Chapter 10
Data Quality in Clinical Research

Meredith Nahm

Abstract Every scientist knows that research results are only as good as the data upon which the conclusions were formed. However, most scientists receive no training in methods for achieving, assessing, or controlling the quality of research data—topics central to clinical research informatics. This chapter covers the basics of collect and process research data given the available data sources, systems, and people. Data quality dimensions specific to the clinical research context are used, and a framework for data quality practice and planning is developed. Available research is summarized, providing estimates of data quality capability for common clinical research data collection and processing methods. This chapter provides researchers, informaticists, and clinical research data managers basic tools to plan, achieve, and control the quality of research data.

Keywords Clinical research data • Data quality • Research data collection • Processing methods • Informatics • Management of clinical data • Data accuracy

Clinical Research Data Processes and Relationship to Data Quality

Data quality is foundational to our ability to human research. Data quality is so important that an Institute of Medicine report [1] was written on the topic. Further, two key thought leaders in the quality arena, W. E. Deming and A. Donabedian, specifically addressed data quality [2–4].

Failing to plan for data quality is an implicit assumption that errors will not occur. Emphasizing that failing to plan for data quality further threatens data quality by inhibiting the detection of errors when they do occur, Stephan Arndt et al. state,

M. Nahm, Ph.D.
Informatics, Duke Translational Medicine Institute, Duke University,
2424 Erwin Road, Durham, NC 27705, USA
e-mail: meredith.nahm@duke.edu

R.L. Richesson, J.E. Andrews (eds.), *Clinical Research Informatics*, Health Informatics,
DOI 10.1007/978-1-84882-448-5_10,

"Ironically, there is a major difference between a process that is presumed through inaction to be error-free and one that monitors mistakes. The so-called error-free process will often fail to note mistakes when they occur" [5].

Quality is broadly defined by Juran as fitness for use [6]. Unfortunately, for clinical investigators and research teams, the use varies from study to study. In clinical research, data collection processes are often customized according to the scientific questions and available resources, resulting in different processes for individual studies or programs of research. Because data quality assurance and control are largely dependent on how data are collected and processed, they are complicated by this mass customization. (The label *mass customization* used to describe clinical research by Karen Koh in a meeting at Duke Clinical Research Institute.) Given the likely persistence of science-driven customization, an antidote may lie in methods for data quality planning. It is only when a planning framework exists and is used that knowledge gained from work on prior projects can translate to new projects with different data sources, processes, and people.

The types of data collected in clinical research include data that are: manually abstracted or electronically extracted from medical records, observed in clinical exams, obtained from laboratory and diagnostic tests, or from various biological monitoring devices, and patient-reported items. Each data source is associated with a method by which the data were acquired. After acquisition, these data are subject to further processing. Whether data are collected specifically for a research project, or whether data collected for other purposes are used, a data quality plan should take into account the data source, precollection processing, the data acquisition method, and, finally, postprocessing. While these elements of the data quality plan apply regardless of where the data were collected, the data sources will likely influence the plan. In other words, one method does not fit all. Using the same method to treat all data will overlook both errors and opportunities to prevent them. For example, data recorded on a form may be retrospectively abstracted from medical records, may be written directly onto the form by the patient, or may be recorded directly on the form by a provider during a study visit. Each of these data acquisition processes is subject to different sources of error and, therefore, may benefit from different error prevention or correction methods, thus the need to take into account the data source, precollection processing, data acquisition, and postprocessing in data quality planning. This chapter is primarily concerned with how to accomplish this and will give the reader a framework to use to assure and control quality regardless of the data source, acquisition method, or processing.

