishing Group; 1986

Other related reading

Tufte, E R. *The Visual Display of Quantitative Information*. 15th printing 1997. Connecticut: Graphics Press; 1983

Tufte, E R. *Envisioning Information*. 5th printing 1995. Connecticut: Graphics Press; 1990

Meyer, C. *Fast Cycle Time*. New York: The Free Press; 1993

Kush, R., et al., *eClinical Trials: Planning and Implementation,* Boston: CenterWatch; 2002

Glossary

The Good Clinical Data Management Practices Document adopts the ICH definitions for terms defined within the ICH guidelines. Unless otherwise noted, these definitions were taken from ICH E6.

Adverse Drug Reaction (ADR) In the pre-approval clinical experience with a new medicinal product or its new usage, particularly as the therapeutic dose(s) may not be established: all noxious and unintended responses to a medicinal product related to any dose should be considered adverse drug reactions. The phrase responses to a medicinal product means that a causal relationship between a medicinal product and an adverse event is at least a reasonable possibility, i.e. the relationship cannot be ruled out. Regarding marketed medicinal products: a response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of diseases or for modification of physiological function (see the ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting).

Adverse Event (AE) Any untoward medical occurrence in a subject or clinical investigation subject administered a pharmaceutical product and which does not necessarily have a causal relationship with this treatment. An adverse event (AE) can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product (see the ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting).

Amendment (to the protocol) See Protocol Amendment.

Applicable Regulatory Requirement(s) Any law(s) and regulation(s) addressing the conduct of clinical trials of investigational products.

Approval (in relation to Institutional Review Boards) The affirmative decision of the IRB that the clinical trial has been reviewed and may be conducted at the institution site within the constraints set forth by the IRB, the institution, Good Clinical Practice (GCP), and the applicable regulatory requirements.

Audit A systematic and independent examination of trial related activities and documents to determine whether the evaluated trial related activities were conducted, and the data were recorded, analyzed and accurately reported according to the protocol, sponsor's standard operating procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s).

Audit Certificate A declaration of confirmation by the auditor that an audit has taken place.

Audit Report A written evaluation by the sponsor's auditor of the results of the audit.

Audit Trail Documentation that allows reconstruction of the course of events.

Blinding/Masking A procedure in which one or more parties to the trial is kept unaware of the treatment assignment(s). Single-blinding usually refers to the subject(s) being unaware, and double-blinding usually refers to the subject(s), investigator(s), monitor, and, in some cases, data analyst(s) being unaware of the treatment assignment(s).

Case Report Form (CRF) A printed, optical, or electronic document designed to record all of the protocol-required information to be reported to the sponsor on each trial subject.

CDISC Clinical Data Interchange Standards Consortium.

Checklist (ASQ) A tool used to ensure that all important steps or actions in an operation have been taken. Checklists contain items that are important or relevant to an issue or situation. Checklists are often confused with check sheets and data sheets (see individual entries).

Clinical Trial/Study Any investigation in human subjects intended to discover or verify the clinical, pharmacological and/or other pharmacodynamic effects of an investigational product(s), and/or to identify any adverse reactions to an investigational product(s), and/or to study absorption, distribution, metabolism, and excretion of an investigational product(s) with the object of ascertaining its safety and/or efficacy. The terms clinical trial and clinical study are synonymous.

Clinical Trial/Study Report A written description of a trial/study of any therapeutic, prophylactic, or diagnostic agent conducted in human subjects, in which the clinical and statistical description, presentations, and analyses are fully integrated into a single report (see the ICH Guideline for Structure and Content of Clinical Study Reports).

Common causes (ASQ) Causes of variation that are inherent in a process over time. They affect every outcome of the process and everyone working in the process (see also "special causes").

Comparator (Product) An investigational or marketed product (i.e., active control), or placebo, used as a reference in a clinical trial.

Compliance (in relation to trials) Adherence to all the trial-related requirements, Good Clinical Practice (GCP) requirements, and the applicable regulatory requirements.

Confidentiality Prevention of disclosure, to other than authorized individuals, of a sponsor's proprietary information or of a subject's identity.

Conformance (ASQ) An affirmative indication or judgment that a product or service has met the requirements of a relevant specification, contract, or regulation.

