Building Data Quality into Clinical Trials

by William J. Crerand, MS, Jana Lamb, Vera Rulon, RHIT, CCS, Bilun Karal, MS, and Jack Mardekian, PhD

A cornerstone of meaningful data is the collection process. Pharmaceutical companies employ several strategies to ensure the high quality of data collected in clinical trials. Learn more about these practices in this article.

As HIM professionals strive to move data collection activities closer to the point of care, they are guided by the principle that the quality of health data begins at the source. This principle can also be applied to data management activities in the pharmaceutical industry, which are guided by quality management expert and consultant W. Edwards Deming's teaching: "Cease dependence on mass inspection to achieve quality. Improve the process and build quality in the product in the first place."[1]

Data management's mission is to construct databases that are complete and high quality to meet a study's analytic objectives and comply with regulatory standards. In this article, we'll provide an overview of basic procedures and approaches used in the pharmaceutical industry to "build quality in the first place" into the data management process by concentrating on case report form (CRF) development through database release.

A Clinical Trial's Version of the Medical Record

Most HIM professionals are familiar with data management for healthcare providers or payers. When applying data management principles to the pharmaceutical industry, you'll find that the environment is similar in many ways. For example, CRFs function as medical records of the subjects participating in a given clinical trial. Similar to medical records in a facility, the quality of documentation directly relates to data quality. Although the information collected may vary and regulations and standards for documentation are somewhat different, the bottom line is the same. For quality data, quality checks must be performed throughout the entire process: from creation of forms to verification of documentation to coding and data quality.

Building quality into the data management process begins with designing a CRF based on the clinical trial protocol. Collecting only data that are required for the clinical trial is simple, sound logic that is sometimes ignored. The CRF designer studies the protocol to understand the type of data required for a study and then selects the appropriate standard safety and efficacy CRF modules to build the form. If no module in the standard module library exists to fit the study's specific needs, a new case report module is developed to collect the required data. A library of standard modules, together with an associated change control process, helps maintain quality.

The draft CRF is distributed for careful review by the clinical and biometrics teams. This team consists of key stakeholders in the data quality management process such as a CRF developer, a database programmer, a data manager, a clinical study manager or research associate, a statistical programmer, a statistician, and a medical monitor. A CRF design meeting is held to review and approve the CRF format and content and focuses on the data acquisition required for primary safety and efficacy variables. The design meeting allows the key stakeholders to ensure that the CRF is designed to collect the data specified by the protocol.

Start-up Phase Tests Database

The next step is designing the clinical trial database according to the approved CRF. The database developer reviews the approved CRF and identifies database modules that are appropriate for the study. Data entry screens are built using these selected modules. For new data types, the database developer requests the creation of new variables from the standards librarian so that new data modules can be developed.

Upon completion of the data entry screens, the database developer creates a series of testing scripts that are used by the data manager to test the data entry screens. The data manager ensures that the screens meet the study specifications and formally asserts that the data entry screens are of the appropriate quality.

To make certain that the data entered into the database are clean, accurate, and high quality, a series of programs referred to as validation procedures or edit checks are developed to check the data entered. (See "Sample Edit Checks" which is based on "Sample CRF Module".) Edit checks are programmed from the specifications created by the data manager in a data validation document. This document contains standard, computerized edit checks and data listings that are generated for every study plus study-specific edit checks and data listings that are primarily used to check the efficacy data unique to the current study.

The data validation document is reviewed by the clinical and biometrics teams to ensure that the appropriate checks and data listings are implemented. The database developer programs the edit checks and data listings in the data validation document and tests the edits and data listings using test data. The output of the testing is reviewed and approved by the data manager. The data manager may also create additional test data to further test some of the edits and listings. The process produces a database that has been approved for receipt of clinical trial data through either conventional data entry or electronic data capture.

