



## THE CASE FOR CLINICAL DATA QUALITY

*Kaye H. Fendt, MSPH  
Director, DQRI*

*Associate Professor, School of Medicine, University of North Carolina at Chapel Hill*

*This paper originally appeared in the Summer 2004 issue of Data Basics, the publication of the Society for Clinical Data Management.*

### Introduction

Over the last 2 decades, as regulatory requirements have increased and health-related research has focused more on rare diseases and conditions, the number of trials, patients and clinical procedures required to demonstrate significant differences has also increased. As a result, we have seen a large increase in the amount of data requiring corporate staff to electronically capture, review, analyze, interpret, and summarize in clinical research reporting, whether reported in peer reviewed journal articles or in support of a marketing application with a format regulated by the various biopharmaceutical regulatory agencies. Unfortunately, methods and metrics for assessing the quality of these data and the impact of that quality on the data-based decisions made have not developed at the same pace. As data management has matured over the last 40 years, increasing emphasis has been placed on the procedure to eliminate errors primarily from the point of database entry to the extraction of data from the database. Efforts to design quality into the procedures have developed in a variety of ways across the research community. Today there is an emerging need to define, measure, and evaluate the quality of data and the impact of the quality of data used for decision making throughout the entire life-cycle of the data. Also, there is a growing need for further research into how these methods can be standardized, reported in a way to allow responsible use of the data in the future, cost-contained, and implemented in an efficient manner.

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

### The Impact of the Quality of Regulated Clinical Data

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

#### Data Quality Research Institute

University of North Carolina at Chapel Hill • CB#7226 • 200 TimberHill Place, Suite 201 • Chapel Hill, NC 27599-7226  
Primary Phone: (919) 843-6615 • Cell: (919) 417-3540 • Email: [info@dqri.org](mailto:info@dqri.org)

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

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

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

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

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

#### Data Quality Research Institute



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

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

## Current State of the Quality of Biopharmaceutical Data

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

## Standards

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

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

The Good Clinical Data Management Practices (GCDMP) ([www.scdm.org](http://www.scdm.org)) provides guidance and thus establishes standards on accepted practices for the many areas of Clinical Data Management (CDM) that are not covered by existing regulations

### Data Quality Research Institute

University of North Carolina at Chapel Hill • CB#7226 • 200 TimberHill Place, Suite 201 • Chapel Hill, NC 27599-7226  
Primary Phone: (919) 843-6615 • Cell: (919) 417-3540 • Email: [info@dqri.org](mailto:info@dqri.org)



and guidance documents. The intent is to remain consistent with regulatory practices in related areas of clinical research and to apply the concepts contained in those regulations and associated guidance documents to Clinical Data Management. The GCDMP recommends guidelines and provides practical suggestions and proven means for meeting the guidelines.

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

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

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

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

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

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

<sup>2</sup> English, Larry P. Improving Data Warehouse and Business Information Quality, New York: Wiley, 1999.