

Assuring Data Integrity and Data Quality in Sponsor Submissions

Paul Schuette, US FDA, Office of Biostatistics, Silver Spring MD, USA

ABSTRACT

The integrity and quality of data in sponsor submissions is a critical concern of the FDA. Poor data quality, whether due to sloppiness or fabricated data, has the potential to undermine the ability of the FDA to provide appropriate analyses as part of the medical product approval process, as well as public trust in those analyses. This paper will outline some of the methods by which sponsors and the FDA can promote and improve data quality and data integrity. Among these methods and approaches are the quality by design paradigm employed by the FDA, source verification, and centralized statistical analyses. Additionally, approaches used to facilitate the detection of data anomalies, and attendant issues will also be discussed.

INTRODUCTION

Evidence based review systems are integral to FDA's decision making processes. Clinical trial data from sponsor submissions are a critical component for medical product approvals and decision making in the Center for Drug Evaluation and Research (CDER) within the FDA. Data quality and data integrity are critical components underlying FDA's evidence based reviews. Part of the purpose of filing meetings with industry is to determine whether the submitted data can support statistical and other analyses. Missing data, poorly collected data, as well as poorly cleaned and poorly processed data, can undermine analyses and even render them invalid. Thus, FDA, regulated industry, and academia all have a vested interest in assuring data quality and data integrity throughout the drug development process and the entire drug life cycle.

CASE STUDIES

Several case studies illustrate some of the issues and adverse outcomes for sponsors and global public health when data quality and integrity are compromised:

1. An inspector reported an incident in which several members of the same family were all enrolled in a clinical trial on a Friday evening. Other potentially questionable practices were found to have occurred at this site, which was the largest site in the trial. The sponsor withdrew the product from consideration after FDA expressed concerns regarding the site and the trial.
2. Semler Research Private Limited. A 2015 inspection in Bangalore, India found significant instances of misconduct and violations of federal regulations, including the substitution and manipulation of study subject samples. As part of a coordinated action with WHO and EMA, letters were sent to sponsors notifying them of the need to repeat all bioavailability and bioequivalence studies performed by Semler, [1]. This affected U.S. President's Emergency Plan for AIDS Relief (PEPFAR).
3. A clinical reviewer noticed that more adverse events were reported in a low dose arm than in a high dose arm in a recent clinical trial submission. When queried, it was discovered that the sponsor's CRO had incorrectly coded the treatment arms.
4. A reviewer reported that a sponsor submitted a legacy data set in which the data quality was so poor as to be incapable of supporting a review. Among other findings, unique subject identifiers were not employed by the sponsor. Prior to receiving a complete response letter, the sponsor withdrew the product from consideration.

Items 1 and 2 above illustrate problems occurring at a single site. Item 3 illustrates a problem that occurred after data collection and during data processing and analysis.

DEFINITIONS

One of the issues with data quality is that there does not appear to be a single, well-defined, definition of data quality. ISO 8000 [2] states "*Data quality* is the degree to which data meets user requirements" (Italics added). Furthermore, it is also stated that:

"The following principles of data quality underlie ISO 8000:

PhUSE Connect 2018

- a. Data quality involves data being fit for purpose; i.e., the decision it is used in.
- b. Data quality involves having the right data, in the right place, at the right time.
- c. Data quality involves meeting agreed customer data requirements.
- d. Data quality involves preventing the recurrence of data defects by improving processes to prevent repetition and eliminate waste."

This definition of data quality as "fit for purpose" involves a potentially subjective evaluation of fitness.

Data integrity has been defined as a separate concept. A MHRA guidance [3] states:

"Data integrity is the degree to which data are complete, consistent, accurate, trustworthy, reliable and that these characteristics of the data are maintained throughout the data life cycle. The data should be collected and maintained in a secure manner, so that they are attributable, legible, contemporaneously recorded, original (or a true copy) and accurate. Assuring data integrity requires appropriate quality and risk management systems, including adherence to sound scientific principles and good documentation practices." (italics in original)

The acronym ALCOA is commonly used to describe data that are **attributable**, **legible**, **contemporaneously recorded**, **original** (or a true copy) and **accurate**.

