JOURNAL OF CLINICAL RESEARCH BEST PRACTICES

Vol. 3, No. 6, June 2007 "Can You Handle the Truth?"

A Risk-Based Approach for Assessing Data Quality

By Kit Howard

Introduction

There is a long-standing mantra in the world of industry-sponsored clinical trials that data must be absolutely clean before a database is locked and an analysis conducted. Tremendous time and resources are devoted to ensuring the quality of the data, both at the site and after the data are entered into the database. Data are checked for accuracy, completeness, consistency and so on, all in the pursuit of ensuring "quality data," but few organizations have defined what they mean by "data quality." Is quality an absolute or relative attribute of the data? In other words, is it possible to achieve "clean enough" data, or is the quest for quality destined to be open-ended?

In 1999, the Institutes of Medicine (IOM) conducted a roundtable workshop on data quality that defined quality data as "data that support the same interpretations and conclusions as those derived from error-free data." [1] According to this definition, error-free data is not required. It must only be "clean enough." The FDA, in its Critical Path Initiative, is encouraging all participants in clinical research to apply carefully considered risk models in order to speed the development and marketing of new drugs.[2] This recommendation was confirmed at the May 2007 FDA/DIA (Drug Information Association) meeting "Defining and Implementing Quality in Clinical Investigations from Design to Completion," where Janet Woodcock, the Deputy Commissioner for Operations at the FDA, stated that she is strongly encouraging all parties conducting clinical research to recognize that quality is a "system" characteristic, and that quality cannot be inspected into a trial but instead must be designed into the process.[3] She proposed that researchers should prospectively define the required level of quality, measure the quality achieved, and present both the plan and the results as part of the research report.[4] Thus far, little has been done to apply these principles to cleaning clinical research data after site monitoring has occurred. Combining the IOM definition, the Critical Path Initiative's perspective, and Dr. Woodcock's comments provides a good foundation for developing a risk-based approach to achieving data quality.

There is a significant body of knowledge on cleaning individual data points. Site monitoring ensures that data have been accurately transferred from source documentation to the data capture tool (e.g., CRF or EDC). Edit checks identify outlying illogical and implausible values and check for consistency between data values (e.g., comparing systolic to diastolic blood pressure values). These checks typically occur throughout the study, and are well-suited for identifying random and fairly obvious errors in the data.

However, these random errors may have little or no affect on the results of the study. By definition, statistics is the art and science of making reliable generalizations about variable data, and as random data errors are equally likely to occur in any treatment group, the main effect of these errors is to increase the variability of the data. While such errors do decrease the power of the study, they are unlikely, in moderation, to change the outcome of most studies.

The more serious threats to study data are less obvious, where individual data values appear plausible but contain systemic patterns of errors. Systemic errors can be due, for example, to site personnel misunderstanding how to administer a measure, misinterpretation of protocol requirements, or even to data falsification. In these cases,

traditional data cleaning tools will not identify the issues because each subject's data is internally consistent and accurately reflects the source documents. While some of these problems are identified by statisticians as they prepare the data for analysis, this review typically happens very late in the study, or even after database lock. If these issues were identified earlier in the study, it might be possible to take corrective action at problem sites or replace them. If the problem is occurring at multiple sites, corrective action can save the entire study.

Examples

Detecting these systemic errors requires a very different approach than traditional within-subject data checking. In this approach, data are aggregated and compared across sites, time, geography, or even monitoring or data management personnel (e.g., excessive queries being generated by one data manager). Unusual patterns can then be identified and further investigated. Many types of data lend themselves to such analyses, including subject data such as adverse events (AEs) and survey scores, and also administrative data such as visit dates, query rates and even system login times. Figures 1 through 4 illustrate some examples.