Similar to the decreased property value of a house with a serious foundation problem, it is no surprise that research conclusions are only as good as the data upon which they were based. As plans and construction of a house help determine quality, well-laid research protocols must address data quality considerations, for example, by specifying a consistent suitable collection method, planning interrater reliability assessments for subjective assessments, or other collection of independent data. The resulting degree of data quality affects how data can be used and, ultimately, the level of confidence that can be reposed in research findings or other

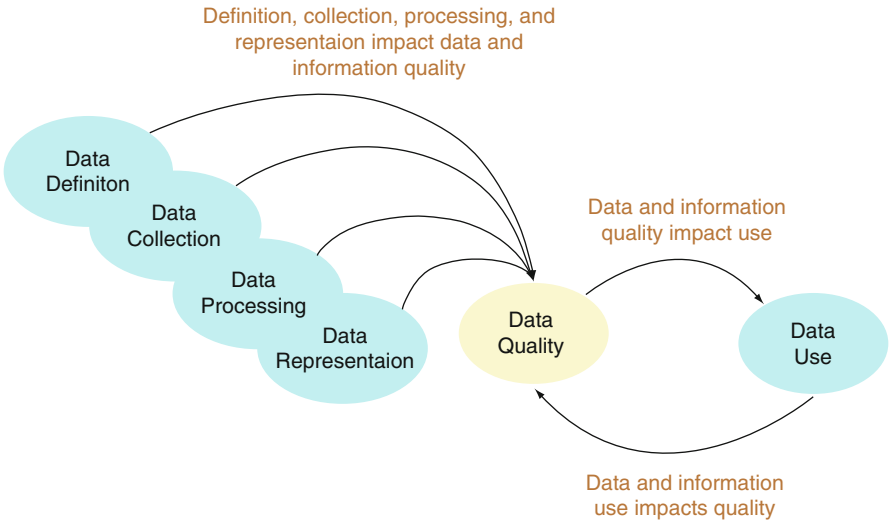

Fig. 10.1 Impacts of data generation and handling features on data and information quality. The way data and information are handled impacts the quality of that data and information. The quality of data and information impacts our willingness and ability to use it. Use of data and information causes more care to be taken in their handling, increasing the quality

decisions based on the data. Thus, protocol and Case Report Form (CRF) design, including data capture methods, must be concerned with data quality assurance measures from the start.

Data quality and the discipline of informatics are inextricably linked. Data definition, collection, processing, representation, and use are central to informatics (Fig. 10.1). Definition, collection, processing, representation, and use impact data and information quality, and data and information quality impact use. In turn, data and information that are used are more likely to have higher quality. In clinical research, data can be collected both prospectively and retrospectively, depending on the protocol and local procedures at the clinical investigational site. Therefore, information use in clinical care as well as information use in the study may impact data quality.

Each step in the collection, handling, and processing of data affects data quality. International Conference on Harmonization (ICH) guidelines state, "Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly" [7]. We suggest a less literal interpretation of the ICH E6 guidance document. The gold standard in achieving quality is prevention rather than after-the-fact finding and fixing errors; thus, interventions aimed at preventing errors are typically designed into data collection and handling processes, i.e., part of the process rather than an after-the-fact checking activity applied to a data handling step. Similarly, methods for monitoring data quality are built into data collection and handling processes.

Assuring and controlling data quality are largely a focus on the presence of *data errors*, defined here as a data value that does not accurately reflect the true state of the thing being represented. Data represent things (or states of things) in the real world. Since all things change with time, so does the accuracy with which a data value represents the true state of the represented thing. Thus, data that are correct will not necessarily remain so with the passage of time. The qualifiers necessary for a given datum to remain accurate over time are often referred to as context, for example, patient age as of the first study visit; or air temperature in degrees Celsius at latitude 35.620252°N, longitude −82.542933°W, at an elevation of 2,310 ft at noon on May 23, 2009; or medications taken within the 10-day time window before the blood draw (see discussions of reliability and validity in Chaps. 4 and 11). The use of a broader definition than "inaccuracies created in data processing," or "nonconformance to data specifications," is intentional because inaccuracies from any source may render data values incorrect. Data quality can be compromised at any point along the continuum of data collection and processing, as demonstrated by the following examples adapted from actual cases. In this chapter, we develop and apply a framework for preventing and controlling data errors in the context of clinical research. The following examples come from the Society for Clinical Data Management [8].