Contract A written, dated, and signed agreement between two or more involved parties that sets out any arrangements on delegation and distribution of tasks and obligations and, if appropriate, on financial matters. The protocol may serve as the basis of a contract.

Coordinating Committee A committee that a sponsor may organize to coordinate the conduct of a multi-center trial.

Coordinating Investigator An investigator assigned the responsibility for the coordination of investigators at different centers participating in a multi-center trial.

Contract Research Organization (CRO) A person or an organization (commercial, academic, or other) contracted by the sponsor to perform one or more of a sponsor's trial-related duties and functions.

Control Chart (ASQ) A chart with upper and lower control limits on which values of some statistical measure for a series of samples or subgroups are plotted. The chart frequently shows a central line to help detect a trend of plotted values toward either control limit.

Corrective Action (CA) (ASQ) the implementation of solutions resulting in the reduction or elimination of an identified problem.

Database Lock A database is locked when an organization's pre-specified database closure procedures have been completed or otherwise approved. At the time of lock, to the best of the sponsor's knowledge, the data is complete, meets pre-specified acceptance criterion, and is acceptable for analysis. Access granted to database users has been restricted to "read-only."

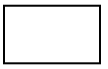
Direct Access Permission to examine, analyze, verify, and reproduce any records and reports that are important to evaluation of a clinical trial. Any party (e.g., domestic and foreign regulatory authorities, sponsor's monitors and auditors) with direct access should take all reasonable precautions within the constraints of the applicable regulatory requirement(s) to maintain the confidentiality of subjects' identities and sponsor’s proprietary information.

Documentation All records, in any form (including, but not limited to, written, electronic, magnetic, and optical records, and scans, x-rays, and electrocardiograms) that describe or record the methods, conduct, and/or results of a trial, the factors affecting a trial, and the actions taken.

e-CRF An auditable electronic record designed to record information required by the clinical trial protocol to be reported to the sponsor on each trial subject.

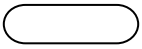
Essential Documents Documents which individually and collectively permit evaluation of the conduct of a study and the quality of the data produced (see 8. Essential Documents for the Conduct of a Clinical Trial).

Flow diagram, flow chart (Juran's Quality Control Handbook) A graphic means for depicting the steps or activities that constitute a process. The flow diagram [flow chart] is constructed from standard symbols. [The delay and database symbols have been added to Juran's List.]

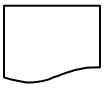


The activity symbol is a rectangle which designates an activity. Within the rectangle is a brief description of that activity.

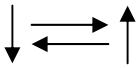
The decision symbol is a diamond which designates a decision point from which the process branches into two or more paths. The path taken depends on the answer to the question which appears within the diamond. Each path is labeled to correspond to an answer to the question.



The terminal symbol is a rounded rectangle which unambiguously identifies the beginning or end of a process. "Start" or "begin" is used o designate the starting point of a process flow; "stop" or "end" is used to designate the end of process flow.



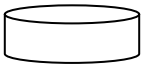
The document symbol is a document pertinent to the process.



The flow line represents a process path which connects process elements. The arrowhead indicates the direction of the flow.

The connector is a circle which is used to indicate a continuation of the flow diagram.

The delay symbol is a rectangle rounded on one side which identifies a waiting point or delay in process flow.



The database symbol is a cylinder which represents a database application and the contained data.

Good Clinical Practice (GCP) A standard for the design, conduct, performance, monitoring, auditing, recording, analyses, and reporting of clinical trials that provides assurance that the data and reported results are credible and accurate, and that the rights, integrity, and confidentiality of trial subjects are protected.

Hard Coding Computer programs utilize logic and machines to allow dynamic responses based on user input. For example, a website calculates the total when books are selected on line, or tabulates the average weight of the patients in the active treatment arm each time you run the program on a dataset. Hard coding is the limiting of the dynamic response by actually typing the data IN the computer program itself, instead of letting the data come from the dataset or the user. This can be dangerous, because this is not visible in the analysis tables and listings, to the regulatory authorities, and is easily forgotten once typed into the computer program.