Several other key documents are developed during the start-up phase prior to the first patient enrollment:

- **CRF completion guidelines**, which provide instructions on how each field of the CRF should be completed by the site
- **monitoring guidelines**, which outline how the CRF should be monitored at the study sites
- **data entry guidelines**, which define the process of data entry and how to handle various data scenarios that appear on the CRF
- **self-evident corrections document**, which clearly specifies changes that can be made to the database without site approval
- **data flow management plan**, which describes the flow of CRFs from retrieval at the site through database release and specifies quality metrics such as "data from a CRF received from a site will be entered within 48 hours of receipt"

Address Discrepancies Close to the Source

The process used to produce high-quality data during a clinical trial has undergone a paradigm shift. In the past, processes revolved around collecting data at the study sites and bringing them in-house to be processed in large batches or all at once. The data managers would be responsible for cleaning the data, and the quality of the data was dependent on the initial knowledge of the site personnel and the clinical monitor. This process of data collection and cleaning resulted in sub-par data quality, and the length of time to database release (when the data are ready to be used for analysis) was unacceptable.

The current philosophy is to improve data quality early in the process and to move data cleaning closer to the source. One of two methods are used:

- **Reviewing the first few CRFs upon receipt** and providing timely feedback to the clinical monitors and site personnel. Data managers should perform trend analyses on discrepancies. This review allows the clinical site to correct any misunderstandings and also provides the data manager with an opportunity to make necessary changes to the CRF, CRF guidelines, data entry guidelines, database structure, and edit checks.
- **Performing real-time data processing**. The goals are to process data upon receipt and to issue all data queries quickly.

These two strategies can have a great impact on the overall data quality for the clinical trial because a better understanding is achieved. Further, changes to documentation are made prior to the collection of a majority of study data.

The pharmaceutical industry is moving toward paperless processing. The completed CRF is received, tracked, and scanned into an imaging system. Each CRF and data query is indexed according to predefined guidelines so that they can be retrieved for the purposes of data entry, data review, and data processing.

Maintaining Quality at Every Step

There are a number of quality steps to make certain that proper indexing and timely data entry occur. Reports are generated from the imaging system, which alerts the data manager to quality issues such as missing or errant header information and missing pages. The data manager reviews and facilitates the correction of these issues prior to the data being entered into the database.

Data entry incorporates quality checks through a two-step process. The first operator enters the data from the image using data entry guidelines. The data entry guidelines define the process of entry and describe how to handle data that may appear on the CRF. Additionally, a more experienced operator enters the data a second time. The more experienced data entry operator or a supervisor resolves any discrepancies between the two entries.

The data manager is then responsible for cleaning the data. Upon entry of the data into the database, computerized edit checks and manual reviews of data listings are executed against the database to produce discrepancies. The data manager reviews each discrepancy and determines the reason the discrepancy was generated. Some discrepancies are due to data entry errors and are immediately corrected. Some discrepancies can be classified as self-evident on the basis of existing information on the CRF; these can be corrected by following procedures defined in the self-evident corrections document. For example, if the patient is male and childbearing potential is recorded, then the correction can be made following procedures outlined in the self-evident corrections document.

After review by the clinical team, the remaining discrepancies result in queries that need to go to the investigator site for clarification. The discrepancy management process is an ongoing process that begins with the receipt of the first subject data and continues throughout the clinical trial until all the data have been received and the database is released. The discrepancy management process is one way to review and clean clinical trial data to ensure data quality.

Next, data managers use data listings to review critical fields. Listings are different from edit checks because they are usually across subjects rather than within the same subject. The combination of discrepancy management and listing review gives the data manager confidence that the desired quality has been achieved for a study. During FDA audits of clinical studies, data listings generated by the sponsor are compared with CRFs and source material at the clinical site.

Minimum quality standards should be satisfied before the delivery of the CRFs in-house, which makes this exchange a quality checkpoint. The farther away from CRF completion a discrepancy is found, the greater the expense to remedy it. A discrepancy that goes undetected and is found during study report generation costs more than if the discrepancy were detected during the clinical monitoring at the study site. The expense and increased effort involved in discrepancy management emphasize the need to get it right the first time.