Example 1
A large multisite clinical trial was sponsored by a pharmaceutical company to obtain marketing authorization for a drug. During the final review of tables and listings, an oddity in the electrocardiogram (ECG) data was noticed. The mean heart rate, QT interval, and other ECG parameters for one research site differed significantly from those from any other site; in fact, the values were similar to ones that might be expected from rather than human subjects. The data listed on the data collection form were checked and were found to match the data in the database, thereby ruling out data entry error; moreover, there were no outliers from that site that would have skewed the data. After further investigation, it was discovered that a single ECG machine at the site was the likely source of the discrepant values. Unfortunately, the site had been closed, and the investigator could not be contacted.

Example 2
In the course of a clinical research study, data were single entered at a local data center into a clinical data management system. During the analysis, the principle investigator noticed results for two questions that seemed unlikely. The data were reviewed against the original data collection forms, and it was discovered that on roughly half of the forms, the operator entering the data had transposed "yes" and "no." Closer examination failed to identify any characteristics particular to the form design or layout that might have predisposed the operator to make such a mistake; rather, the problem was due to simple human error, possibly from working on multiple studies with differing form formats.

Example 3
A clinical trial of subjects with asthma was conducted at 12 research sites. The main eligibility criterion was that subjects must show a certain percentage increase in peak expiratory flow rate following inhalation of albuterol using the inhaler provided in the drug kits. Several sites had an unexpectedly high rate of subject eligibility compared with other sites. This was noticed early in the trial by an astute monitor, who asked the site staff to describe their procedures during a routine monitoring visit. The monitor realized that the high-enrolling sites were using nebulized albuterol (not permitted under the study protocol), instead of the albuterol inhaler provided in the study kits for the eligibility challenge. Because nebulized albuterol achieves a greater increase, these sites enrolled patients who would not otherwise have been eligible. Whether due to misunderstanding or done deliberately to increase their enrollment rate (and financial gain), the result was the same: biased and inaccurate data.

Each of these scenarios describes a data quality problem, one in device-based data collection, one in data processing, one in measurement procedure. Despite the differences in setting and in the sources of the errors, the end result was the same: inaccurate data.

The 1999 Institute of Medicine Report [1] emphasized the importance of data quality to regulatory decision-making, i.e., drawing conclusions from clinical trials. At the time, there was little in the literature base to synthesize in the report. Further, since the IOM report, there has been scant methodological progress toward data quality assurance, assessment, and control in clinical research. The framework presented here draws from a synthesis of experience and first principles.

Errors Exist

Errors occur naturally by physical means and human fallibility. Some errors cannot be prevented or even detected, for instance, a study subject who deliberately provides an inaccurate answer on a questionnaire or a measurement that is in range but due to calibration drift or measurement error. Nagurney reports that in a recent study, up to 8% of subjects could not recall historical items and up to 30% gave different answers on repeat questioning [9]. A significant amount of clinical data consists of information reported from patients. Further, as Feinstein eloquently states,

> In studies of sick people, this [data accuracy] problem is enormously increased because (1) the investigator must contemplate a multitude of variables, rather than the few that can be isolated for laboratory research; (2) the variables are often expressed in the form of verbal descriptions rather than numerical dimensions; (3) the observational apparatus consists mainly of human beings, rather than inanimate equipment alone [10].

Even with clinician observation, reading test results, or interpreting images, human error and variability remain as factors. Simply put, where humans are involved, human error exists [11]. For most types of assessment, observation, or interpretation of test results, reports of error or agreement rates can be found in the literature. These known and real errors and inconsistencies are often not accounted for in data quality planning in clinical research.

Moreover, in every process, nature affects every project every day, increasing disorder. As time passes, natural forces cause machines to wear, settings to drift, and attention to wander. Thus, while measurements and processes capable of achieving the desired levels of quality are often sought and employed in a research project, energy and vigilance must continuously be applied to maintain them.