In theory, it would appear that one could have data quality without data integrity, but that seems unlikely since data integrity assures the provenance of the data. In practice, data integrity is necessary to ensure that high quality data is available.

QUALITY BY DESIGN

The US Food and Drug Administration, along with other regulators, has embraced the *quality by design* paradigm. Implementation of the quality by design paradigm has extensively relied upon GxP guidelines, where GxP stands for Good x Practice. For example, there is Good Clinical Practice (GCP), Good Manufacturing Practice (GMP), Good Clinical Data Management Practices (GCDMP, [4]), and PhUSE's own Good Programming Practice (GPP, [5]). FDA has recently (March 2018) distributed E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1) Guidance for Industry [6], which contains an overview of many salient topics, some of which are subjects of separate guidance documents. Of particular note are the following:

"4.1.4 The investigator/institution should permit monitoring and auditing by the sponsor, and inspection by the appropriate regulatory authority(ies)."

"4.9.7 Upon request of the monitor, auditor, IRB/IEC, or regulatory authority, the investigator/institution should make available for direct access all requested trial-related records."

Section 5 of the E6(R2) guidance deals with sponsors, and has subsections for:

- Quality Management
- Quality Assurance and Quality Control
- Trial Design
- Trial Management, Data Handling, and Recordkeeping
- Monitoring
- Audit
- Noncompliance

In particular,

"5.1 Quality Assurance and Quality Control

5.1.1 The sponsor is responsible for implementing and maintaining quality assurance and quality control systems with written SOPs to ensure that trials are conducted and data are generated, documented (recorded), and reported in compliance with the protocol, GCP, and the applicable regulatory requirement(s).

5.1.2 The sponsor is responsible for securing agreement from all involved parties to ensure direct access (see section 1.21) to all trial-related sites, source data/documents, and reports for the purpose of monitoring and auditing by the sponsor, and inspection by domestic and foreign regulatory authorities.

PhUSE Connect 2018

5.1.3 Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly."

Thus, sponsors have a primary responsibility for ensuring the data quality and data integrity of submitted data. Furthermore, sponsors should ensure that source data from individual investigators is available to regulators for audit and inspection purposes.

The E6(R2) guidance [6] discusses monitoring at some length, and calls for monitoring plans to employ a risk-based approach, including on-site source validation and centralized statistical monitoring. As noted in the guidance:

"Review that may include statistical analyses of accumulating data from centralized monitoring can be used to:

- (a) Identify missing data, inconsistent data, data outliers, unexpected lack of variability and protocol deviations
- (b) Examine data trends such as the range, consistency, and variability of data within and across sites
- (c) Evaluate for systematic or significant errors in data collection and reporting at a site or across sites; or potential data manipulation or data integrity problems
- (d) Analyze site characteristics and performance metrics
- (e) Select sites and/or processes for targeted on-site monitoring."

Centralized statistical monitoring can be an important component of a risk-based monitoring plan, allowing for greater exploration of potential data quality issues, while also potentially reducing monitoring costs.

CDER Efforts

Within the US FDA, the Center for Drug Evaluation and Research (CDER) has established the CDER Data Integrity Advisory Board (CDIAB) to work on data integrity and data quality issues dealing with both clinical trial and manufacturing data. Additionally, the Office of Translational Sciences has established two working groups to address ongoing issues with data from clinical trial sites with clinical endpoints (NDAs, BLAs and some ANDAs) and with PK/PD data (ANDAs).