Database Release Signals Readiness

Database release is a major milestone in data management. It indicates that all known data discrepancies and data issues have been addressed and the database has been cleaned and is ready for statistical analysis. A checklist is maintained to document and verify that all steps of the database release process have been completed. Key components of the database release checklist include:

- CRFs for all patients enrolled in the study are entered
- final edit checks are executed and all discrepancies are resolved
- the CRF database is reconciled with other company databases, such as a serious adverse event database
- all electronic data captured from central laboratories are loaded into the database
- a clinician uses standard dictionaries to review medical coding, such as coding of adverse events

In addition, an independent quality control check of the database is performed, which is reviewed by the biometrics and clinical team prior to proceeding to database release. Documentation is filed in a document management system according to regulatory and company standard operating procedures.

Metrics aimed at continuously improving quality, such as the number of discrepancies generated per patient, the number of discrepancies requiring a database change per patient, the number of database re-releases required for the study, the percentage of discrepancies by type, and the time from the last patient visit to the database release, are maintained for trend analyses across studies.

The Same Goal across Settings

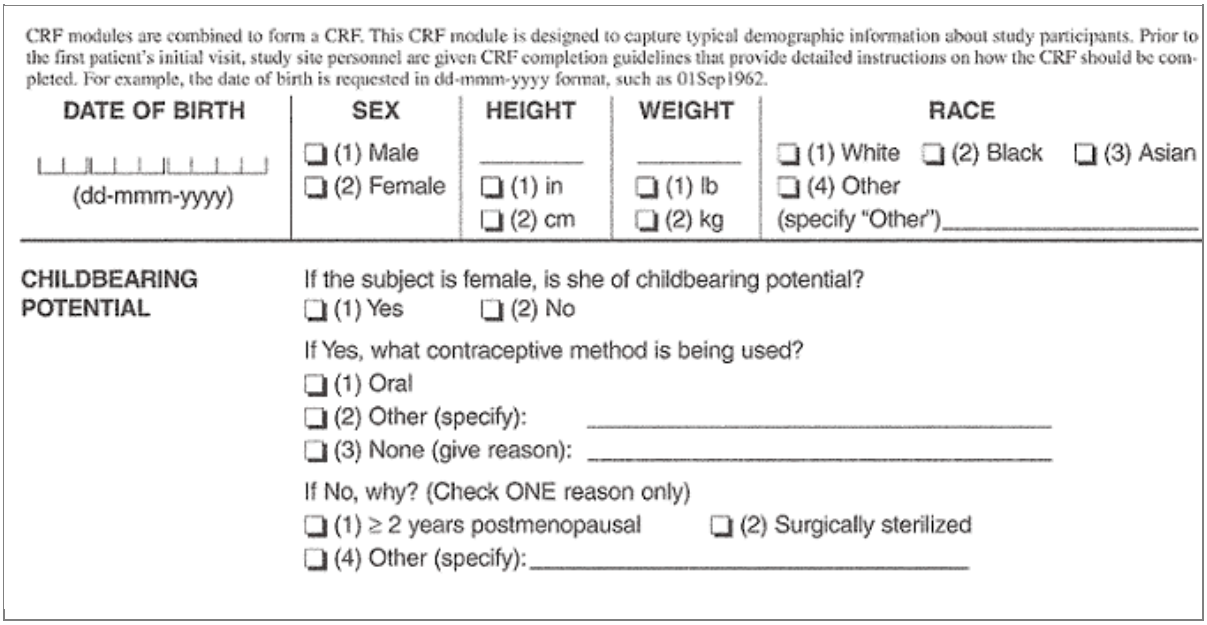
In many ways, managing data quality in clinical trials mirrors managing clinical data in general. In the provider world, agencies like the Joint Commission keep an eye on quality. In the world of clinical trials, pharmaceutical companies must comply with regulatory bodies like the Food and Drug Administration and the International Coalition of Harmonization. However, when an HIM professional might directly contact physicians in a facility, the clinical teams in a pharmaceutical company communicate via CROs and CRAs to ensure that any changes to documentation is agreed to by the investigator.

