

---

# Measuring Data Quality

September 2008

## *Abstract*

Data collected during a clinical trial must have as few errors as possible to be able to support the findings or conclusions drawn from that trial. Moreover, proof of data quality is essential for meeting regulatory requirements. This chapter considers the challenges faced by clinical data management professionals in determining a dataset's level of quality, with an emphasis on the importance of calculating error rates. An algorithm for calculating error rates is presented in this chapter and is asserted to be the preferable method for determining the quality of data from a clinical trial.

## *Introduction*

This chapter concentrates on identifying, counting and interpreting errors in clinical trial data. Data quality measurement methods are very important and should be applied to clinical trial operations as part of an overall planned approach to achieving data quality. Although measuring data quality is important, it is equally if not more important to focus on preventing errors early in the protocol development and data handling process design stages. Error prevention will be addressed in the “Assuring Data Quality” chapter of the GCDMP.

Federal regulations and guidelines do not address minimum acceptable data quality levels for clinical trial data, therefore it is left up to each organization to set their own minimum acceptable quality level and methodology for determining that level. As a result, differences in methodology for determining data quality and estimated error rates are often not comparable between different trials, vendors, auditors or sponsors. It is important that data management professionals take a proactive role to set appropriate standards

for acceptable data quality levels, to utilize methods for quantifying data quality, and to implement practices to assure data quality.

## **Scope**

This chapter provides minimum standards, best practices, and methods for measuring data quality.

The Institute of Medicine (IOM) defines “quality data” as data that support conclusions and interpretations equivalent to those derived from error-free data<sup>1</sup>. To make the IOM definition of data quality operational, organizations must understand sources of errors, identify errors through inspections, use inspection results to measure data quality, and assess the impact of the data quality on conclusions drawn from the trial.

## **Minimum Standards**

- Use statistically appropriate inspection sample sizes for decision making.
- Document the method and frequency of data quality assessments in the study’s data management/quality plan.
- Perform at least one quality assessment of the study data prior to final lock.
- Document data quality findings and corrective actions, if needed.
- Determine acceptable error rates for primary and secondary safety and efficacy (also known as “critical”) variables.

## **Best Practices**

- Use quantitative methods to measure data quality.

NOTE: Quantitative methods for measuring data quality involve classifying the data, counting the data, and constructing statistical models to help explain database quality, database errors, and patterns of errors. Database errors, or “findings”, can be generalized to the entire data set, and direct comparisons can be made between the sample and the whole data population as long as valid sampling and significance techniques are used. Quantitative methods

help differentiate between data errors that might be pervasive in the data set and errors that are merely random occurrences.

- Compare trial data and processes in the beginning, middle, and end stages of the trial.
- Work with clinical operations to predefine criteria to trigger site comparisons based on monitoring reports.
- Perform quality control on 100% of key safety and efficacy (critical) variables.
- Monitor aggregate data by site to detect sites whose data differ significantly so that appropriate corrective actions can be taken.
- Perform quality control prior to release of data used for decision making.

### ***Other Best Practice Considerations***

- “When a long series of data processing steps occurs between the source document and the final summaries (as when the source document is transcribed to a subject’s chart, transcribed onto a case report form, entered into a database, and stored in data tables from which a narrative summary is produced)”<sup>2</sup> compare the final summaries directly against the source document, at least on a sample of cases.
- Streamline data collection and handling to limit the number of hand-offs and transfers.
- Perform a data quality impact analysis. Impact analysis in data quality is a methodical approach used to assess the impact of data errors or error patterns on the trial or project. Through impact analysis, potential risks or opportunities can be identified and analyzed. Impact analysis can provide key information to aid in decision making.
- Evaluate the results of the impact analysis and propose system and process changes.

- Perform the appropriate level of risk assessment to ensure data quality based on the type and purpose of the trial. For more on this, see the “Assuring Data Quality” chapter.