Natural laws, logic, and empirical evidence together suggest that it is unwise to assume any data set is truly error-free. Still, respondents to a data quality survey conducted by the author [12] and others [13] noted perfect data as their acceptance criterion. References to fear of consequences from regulators and potential data users observing obvious errors [1], such as a diastolic blood pressure of 10, suggest that the real concern may be the doubt that a user-discovered data error casts on the rest of the data set. Such concern should be taken into account in data quality planning; for example, many organizations perform a review of blinded tables, listings, and figures prior to closing a database, to identify such obvious errors. The concern of obvious errors discrediting a data set will likely increase with more public data sharing, so methods such as looking at descriptive statistics, outliers, frequencies, and distribution graphs to efficiently scan a data set will persist.

It is important to note that cleaner data can save time in programming and data use, but this is likely concomitant with additional costs. As such, and within the context of a given research project, pursuing data quality to a greater extent than needed to support the conclusions is unnecessary. Thus, data quality plans must be informed by the necessary level of data quality and must target the necessary level of data quality in the most cost effective way. Two questions naturally result from this line of thought:

1. How clean do the data need to be to support the intended analysis?
2. What is the best method, given the study context, to achieve this?

The first is a statistical question, and the second is for the experienced informaticist to explore.

Defining Data Quality

The Institute of Medicine (IOM) defines quality data as “data strong enough to support conclusions and interpretations equivalent to those derived from error-free data” [1]. Like Joseph Juran’s famous “fitness for use” definition [6], the IOM definition is use dependent. Further, the robustness of statistical tests and decisions to

data errors differs. Thus, applying the IOM definition requires a priori knowledge of how a statistical test or mode of decision-making behaves in the presence of data errors. For this reason, in clinical research, it is most appropriate that a statistician set the acceptance criterion for data quality.

Further specification of the IOM definition of data quality is necessary for operational application. Other authors who have discussed data quality define it as a *multidimensional concept* [14–20]. In clinical research, the dimensions most commonly considered are *reliability*, *validity*, *accuracy*, and *completeness* [21]. Reliability and validity address the underlying concept being measured, i.e., is this question a reliable and valid measure of depressive mood? Accuracy is important with respect to and intrinsic to the data value itself. For example, does a heart rate of 92 represent the patient's true heart rate at the time of measurement? That is, *is it correct?* And completeness is a property of a set of data values; i.e., *are all of the data there?* More recently, as research methods have matured and data are increasingly used for monitoring and decision-making during the trial (as in the case of data and safety monitoring boards), *timeliness* has emerged as an important data dimension. Further, regulatory authorities are concerned with trustworthiness of the data and initially identified the following data quality dimensions for clinical research: "electronic source data and source documentation must meet the same fundamental elements of data quality (e.g., attributable, legible, contemporaneous, original, and accurate) that are expected of paper records and must comply with all applicable statutory and regulatory requirements" [22].

These "fundamental elements," *attributable*, *legible*, *contemporaneous*, *original*, and *accurate*, are commonly referred to as ALCOA. Registries commonly report data quality in terms of accuracy and completeness [23]. As secondary use of data has grown, so has the need for data to be *specified*, *accessible*, and *relevant*. Similarly, the dimension of *volatility*, or how quickly the data change, becomes a concern; for example, studies in adult populations seldom collect height at annual study visits, but studies in pediatric populations are likely to do so. These fundamental dimensions are attributes, or descriptors of data quality, allowing users, especially secondary users, to evaluate the likelihood that data will support their specific (secondary) use. As we begin to see an increase in secondary, particularly research, uses of clinical data, the need for fundamental dimensions of data quality will become a necessary data itself.