For the HIM professional, the application of the data quality management skill set in clinical trials is both challenging and exciting. It's a matter of thinking creatively and applying HIM concepts to new arenas with an eye toward the value of quality.

Sample Edit Checks

CRF Module	Edit Check Type	Specification	Discrepancy Generated by Edit Check
Demography	Univariate	Weight unit is not kg or lb.	Weight unit is not kg or lb. Please provide corrected weight/unit.
Demography	Multivariate	Age is less than 19 or greater than 72.	Age is out of the expected range 19-72. Please confirm or correct the date of birth.
Demography	Multivariate	Patient is male and childbearing potential is recorded	Patient is male, but childbearing potential . Please correct.

Sample CRF Module



CRF modules are combined to form a CRF. This CRF module is designed to capture typical demographic information about study participants. Prior to the first patient's initial visit, study site personnel are given CRF completion guidelines that provide detailed instructions on how the CRF should be completed. For example, the date of birth is requested in dd-mmm-yyyy format, such as 01Sep1962.

DATE OF BIRTH	SEX	HEIGHT	WEIGHT	RACE
|_|_|_|_|_|_|_|_|_|_|_| (dd-mmm-yyyy)	❑ (1) Male ❑ (2) Female	_____ ❑ (1) in ❑ (2) cm	_____ ❑ (1) lb ❑ (2) kg	❑ (1) White ❑ (2) Black ❑ (3) Asian ❑ (4) Other (specify "Other")_____

CHILDBEARING POTENTIAL

If the subject is female, is she of childbearing potential?
❑ (1) Yes ❑ (2) No

If Yes, what contraceptive method is being used?
❑ (1) Oral
❑ (2) Other (specify): _____
❑ (3) None (give reason): _____

If No, why? (Check ONE reason only)
❑ (1) ≥ 2 years postmenopausal ❑ (2) Surgically sterilized
❑ (4) Other (specify): _____

Outsourcing Requires Clear-cut Standards

Faced with limited in-house resources, many companies choose to outsource clinical data management to contract research organizations (CROs). Dependent on each other, CROs and pharmaceutical companies work to produce a quality product that will satisfy both of their needs. Most companies choose CROs as an alternative resource when faced with time constraints. However, unless the relationship between the pharmaceutical company and the CRO is a partnership working toward a common outcome, data quality can suffer.

The overall definition of quality data must be shared between the sponsor company and the CRO. A simple definition is information that is complete, consistent, and accurate. Quality control steps must be developed in partnership with the CRO to ensure the data possess all three of these characteristics. Below is a closer look at each characteristic.

Complete: The sponsor company should communicate its expectations regarding what makes up a complete patient record before subject data come in from the investigator site. Then, the CRO can act accordingly when a CRF does not match the defined parameters and resolve any discrepancies. Without a common understanding of what constitutes a standard subject record, accurate identification of a discrepancy is not possible. Because most studies have multiple exceptions to the standard, the CRO needs an accurate description of what the sponsor considers incomplete.

Consistent: Data must be logical and consistent within and across all subjects. One way to accomplish consistency is standardization. This can be achieved by using standardizing processes among biometrics and clinical teams at the sponsor company and reducing the number of different CROs used by the sponsor company. Another procedure is to have the CRO log directly into the sponsor company's computer systems to perform data management and statistical analysis. Appropriate training must be provided in executing all relevant processes. The sponsor company should be aware that in any task, a learning curve should be anticipated. Once key personnel are trained in sponsor company processes and with repeated use of a CRO team, working relationships and communication of expectations will become second nature.