Independent Data-Monitoring Committee (IDMC) (Data and Safety Monitoring Board, Monitoring Committee, Data Monitoring Committee) An independent data-monitoring committee that may be established by the sponsor to assess at intervals the progress of a clinical trial, the safety data, and the critical efficacy endpoints, and to recommend to the sponsor whether to continue, modify, or stop a trial.

Investigational New Drug application (IND) The IND regulations can be found at the following link: www.fda.gov/cber/ind/ind.htm "IND" is synonymous with "Notice of Claimed Investigational Exemption for a New Drug". An IND is submitted to the FDA when a sponsor or investigator wishes to initiate trials in humans.

Impartial Witness A person, who is independent of the trial, who cannot be unfairly influenced by people involved with the trial, who attends the informed consent process if the subject or the subject's legally acceptable representative cannot read, and who reads the informed consent form and any other written information supplied to the subject.

Independent Ethics Committee (IEC) An independent body (a review board or a committee, institutional, regional, national, or supranational), constituted of medical professionals and non-medical members, whose responsibility it is to ensure the protection of the rights, safety and well-being of human subjects involved in a trial and to provide public assurance of that protection, by, among other things, reviewing and approving / providing favorable opinion on, the trial protocol, the suitability of the investigator(s), facilities, and the methods and material to be used in obtaining and documenting informed consent of the trial subjects. The legal status, composition, function, operations and regulatory requirements pertaining to Independent Ethics Committees may differ among countries, but should allow the Independent Ethics Committee to act in agreement with GCP as described in this guideline.

In-control process (ASQ) A process in which the statistical measure being evaluated is in a state of statistical control, i.e., the variations among the observed sampling results can be attributed to a constant system of chance causes (see also "out-of-control process").

Informed Consent A process by which a subject voluntarily confirms his or her willingness to participate in a particular trial, after having been informed of all aspects of the trial that are relevant to the subject's decision to participate. Informed consent is documented by means of a written, signed and dated informed consent form.

Inspection 1) (ICH)The act by a regulatory authority(ies) of conducting an official review of documents, facilities, records, and any other resources that are deemed by the authority(ies) to be related to the clinical trial and that may be located at the site of the trial, at the sponsor's and/or contract research organization's (CRO's) facilities, or at other establishments deemed appropriate by the regulatory authority(ies).
2) (ASQ) measuring, examining, testing, and gauging one or more characteristics of a product or service and comparing the results with specified requirements to determine whether conformity is achieved for each characteristic.

Institution (medical) Any public or private entity or agency or medical or dental facility where clinical trials are conducted.

Institutional Review Board (IRB) An independent body constituted of medical, scientific, and non-scientific members, whose responsibility is to ensure the protection of the rights, safety and well-being of human subjects involved in a trial by, among other things, reviewing, approving, and providing continuing review of trial protocol and amendments and of the methods and material to be used in obtaining and documenting informed consent of the trial subjects.

Interim Clinical Trial/Study Report A report of intermediate results and their evaluation based on analyses performed during the course of a trial.

Investigational Product A pharmaceutical form of an active ingredient or placebo being tested or used as a reference in a clinical trial, including a product with a marketing authorization when used or assembled (formulated or packaged) in a way different from the approved form, or when used for an unapproved indication, or when used to gain further information about an approved use.

Investigator A person responsible for the conduct of the clinical trial at a trial site. If a trial is conducted by a team of individuals at a trial site, the investigator is the responsible leader of the team and may be called the principal investigator. See also Subinvestigator.

Investigator / Institution An expression meaning "the investigator and/or institution, where required by the applicable regulatory requirements".

Investigator's Brochure A compilation of the clinical and non-clinical data on the investigational product(s) which is relevant to the study of the investigational product(s) in human subjects (see 7. Investigator's Brochure).

ISO (ASQ) International Organization for Standardization

ISO 9000 series standards (ASQ) A set of five individual but related international standards on quality management and quality assurance developed to help companies effectively document the quality system elements to be implemented to maintain an efficient quality system. The standards, initially published in 1987, are not specific to any particular industry, product, or service. The standards were developed by the International Organization for standardization (ISO), a specialized international agency for standardization composed of the national standards bodies of 91 countries.