## **Data Errors**

A clinical research study is a complex project involving many processing steps. Each step where data are transcribed, transferred, or otherwise processed has an error potential associated with it.

A data error is defined as a data point that inaccurately represents a true value. There are many sources or causes of data errors, including but not limited to, incorrect transcription at a clinical site, incorrect data processing, unintended responses based on an ambiguous question, or collection of data outside a required time window.

Common errors in clinical trial data are compiled from several references and are shown in Table 1.<sup>2, 3, 4, 5</sup> Table 1 also suggests some detection methods data managers can employ to identify data errors.

### **Error Detection**

It is not practical, necessary, or efficient to design a quality check for every possible error, or to perform a 100% manual review of all data. There will always be errors that are not addressed by quality checks or reviews, and errors that slip through the quality check process undetected.

Programmatic checks (data validation and/or edit checks) should be applied consistently across trial data, and all errors that are identified in this manner may be corrected. At a minimum, these checks should target fields critical to the analysis where errors may have a greater impact on the outcome of the study. However, not all errors can be detected using these methods. For example, unreported adverse events may be difficult to identify using programmatic checks.

**Table 1. Common Sources of Error and Primary Detection Methods**

<b><i>SOURCES OF ERROR</i></b>	<b><i>DETECTION METHODS</i></b>				
	Programmatic Data Checks	Source Data Verification	Data Validation	Aggregate Statistics	CRF-to-Database Inspection
Subject completes questionnaire incorrectly or provides incorrect or incomplete answers to questions (lack of tool validation or bad form design)			<b>X</b>		
Subject does not follow trial conduct instructions		<b>X</b>			
Inadequate instructions given to the subject				<b>X</b>	
Site personnel trial conduct error (protocol violation)		<b>X</b>		<b>X</b>	
Data captured incorrectly on the source	<b>X</b>	<b>X</b>			
Site personnel transcription error	<b>X</b>	<b>X</b>	<b>X</b>		
Site equipment error				<b>X</b>	
Human error in reading equipment or print out or inter-rater-reliability		<b>X</b>			
Data entry error	<b>X</b>	<b>X</b>	<b>X</b>		<b>X</b>
Electronic data acquisition error (power glitch, back up that didn't run, lead not attached securely)			<b>X</b>		<b>X</b>
Data linked to the wrong subject		<b>X</b>	<b>X</b>		<b>X</b>
Database updated incorrectly from data clarification form or query					<b>X</b>
Missing data	<b>X</b>	<b>X</b>			
Outliers	<b>X</b>				
Data inconsistencies	<b>X</b>	<b>X</b>			
Programming error in user interface or database or data manipulations					<b>X</b>
Lost data		<b>X</b>	<b>X</b>		
Fraud		<b>X</b>		<b>X</b>	

Errors caused by fraud and protocol violations can be difficult to detect without the use of special programming and the use of aggregate statistics.<sup>3, 5, 6, 7, 8, 9</sup> Throughout a trial, aggregate statistics should be available to monitors to facilitate detection of misunderstandings, misconduct and fraud. Data management is the first point in many processes where the data are available

for viewing in aggregate across sites. It is at this earliest point that aggregate statistics should be provided to monitors and other study personnel to quickly identify sites that are behaving differently from the rest. Aggregate data reports may be designed to summarize the performance of individual centers in the areas of recruitment, extent of follow-up, compliance to treatment, completion of procedures, late visits, or data queries.<sup>2</sup>

Source data verification (SDV) may be used to identify errors that are difficult to catch with programmatic checks. For example, a clinical trial monitor at the investigator site performs SDV by comparing the medical record (a subject's chart) to the CRF. Any discrepancies between the two that are not explained by CRF completion instructions, the protocol, or other approved site conventions are counted as errors. In addition, if a study is using electronic data capture (EDC) methods, SDV may be the best way to check for data errors. The scope of SDV can be decided on a trial-by-trial basis and should be determined at the beginning of the trial.