Accurate: As all HIM professionals know, data are only as good as the accurate transcription from the source documentation at the investigator site and the instructions given for CRF completion. Quality data monitoring at the sites heads off the discrepancy management cycle at the back end. If a CRO has been contracted to monitor the study, expectations for the type of monitoring activities must be clearly communicated to the monitoring staff. However, the real timesaver in data accuracy at the source is in the number of handoffs (i.e., back and forth iterations among the study site, the CRO, and the sponsor) required to fix discrepant data once retrieved. Financial incentives for performance metrics, such as maintaining the number of discrepancies requiring investigator signature at less than six, should be considered in the contract language.

The role of the CRO should be greater than simply identifying incorrect data via electronic checks. The sponsor should solicit ideas from the CRO for opportunities to improve protocol and CRF design. CROs process data from a variety of sources and have experience in industry best practices. With continued use of CROs, it is important to establish and maintain long-term relationships with them. A good relationship improves communication and increases the likelihood of reaching the common goal: data quality.

Clinical Trials 101

Following are some key terms used in clinical trials in the pharmaceutical industry:

Case report form (CRF): a repository for all the protocol-required information on each subject in a clinical trial that is reported to the sponsor of the clinical trial. A CRF can be paper or electronic. CRFs are typically organized by modules. A CRF demographic module is used to record a subject's characteristics such as age, gender, and family history of the disease under study

Clinical research associate (CRA): a person employed by a sponsor or a CRO who monitors the progress of investigator sites participating in a clinical study to ensure the study is being conducted in accordance with the protocol (also known as a clinical study manager)

Clinical trial/clinical study: an investigation in human subjects intended to discover or verify the clinical, pharmacokinetic, or other pharmacodynamic effects of an investigational product to ascertain

its safety or efficacy

Contract research organization (CRO): an organization that offers a wide range of services such as clinical trial management, data management, and statistical activities to the pharmaceutical industry

Database release: an important milestone in data management that occurs when all CRF information is entered into a database and the data are determined to be clean. Database release indicates the data are ready for statistical analysis

Edit checks: computer programs and manual reviews that are executed against a CRF database to search for deficiencies and discrepancies in a database such as data entry errors and illogical data. For example, if a subject's birth date in the database matched the subject's date of entry into the clinical trial, a discrepancy would be generated for resolution

Protocol: a document that describes the objectives, design, methodology, statistical considerations, and organization of a clinical trial

Source documents: all information contained in original records and certified copies of results, observations, or other aspects required for the reconstruction and evaluation of the study. Includes original documents, data, and records. Source document examples include pharmacy dispensing records, photographic negatives, laboratory notes, hospital records, and clinical and office charts