Figure 1. Unusually Few Events

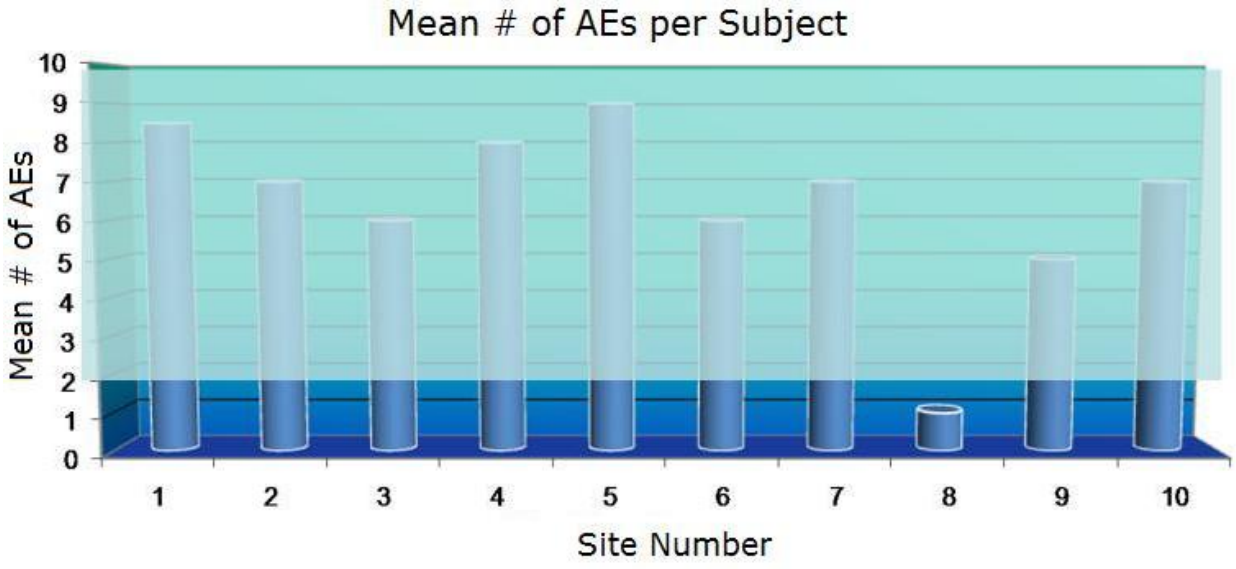

Figure 1 charts the mean number of adverse events recorded per subject at each site. Sites 1 through 10 are represented on the x axis and the means are on the y axis. The shaded band represents the mean plus and minus 2 standard deviations; values outside this range are unlikely to happen by chance. While a certain amount of variability is expected, Site 8 has a suspiciously low incidence of AEs. Perhaps the site has an unusually healthy population, or it is not querying the subjects closely enough about possible events, or it may misunderstand which events are supposed to be recorded. Regardless of the reason, it is a clue that Site 8 is different and should be investigated.