## ***Inspection or Comparison of Data***

ICH E6 defines an inspection as “the act by a regulatory authority(ies) of conducting an official review of documents, facilities, records, and any other resources that are deemed by the authority(ies) to be related to the clinical trial, and that may be located at the site of the trial, at the sponsor's or CRO's facilities or both, or at other establishments deemed appropriate by the regulatory authority(ies).”<sup>10</sup>

The American Society for Quality (ASQ) defines inspection as “measuring, examining, testing, and gauging one or more characteristics of a product or service and comparing the results with specified requirements to determine whether conformity is achieved for each characteristic.”<sup>11</sup> Here the term inspection is used to indicate a scope narrower than a comparison, and is a process where measuring can be performed as a step in the work process with less independence than a comparison. For example, a CRF-to-database inspection may be performed by individuals in the same department or on the same project team as those who did the work, as long as they are not the individuals who performed the work being inspected. In contrast, a comparison is often performed by trained company or sponsor representatives. Many organizations require both ongoing inspections and/or more formal comparisons to assure high quality data for all their trials.

## Data Comparison

Errors can be detected by comparing two representations of data captured at different points in the data handling process. A CRF-to-database comparison is performed by comparing the CRF to the data stored in the database.

Depending on the needs of the study, the comparison may be performed on the clinical database immediately following data entry, or on the analysis-ready datasets at database lock. In either case, an error is defined as a discrepancy between the dataset and the CRF that is not explained by data handling conventions, site signed data clarification forms, or programming conventions defined in the trial analysis plan.

## Sample Size

The best practice for sample size selection is using a statistically appropriate sample size for each inspection. This assures that information obtained from the inspection is representative of the entire database and can be used in decision making. It is important that the data manager work with key study personnel to develop and document a sampling methodology. For studies having a large enough study population, one sample size algorithm commonly used by many organizations is the square root plus one ( $\sqrt{n} + 1$ ) of the total study population. Another approach used is having a sample size equal to ten percent (10%) of the total study population.

## Error Rates

Data quality can be quantified in two ways: (1) raw counts of numbers of errors, and (2) error rates. Calculating an error rate guards against misinterpretation of error counts, can facilitate comparison of data quality across database tables and trials, and therefore is the preferable method. Caution should be used when interpreting raw error count data. These data can be misinterpreted if used to compare the quality of different database tables within the same database, or the data quality of two different trials.

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

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

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

DEMOG error rate = 10,000(Number of Errors Found / Number of Fields Inspected)

DEMOG error rate = 10,000(10/100)

DEMOG error rate = 1000 errors per 10,000 fields

Include the mathematical calculation(s) and the final, calculated error rate(s) in a report that summarizes database quality.

This is just one example of how to express error rates, and error rates can also be expressed through other means, such as by a percentage or a p value.

### **Important Concepts About Error Rates**

- The error rate is only a part of data quality process evaluation. It is important to know if the errors are in critical or noncritical fields. If the high error rates are in noncritical fields, they may have little impact. In this case, an organization may determine that it is not worth the time and effort required to clean these data.
- Knowledge of the error rates can help you choose the process paths and technology that will yield the highest quality for your organization.



## **Acceptable Quality**

In the absence of industry-wide standards for acceptable error rates for CRF to database quality control, “quality data” means different things to different organizations. Popular definitions of an “acceptable quality level” include rates of 50 errors per 10,000 fields overall, and different standards for critical and noncritical variables. These standards range from 0 to 10 errors per 10,000 fields for critical variables, and 20 to 100 errors per 10,000 fields for noncritical variables.