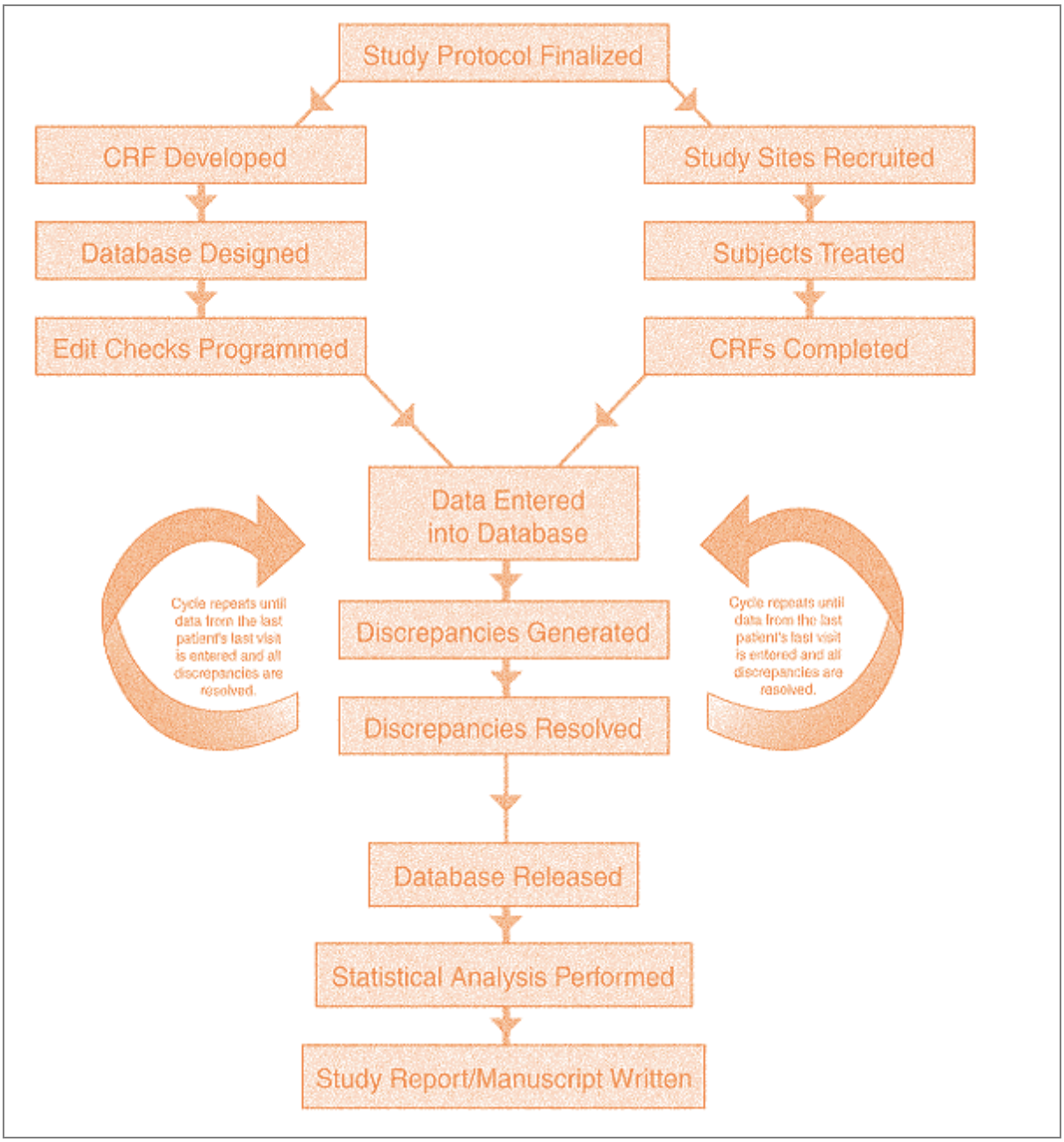
Sponsor: an individual, company, institution, or organization responsible for the initiation, management, or financing of a clinical trial. A pharmaceutical company is a typical sponsor

Study investigator: a person responsible for conducting the clinical trial at a study site. This person is typically a medical professional, usually a physician under whose direction an investigational drug is administered or dispensed

Study site: the location where trial-related activities are actually conducted. This may be in a hospital, private practice setting, or outpatient clinic, depending on the nature of the population under study

Subject: a participant in a clinical trial about whom an investigator obtains private information or data through intervention or interaction

A Clinical Trial, Step by Step



Note

1. Deming, William E. *Out of the Crisis*. Cambridge: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986.

Reference

Spilker, Bert. *Guide to Clinical Trials*. New York: Raven Press, 1991.

William J. Crerand *is Director, Data Operations,* ***Jana Lamb*** *is Manager, Quality Standards,* ***Vera Rulon*** *(vera.rulon@pfizer.com) is Director, Records Management Services/ETMF,* ***Bilun Karal*** *is Director, Headquarters Data Management, and* ***Jack Mardekian*** *is Senior Director, Global Data Operations at Pfizer, Inc.*

Article citation:Crerand, William J., et al. "Building Data Quality into Clinical Trials." *Journal of AHIMA* 73, no.10 (2002): 44ff.