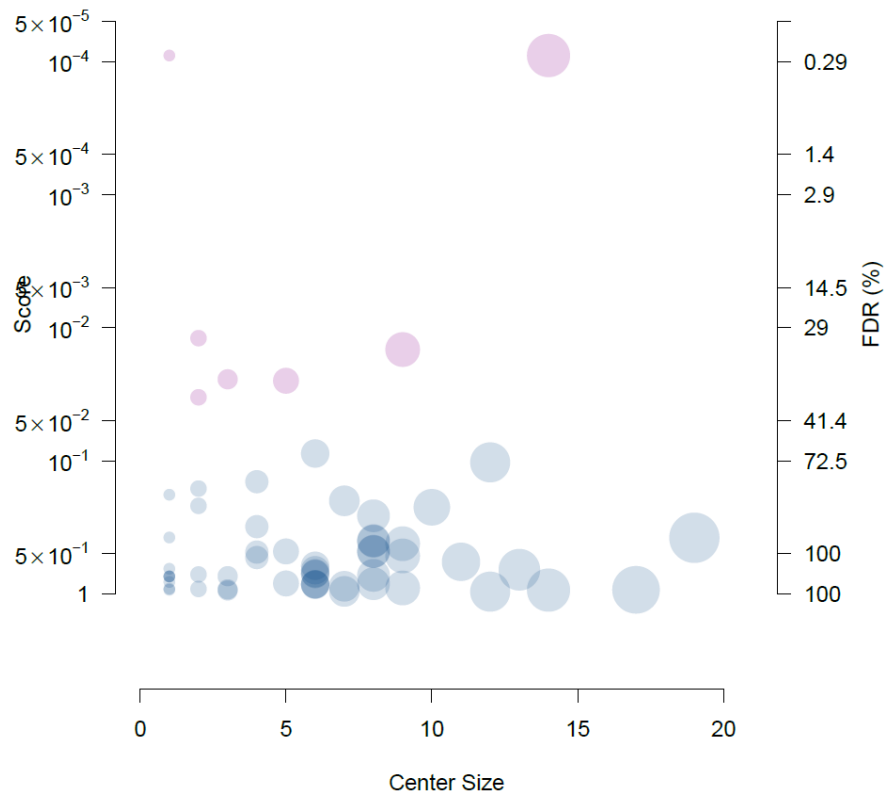
Several tools to assist with data quality and integrity issues have been developed and/or prototyped, in conjunction with staff from the Office of Scientific Investigations (OSI) and the Office of Study Integrity and Surveillance (OSIS).

- Clinical Investigator Site Selection Tool (CISST). CISST is an Excel-based tool that analyzes site-level clinical trial data by study and provides a risk-rated prioritized list of clinical investigator sites for inspection. See Standardized Format for Electronic Submission of NDA and BLA Content for the Planning of Bioresearch Monitoring (BIMO) Inspections for CDER Submissions [7] and Bioresearch Monitoring Technical Conformance Guide [8] for details.
- Core DataFitness Assessments (CoreDF). CoreDF summarize a list of common data quality errors in SDTM data sets into a standalone, self-explanatory report, using Pinnacle 21 software.
- CISST Assist. An OCS pilot for using JMP Clinical in tandem with CISST.
- CluePoints Cooperative Research and Development Agreement (CRADA). The FDA and CluePoints have a three year CRADA in place to evaluate and further develop CluePoints's SMART™ engine for Risk Based and Central Statistical Monitoring. Using subject level data, the SMART engine applies a battery of statistical tests to variables across multiple data sets in a clinical trial and derives a data inconsistency score. Inspections and/or sensitivity analyses can then be conducted on potentially suspect sites.
- CluePoints Cooperative Research and Development Agreement (CRADA). The FDA and CluePoints have a three year CRADA in place to evaluate and further develop CluePoints's SMART™ engine for Risk Based and Central Statistical Monitoring. Using subject level data, the SMART engine applies a battery of statistical tests to variables across multiple data sets in a clinical trial and derives a data inconsistency score. Inspections and/or sensitivity analyses can then be conducted on potentially suspect sites.
- SAGE. A Python based tool for OSIS inspectors.
- Data Anomaly Detection Tool. A R Shiny app to explore and detect excessively similar PK/PD profiles in ANDAs for both reviewers and inspectors.

SMART™ EXAMPLE

The following plot is based on clinical trial data from an approved product, and was produced with R and the CluePoints SMART engine, using a negative log scale for the vertical axis. Potentially problematic sites are shown in magenta, and lie above the 0.05 level. Most of the magenta sites in the referenced clinical trial were located outside of the US, and were not inspected by OSI; data needed for the CISST tool were not submitted for this study. A review of sponsor site audits confirmed that the two largest potentially problematic sites did in fact have issues and were not fully compliant with good clinical practices. This trial evidenced greater treatment benefit at sites external to the United States than those within the United States.