Figure 2. Unusually Low Variability

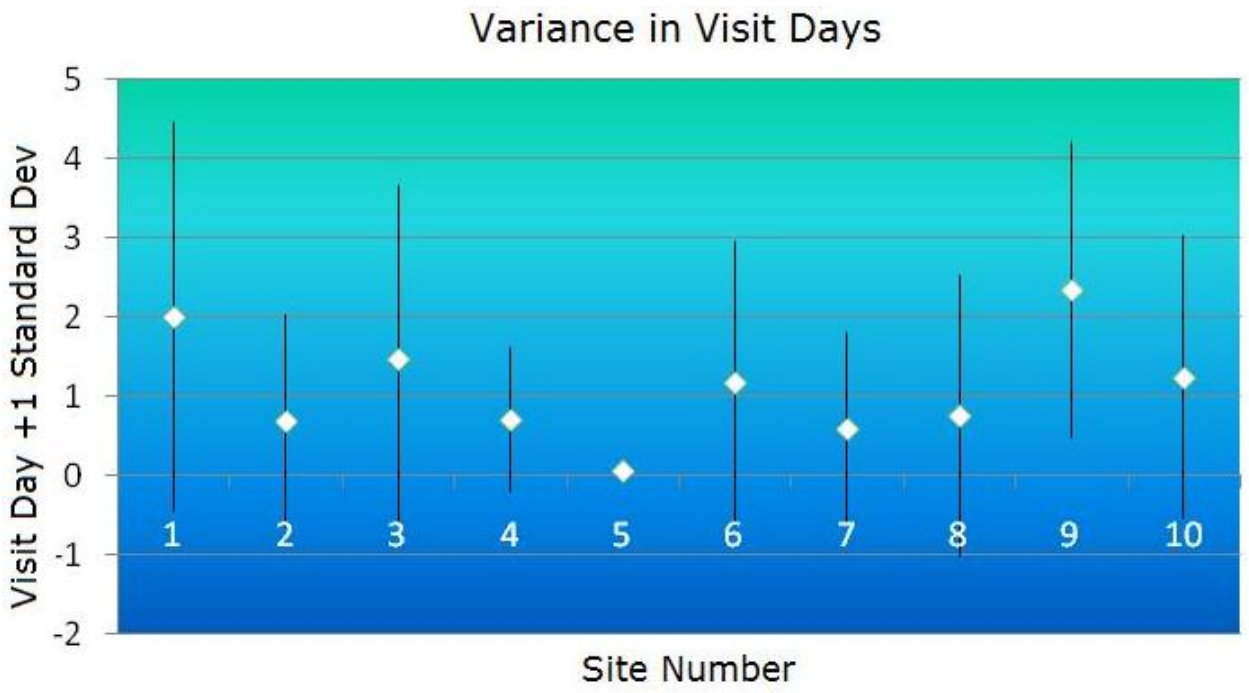

Figure 2 approaches variance from a different perspective. Based on the date of Visit 1, it looks at the difference between subsequent visit dates per protocol schedule and actual visit dates. For example, if Visit 3 was supposed to occur on Study Day 25, and the actual visit occurred on Study Day 30, the difference is +5 days. Taking the mean differences for all visits for all subjects at a site and plotting them with their standard deviations gives a good indication of how well the site adhered to the protocol schedule. If there is too much variation, the site's data may not be comparable to other sites. Too little may indicate that a site drew its subjects from an immobile population such as a retirement community, or it may suggest that the data are not reflective of what really occurred.