There are many ways to quantify data quality and calculate an error rate. While the differences among the methods can be subtle, the differences among the results can be by a factor of two or more. For example, consider the hypothetical situation of two lab data vendors calculating error rates on the same database with three panels. The Protocol Number, Site Number, and Sponsor Number are default fields that do not require data entry, in all of three database panels. Vendor 1 includes each of these default fields in the field count as fields inspected, which results in a denominator of 100,000 fields inspected in the error rate calculation. Vendor 2 does not include them in the field count since they are default fields, for a denominator of 50,000 fields inspected. Both vendors do a data quality inspection and both vendors find 10 errors. When they calculate the error rates, Vendor 1 has an error rate half that of Vendor 2 only because they did not follow the same algorithm for field counts. This example illustrates how important it is for a common algorithm to be followed by all parties calculating error rates. It is imperative that the units in the numerator and denominator be the same. Some other examples of algorithm details that could skew results are:

- Should data errors involving derived fields be counted?
- Is an error in the month and year fields of a derived date one error or two?
- How should errors be counted in a header that are entered one time then electronically populated throughout the study pages?

Data managers, statisticians and other trial personnel must work together to define acceptable data quality levels for trials, to design data collection and processing so as to achieve the desired level of data quality, to measure data quality and monitor it throughout the trial, and to communicate data quality information to key stakeholders, including management and sponsors.

## **Documentation**

Documentation of data quality comparisons should include the number of errors found, the error rate, how the numerator and denominator were defined and the final error rate. Anyone reading the documentation should be able to recreate the sampling and error rate calculations and produce the exact same results. For information addressing how these processes may differ in studies using EDC, please refer to the chapters entitled "Electronic Data Capture—Study Conduct" and "Electronic Data Capture—Study Closeout."

## **Recommended Standard Operating Procedures**

- Measuring Data Quality
- Monitoring Data Quality
- Data Quality Acceptability Criterion

## **References**

1. Davis JR, Nolan VP, Woodcock J, Estabrook RW. eds. Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making. Workshop Report. *Roundtable on Research and Development of Drugs, Biologics, and Medical Devices, Division of Health Sciences Policy, Institute of Medicine*. Washington DC: National Academy Press; 1999.
2. Knatterud GL, Rockhold FW, George SL, et al. Guidelines for Quality Assurance in Multicenter Trials: A Position Paper. *Controlled Clinical Trials*. 1998;19:477-493.
3. Califf RM, Karnash SL, Woodlief LH. Developing Systems for Cost-Effective Auditing of Clinical Trials. *Controlled Clinical Trials*. 1997;18:651-660.
4. DeMets D. Distinctions between Fraud, Bias, Errors, Misunderstanding, and Incompetence. *Controlled Clinical Trials*. 1997;18:637-650.
5. Williams GW. Discussion: Errors, Misunderstandings, and Fraud. *Controlled Clinical Trials*. 1997;18:661-666.

6. Ventura D. Issues in Clinical Trial Management, Part II Corrupt Data: Sources and Solutions. *Research Nurse*. 1999; March/April:12-17.
7. Bauer P. Statistical Process Monitoring During Clinical Trials. In: *Abstracts from the DIA Thirty Fifth Annual Meeting; June 27 – July 1 1999, Baltimore, MD*: Drug Information Association; 1999
8. Knatterude GL. Quality Assurance in Multicenter Clinical Trials: An Overview. *Abstracts from the DIA Thirty Fifth Annual Meeting; June 27 – July 1 1999, Baltimore, MD*: Drug Information Association; 2000.
9. Nahm ML. Optimizing Data Quality Through Statistical Process Control. *Proceedings from the 15 Th. Annual US Clinical Data Management Symposium, March 2000, Philadelphia, PA*: Drug Information Association.
10. International Conference on Harmonisation. Guideline for Good Clinical Practice, E6: Good Clinical Practice; Consolidated Guideline. Geneva, Switzerland: Author; 1996.
11. American Society for Quality, “Glossary”, Available at: <http://www.asq.org/glossary/i.html>. Accessed on April 3, 2008.

### **Further Reading**

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

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

## ***Chapter Revision History***

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