Legally Acceptable Representative An individual or juridical or other body authorized under applicable law to consent, on behalf of a prospective subject, to the subject's participation in the clinical trial.

Monitoring The act of overseeing the progress of a clinical trial, and of ensuring that it is conducted, recorded, and reported in accordance with the protocol, Standard Operating Procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s).

Monitoring Report A written report from the monitor to the sponsor after each site visit and/or other trial-related communication according to the sponsor's SOPs.

Multicenter Trial A clinical trial conducted according to a single protocol but at more than one site, and therefore, carried out by more than one investigator.

New Drug Application (NDA) The documentation submitted to the Food and Drug Administration, FDA.

> *"The goals of the NDA are to provide enough information to permit FDA reviewer to reach the following key decisions: Whether the drug is safe and effective in its proposed use(s), and whether the benefits of the drug outweigh the risks. Whether the drug's proposed labeling (package insert) is appropriate, and what it should contain. Whether the methods used in manufacturing the drug and the controls used to maintain the drug's quality are adequate to preserve the drug's identity, strength, quality, and purity.*

The documentation required in an NDA is supposed to tell the drug's whole story, including what happened during the clinical tests, what the ingredients of the drug are, the results of the animal studies, how the drug behaves in the body, and how it is manufactured, processed and packaged."(1 FDA) The NDA regulations are 21CFR314.

Non-clinical Study Biomedical studies not performed on human subjects.

Opinion (in relation to Independent Ethics Committee) The judgement and/or the advice provided by an Independent Ethics Committee (IEC).

Out-of-control process (ASQ) A process in which the statistical measure being evaluated is not in a state of statistical control, i.e., the variations among the observed sampling results can be attributed to a constant system of chance causes (see also "in-control process").

Original Medical Record See Source Documents.

Protocol A document that describes the objective(s), design, methodology, statistical considerations, and organization of a trial. The protocol usually also gives the background and rationale for the trial, but these could be provided in other protocol referenced documents. Throughout the ICH GCP Guideline the term protocol refers to protocol and protocol amendments.

Protocol Amendment A written description of a change(s) to or formal clarification of a protocol.

Quality Assurance (QA) All those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with Good Clinical Practice (GCP) and the applicable regulatory requirement(s).

Quality Control (QC) The operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled.

Quality Assurance/Quality Control (ASQ) two terms that have many interpretations because of the multiple definitions for the words "assurance" and "control." For example, "assurance" can mean the act of giving confidence, the state of being certain, or the act of making certain; "control" can mean an evaluation to indicate needed corrective responses, the act of guiding, or the state of a process in which the variability is attributable to a constant system of chance causes. (For a detailed discussion on the multiple definitions, see aNSI/ISO/aSQC a3534-2, Statistics--Vocabulary and Symbols--Statistical Quality Control.) One definition of quality assurance is: all the planned and systematic activities implemented within the quality system that can be demonstrated to provide confidence that a product or service will fulfill requirements for quality. One definition for quality control is: the operational techniques and activities used to fulfill requirements for quality. Often, however, "quality assurance" and "quality control" are used interchangeably, referring to the actions performed to ensure the quality of a product, service, or process.

Quality audit (ASQ) A systematic, independent examination and review to determine whether quality activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve the objectives.

Random Sampling (ASQ) A commonly used sampling technique in which sample units are selected in such a manner that all combinations of n units under consideration have an equal chance of being selected as the sample.

Randomization The process of assigning trial subjects to treatment or control groups using an element of chance to determine the assignments in order to reduce bias.

Regulatory Authorities Bodies having the power to regulate. In the ICH GCP guideline the expression Regulatory Authorities includes the authorities that review submitted clinical data and those that conduct inspections (see 1.29). These bodies are sometimes referred to as competent authorities.

Research Misconduct Falsification of data in proposing, designing, performing, recording, supervising or reviewing research or in reporting research results. Falsification includes acts of omission and commission. Deliberate noncompliance with

the regulations can be considered misconduct but is secondary to falsification of data. Research misconduct does not include honest error or differences of opinion.*

* Woolen, Stan, Office for Good Clinical Practice Web Site, FDA. October 2001.