Figure 3. Unusually Regular Times of Data Entry

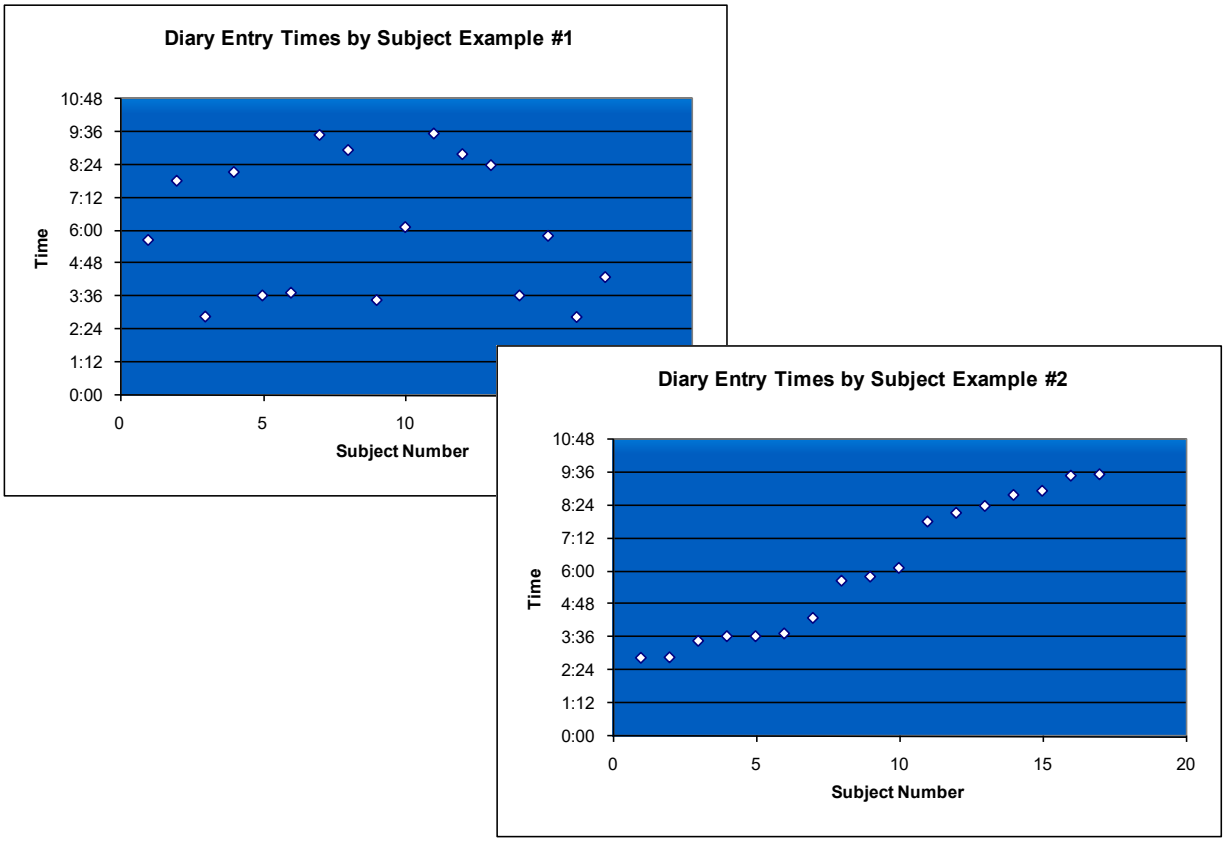

Figure 3 is a plot utilizing study "meta-data", or data about the data. In this case, it shows the times at which diary data were entered into a handheld device. The subjects are on the x axis, and the time of entry is on the y axis. In Example 1, the entry times are fairly random, whereas in Example 2, each subject's data were entered within a few minutes of the previous subject. If the subjects were supposed to enter the diary data between visits then this pattern would be suspicious, but if they completed paper diaries and the coordinator entered the data it might be expected. This example illustrates why it is important to investigate findings with an open mind and not draw conclusions prematurely.

Figure 4: Unusual Distributions of Rating Scores

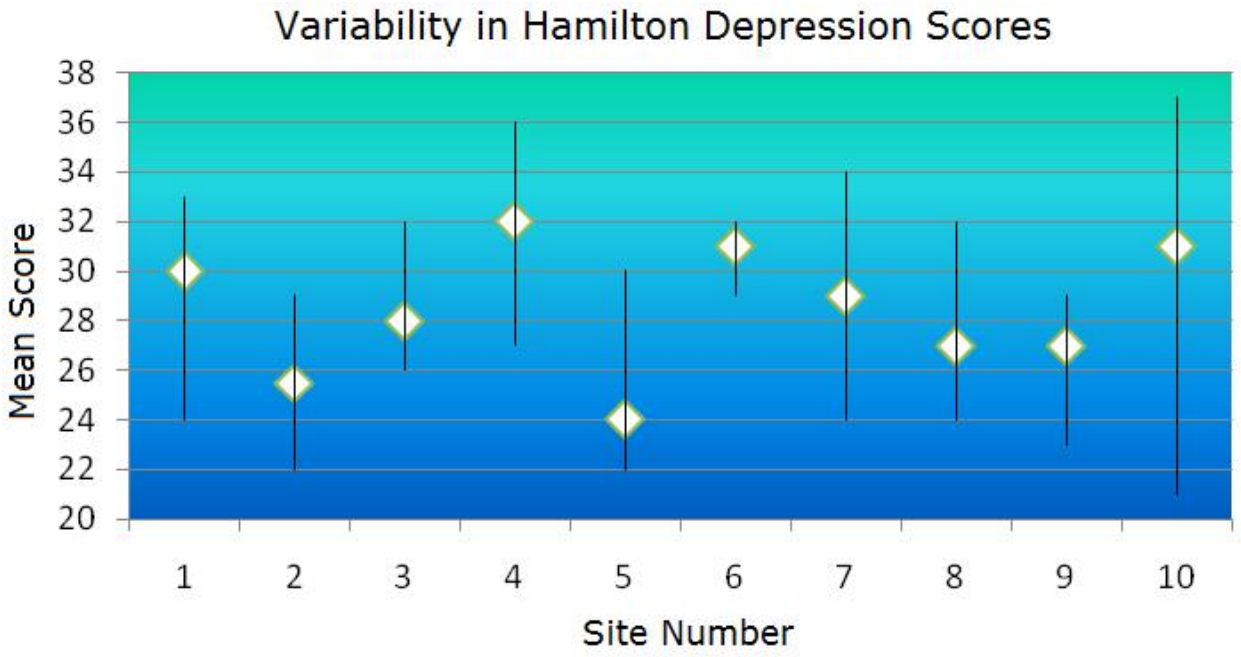

Figure 4 illustrates the range of Hamilton Depression Scale scores at baseline at ten sites. The subjects at Site 10 appear to vary widely in depression severity, whereas those at Site 6 are clustered in a very tight range. Depending on the eligibility requirements and the structure of the study hypotheses, these anomalies may or may not be an issue. Again, some variability is expected, but too much affects the power of the study. The low variability at Site 6 may suggest that the scale is not being administered correctly, or perhaps subjects are being selectively included based on their disease severity. In cases like this, the decision to investigate further is a judgment call and should be based on the importance of the data to the study (e.g., primary endpoint vs. eligibility criterion), the results of other data examinations, and any other factors that shed light on the patterns.

Implementation

Analyzing every systemic aspect of the data could rapidly become overwhelming. Each study has many data points, each of which can be evaluated many ways. When combined with traditional data-point cleaning, the cost could become prohibitive. If, instead, a risk-based approach is adopted, efforts can then be focused on detecting the issues that are potentially the most damaging.