PhUSE Connect 2018



INSPECTIONS AND ASSOCIATED PROCEDURES

Once a potentially problematic site or sponsor has been identified, FDA inspectors and other FDA staff may visit a site to conduct an inspection. Inspected sites receive one of three outcomes: No Action Indicated (NAI), Voluntary Action Indicated (VAI), or Official Action Indicated (OAI). Most inspected sites received an NAI rating, followed by VAI, and OAI. OAI ratings have generally been in the 1%-3% range. FDA resources allow only a small proportion of sites to be inspected. Appropriate tools can assist OSI and OSIS inspectors to prioritize sites and allocate available resources. Additionally, statistical reviewers may use data anomaly detection tools to perform sensitivity analyses by excluding potentially suspect sites.

The ability of CDER to detect issues with clinical trial sites is improving, but challenges remain. Clinical trials for certain generic products may be conducted entirely at a single site administered by a CRO, as was the case with Semler [1]. Compensation to the CRO may be related to the results of the clinical trial, creating a potential conflict of interest between ethical conduct and CRO remuneration. In the aftermath of the Semler case, tools have been developed for use with PK/PD profiles. Additional concerns are related to trial design, data processing, and programming. Poorly designed trials may result in missing data, inaccurate data, and incomplete data. Poor source data collection procedures and poor data management processes can compromise data quality and integrity, while poor programming procedures can compromise analyses. CDER has begun requesting that sponsors submit programs necessary to trace data flow from SDTM data to ADaM analysis data, as well as those programs used to obtain efficacy endpoint analyses [9]. Data management and programming problems can adversely affect statistical analyses used for decision making, while sponsor errors and misconduct can jeopardize drug development programs.

CONCLUSION

The US FDA has taken both passive and active measures to address data quality and data integrity issues. Tools are available for assessing compliance with SDTM standards and are in development to assess data quality and integrity. However, there are continuing challenges with detecting and meaningfully addressing issues occurring at the sponsor/CRO level. Ensuring that data quality and data integrity are maintained supports appropriate and timely decision making within the FDA, while reducing costly errors and missteps for sponsors.

REFERENCES

PhUSE Connect 2018

1. Food and Drug Administration. (2016) Notification to Pharmaceutical Companies: Clinical and Bioanalytical Studies Conducted by Semler Research Are Unacceptable, <https://www.fda.gov/Drugs/DrugSafety/ucm495778.htm>.
2. International Organization for Standardization. (2011) Data Quality-Part 1, Technical Specification ISO/TS 8000-1:2011(E), <https://www.iso.org/standard/50798.html>.
3. Medicines & Healthcare products Regulatory Agency (2018) 'GXP' Data Integrity Guidance and Definitions, <https://www.gov.uk/government/publications/guidance-on-gxp-data-integrity> .
4. Society for Clinical Data Management. (2013) Good Clinical Data Management Practices, <https://www.scdm.org/publications/gcdmp/> .
5. Pharmaceutical Users Software Exchange. (2014) Guidance on Good Programming Practice, http://www.phusewiki.org/wiki/index.php?title=Good_Programming_Practice_Guidance.
6. Food and Drug Administration. (2018) E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1) Guidance for Industry, <https://www.fda.gov/downloads/Drugs/Guidances/UCM464506.pdf>.
7. Food and Drug Administration. (2018) Standardized Format for Electronic Submission of NDA and BLA Content for the Planning of Bioresearch Monitoring (BIMO) Inspections for CDER Submissions Guidance for Industry (Draft), <https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formssubmissionrequirements/ucm332466.pdf>.
8. Food and Drug Administration. (2018) Bioresearch Monitoring Technical Conformance Guide, <https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formssubmissionrequirements/ucm332468.pdf>.
9. Food and Drug Administration. (2018) Study Data Technical Conformance Guide, <https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf>.

CONTACT INFORMATION

Paul H. Schuette, Ph.D.
Food and Drug Administration,
Center for Drug Evaluation and Research,
Office of Translational Sciences,
Office of Biostatistics
10903 New Hampshire Avenue
Silver Spring, MD 20993
301-796-3838
Paul.Schuette@fda.hhs.gov