Serious Adverse Event (SAE) or Serious Adverse Drug Reaction (Serious ADR) Any untoward medical occurrence that at any dose:

- Results in death,
- Is life-threatening,
- Requires in subject hospitalization or prolongation of existing hospitalization,
- Results in persistent or significant disability/incapacity,

  or
- Is a congenital anomaly/birth defect

(See the ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting).

Source Data All information in original records and certified copies of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial. Source data are contained in source documents (original records or certified copies).

Source Documents Original documents, data, and records (e.g., hospital records, clinical and office charts, laboratory notes, memoranda, subjects' diaries or evaluation checklists, pharmacy dispensing records, recorded data from automated instruments, copies or transcriptions certified after verification as being accurate copies, microfiches, photographic negatives, microfilm or magnetic media, x-rays, subject files, and records kept at the pharmacy, at the laboratories and at medico-technical departments involved in the clinical trial).

Special Causes (ASQ) Causes of variation that arise because of special circumstances. They are not an inherent part of a process. Special causes are also referred to as assignable causes (see also "common causes").

Specification (ASQ) A document that states the requirements to which a given product or service must conform.

Sponsor An individual, company, institution, or organization which takes responsibility for the initiation, management, and/or financing of a clinical trial.

Sponsor-Investigator An individual who both initiates and conducts, alone or with others, a clinical trial, and under whose immediate direction the investigational product is administered to, dispensed to, or used by a subject. The term does not include any person other than an individual (e.g., it does not include a corporation or an agency).
The obligations of a sponsor-investigator include both those of a sponsor and those of an investigator.

Standard Operating Procedures (SOPs) Detailed, written instructions to achieve uniformity of the performance of a specific function.

Statistical process control (SPC) (ASQ) The application of statistical techniques to control a process. Often the term "statistical quality control" is used interchangeably with "statistical process control."

Statistical quality control (SQC) (ASQ) The application of statistical techniques to control quality. Often the term "statistical process control" is used interchangeably with "statistical quality control," although statistical quality control includes acceptance sampling as well as statistical process control.

Subinvestigator Any individual member of the clinical trial team designated and supervised by the investigator at a trial site to perform critical trial-related procedures and/or to make important trial-related decisions (e.g., associates, residents, research fellows). See also Investigator.

Subject/Trial Subject An individual who participates in a clinical trial, either as a recipient of the investigational product(s) or as a control.

Subject Identification Code A unique identifier assigned by the investigator to each trial subject to protect the subject's identity and used in lieu of the subject's name when the investigator reports adverse events and/or other trial related data

Trial Site The location(s) where trial-related activities are actually conducted.

Type I error (ASQ) An incorrect decision to reject something (such as a statistical hypothesis or a lot of products) when it is acceptable.

Type II error (ASQ) An incorrect decision to accept something when it is unacceptable.

Unexpected Adverse Drug Reaction An adverse reaction, the nature or severity of which is not consistent with the applicable product information (e.g., Investigator's Brochure for an unapproved investigational product or package insert/summary of product characteristics for an approved product) (see the ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting).

Vulnerable Subjects Individuals whose willingness to volunteer in a clinical trial may be unduly influenced by the expectation, whether justified or not, of benefits associated with participation, or of a retaliatory response from senior members of a hierarchy in case of refusal to participate. Examples are members of a group with a hierarchical structure, such as medical, pharmacy, dental, and nursing students, subordinate hospital and laboratory personnel, employees of the pharmaceutical industry, members of the armed forces, and persons kept in detention. Other vulnerable subjects include subjects with incurable diseases, persons in nursing homes, unemployed or impoverished persons,

subjects in emergency situations, ethnic minority groups, homeless persons, nomads, refugees, minors, and those incapable of giving consent.

Well-being (of the trial subjects) The physical and mental integrity of the subjects participating in a clinical trial.