Risk can be thought of as the combination of the likelihood of an event and its impact if the event occurs. In the context of clinical trials, the likelihood of data errors is largely dependent on the data collection methodology (e.g., paper, electronic or automatic) and upon the complexity of the data. The impact of the errors depends on the role of the data in the study and how common the errors are. The consequences can range from skewing the analyses and damaging the credibility of the study, site or sponsor, to putting subjects at greater risk of harm, to jeopardizing a study, the NDA or even multiple NDAs.

Data in clinical trials can be categorized into domains, that is, groupings of interrelated information such as adverse events, scored assessments, or lab results. Most studies require the collection of numerous domains, and not all are equally important. Some domains are primary in that they provide direct evidence for the study's central hypotheses

and for subject safety. Other domains are secondary, providing information that is analyzed but is not key to the central hypotheses. Examples might include concomitant medications or secondary efficacy measures. Finally, there are tertiary domains that are collected for administrative or other reasons, such as physical exam data, subject diaries, or non-pharmacological treatments.

It is important to note that a given domain can play different roles in different studies. For example, a depression scale may be an eligibility criterion in one study, in which case it would be tertiary, but a key efficacy parameter in another, making it primary. The same errors in the data would have very different implications in the different studies.

Figure 5: Risk Levels by Event Likelihood and Impact

Impact	Likelihood		
Domain Type	Low	Medium	High
Primary (high)	***Medium***	***High***	***High***
Secondary (medium)	***Low***	***Medium***	***Medium***
Tertiary (low)	***Low***	***Low***	***Medium***

Figure 5 suggests an approach for determining the risk of different data types and determining how to handle them. The columns represent the likelihood of errors. The rows represent the domain types as defined above. The cells identify the risk of not cleaning the data thoroughly, remembering that risk means the chance of affecting study conclusions or subject safety. At the extremes, for primary data where there is a high likelihood of error, the risk of not cleaning the data is high, whereas for tertiary data where there is a low probability of error the risk of not cleaning is low. These are obviously the easier cases, with the former demanding greater attention than the latter. The more difficult cases are where the likelihood of errors is low in primary data and high in tertiary data. If errors occur in the primary data, they are likely to have a disproportionately large effect. Conversely, the industry is generally very uncomfortable with presenting dirty data, including tertiary data, even if there is no impact on study outcomes or subject safety. For this reason, utilizing a risk-based methodology and providing clear documentation of the decisions is critical.

As is the case with most changes, there are some challenges to adopting these methods. While statisticians have used some of the above techniques to identify problems in the study database, they are not typically involved in data review early in the study. The data managers and monitors who review the data are not usually trained in statistics and are not familiar with these methods, so implementation will require a degree of cross-functional cooperation that is awkward for many companies. In addition, given that there are many data points in most studies, and many ways to examine them, implementing the above approach could rapidly become overwhelming. A risk-based methodology will assist study teams in choosing the checks that most effectively manage the risks.

Conclusion

Applying this type of risk-based approach to assessing the quality of study data has numerous benefits to both sponsors and sites. Early detection of problems allows for timely interventions to retrain, change processes, or take other actions to minimize the consequences. It permits monitors and data managers to focus their attention on quality

assurance activities that identify significant threats to subject safety or data integrity, rather than concentrating on reducing the "noise" in the data. Finally, if no problems are detected, then it significantly increases confidence in the data, while streamlining processes and reducing study costs, all of which serve to deliver more cost-effective treatments more quickly to the patients who need them.

Acknowledgement

Thanks are due to Alec Vardy of CV Therapeutics for providing the examples in Figures 2 and 4.

References

1. "Assuring Data Quality and Validity In Clinical Trials For Regulatory Decision Making", Workshop Report. Jonathan R. Davis, Vivian P. Nolan, Janet Woodcock, and Ronald W. Estabrook, Editors. Division of Health Sciences Policy, Institute Of Medicine. National Academy Press Washington, D.C. 1999
2. http://www.fda.gov/oc/initiatives/criticalpath
3. "Defining and Implementing Quality in Clinical Investigations from Design to Completion", joint FDA/DIA meeting, 10-11 May 2007, Washington DC. From personal notes.
4. Good Clinical Data Management Practices, Society for Clinical Data Management, http://www.scdm.org

Author

Kit Howard is a Principal at Kestrel Consultants, a firm specializing in the design and implementation of Data Lifecycle Plans, which are clinical data and process standards that span protocol development through clinical study reports. Contact her at 1.734.576.3031 or kit@kestrelconsultants.com.