Appendix A: Sample CRF Printing Specifications Checklist

Total # of CRF binders to be printed _____
Total # of diaries to be printed _____
Total # of CRF pages per binder _____
# of NCR pages per binder _____
# of non-NCR pages per binder _____
# of diary pages per binder _____

Page formats: 2-part NCR with 2nd part cardstock, or specify other (The first part NCR should be white paper of weight 26):

Specify page format for diary pages and diary covers (ex. Tri-fold):

Tabs: specify # of banks, # tabs/bank, #tabs with printed instructions on back, mylar-laminated or not and mylar color:

Does printer need to add page numbers? : Y N

Binders (specify): Color: _____ Attach spine label □
Width: _____
inside pockets: _____ Attach cover artwork □

Timetable:
Final master copy to printer: _____/_____/_____
Prototype to XXX/XXX for review & approval_____/_____/_____
Shipment to arrive on site or at XXX/XXX _____/_____/_____

Packaging instructions:
_____
_____

Shipping Instructions: # of CRFs to be shipped per site _____
# of diaries to be shipped per site _____
will printer store backup supply? Y N

Specify ground or air and # of days to arrive _____

Attach bar code specifics □
Attach list of sites, addresses, and shipping priorities □
Attach final camera ready artwork for the CRF, tab back instructions, diary □
Attach specifications of layout

Additional comments: _____

Appendix B: Sample SAE Data Reconciliation Form

Sponsor Name
Protocol Number
Project Work Code
Reconciliation Session Date

Investigator/ Patient Number	Field Name	Description of Inconsistency		Description of Resolution/Action Required	Resolved By (initial and date)
		Drug Safety Database	Clinical Database		

Appendix C: Potential Metrics and Definitions

Because metrics should be based on goals and objectives set by a group or organization for their particular needs, it is difficult to recommend specific metrics that would fit the needs of every department or organization. However, most goals and objectives set by groups or organizations within our industry revolve around the interdependent areas of quantity, cost, time and quality associated with clinical trials. The table below provides a list of potential metrics and definitions to consider as a starting point for establishing a metrics program. This list of possible key metrics includes processes that would most likely be controlled across functional groups. Additional or different metrics could be included, as deemed appropriate by a CDM department or by company for their specific purposes.

Type of Metric	Proposed Metric	Potential Definition and Comments
Quantity	Amount of data cleaned	The amount (e.g. field or percentage of fields) of data deemed clean in a specified mount of time by an individual or a group of individuals
Quantity	Percentage of data entered	The amount (e.g. fields or percentage of fields) of data entered in a specified period of time by an individual or group of individuals
Cost	Resource Requirements	Total FTE for trial (projected vs. actual) or Total manpower hours for trial (projected vs. actual)
Cost	Activity-based Costing	Actual time (hours or days or FTE) for a specific task or functional area multiplied by the salary for those performing the task for that time frame multiplied by the overhead rate OR Time (hours or days or FTE) for a specific period of time (e.g. planning stage of trial) multiplied by the manpower required multiplied by the salaries for resources multiplied by the overhead rate
Time	Subject Visit to Data Accessible by Sponsor (or CRO)	Subject Visit Date (and Time, if possible) to Date (and Time, if possible) Sponsor (or CRO) has data in a reviewable form (i.e. data in database). The median, minimum and maximum elapsed times can be reviewed across all subjects. Site performance can be determined by grouping the data accordingly.
Time	Subject Visit to Clean Data Obtained	Time from Subject Visit to Data Locked (may be 'locked' per subject visit or per completed subject participation in trial)

Time	Last Subject Complete to Database Lock	Last Subject Complete (last subject in trial completed last visit of protocol) to Database Lock (database has received sufficient review and approval to allow for the application of the randomization codes, i.e., unblinding of the data and no further changes to the database are permitted without Senior Management approval.) For extension studies, include the last subject to be included in the study report.
Time and Quality	Inline Status Information	Query rate by site (overall median vs. actual) Percent data received (planned vs. actual) Percent outstanding queries (total vs. outstanding) Age of outstanding queries by site (all of the above best shown as timely graphs)
Quality	Number of Queries on Incoming Data	Total number of queries that require resolution by Site personnel after the data are reviewed by the Sponsor (or CRO)
Quality	Type and Distribution of Queries	For example, query rate by data module or queries requiring no change to database.
Quality	Error Detection in Final Database	Defined per QA and QC procedures to indicate number of errors detected in final database check; may be expressed as a total or as errors in pre-selected critical variable fields; if deemed unacceptable, the database cannot be locked.