The multidimensionality data quality causes ambiguity because any given use of the term might refer to a single dimension or to a subset of possible dimensions. Further, different data users may emphasize some dimensions while excluding others; for instance, the information technology (IT) sector tends to assess data quality according to conformance to data definitions stated business rules, while regulatory authorities are concerned with attribution and verifiability [22]. Although accuracy and completeness historically have been emphasized in the clinical research literature, multiple dimensions ultimately affect and determine the usefulness of data. Each individual dimension describes an element of quality that is necessary but usually not sufficient for data to be useful for their intended purpose.

When maintained as metadata, can be used to assess the quality of the data for primary and secondary uses.

All dimensions apply to any use of data, but often the circumstances surrounding a given (or the primary) use include built-in processes that assure a relevant dimension is present and addressed. For example, in a clinical trial, those who use data often have a role in defining it, meaning the *definition* is of little concern. However, when data are considered for secondary uses, such as a pooled analysis spanning a number of studies, *relevance* and *definition* become primary concerns. By employing a dimension-oriented approach to data quality, these assumptions become transparent, helping us to avoid overlooking important considerations when working in new situations. In other words, carving data quality up into dimensions helps us design for, measure or assess, control, and increase data quality. A consensus set of dimensions for clinical research does not yet exist. Here, we will primarily address the dimensions of *accuracy*, *completeness*, *timeliness*, *accessibility*, *relevance*, *and volatility*. *Reliability* and *validity* are addressed in Chaps. 4 and 11, as noted, and data *definition* (full specification) is addressed in Chap. 13.

Using multiple dimensions to characterize data quality, and measuring those dimensions to assess data quality, requires both operational definitions and acceptance criteria for each dimension of quality. An approach that will allow collaboration across studies and domains includes standard operational definitions for dimensions, with project-specific acceptance criteria. For example, *timeliness* can be operationally defined as the difference between the date a given set of data is needed and the actual date it is available. The acceptance criterion—"How many minutes, days, or weeks late is too late?"— is set based on study needs. Further, some dimensions are inherent in the data, i.e., characteristics of data elements or data values themselves, while others are context dependent. Table 10.1 contains common clinical research data quality dimensions, labels each dimension as inherent or context sensitive, labels the level at which it applies, and suggests an operational definition.

Framework for Data Quality Planning

Over the past decade or more, the number and diversity of both new technology and new data sources have increased. Managing new technology or data sources on a given project is now a normal aspect to clinical research data management. One of the largest problems is preparing data managers to work with new technology and data sources. Simply put, a framework is needed that will enable data managers to assess a given data collection scenario, including new technology and data sources, and systematically evaluate that scenario, apply appropriate methods and processes, and achieve the desired quality level.

A dimension-oriented approach provides a framework that practitioners can rely on when handling data in a novel situation (e.g., data from a different source, in a

Table 10.1 Data quality dimensions for clinical research

Dimension	Type	Natural language definition	Operational definition/ metric
Accuracy	Inherent	*States in the data match the intended state in the real world*	Number of errors divided by number of fields inspected (implies comparison with gold standard)
Currency	Inherent	Length of time a data value has been stored (since last update)	Use/need date minus date data last updated
Completeness	Inherent	*The extent to which every represented real-world state is reflected in the data*	Number of missing values divided by number of fields assessed
Consistency (internal)	Inherent	Data values representing the same real-world state are not in conflict	Number of discrepant values divided by number of values subject to data consistency checks
Timeliness	Context dependent	*Length of time from a change in the real-world state to the time when the data reflect the change*	Data need date minus date data ready for intended use
Relevance	Context dependent	Data can be used to answer a particular question	Percentage of data values applicable to intended use
Granularity	Context dependent	Level of detail captured in data	Percentage of values at level of detail appropriate for intended use
Specificity (nonambiguity)	Inherent	*Each state in the data definition* (metadata) *corresponds to one* (or no) *state of the real world*	Number of values with full ISO 11179 metadata including definition divided by number assessed
Precision	Context dependent	Number of significant digits to which a continuous value was measured (and recorded); for categorical variables, the resolution of the categories	Percentage of values with precision appropriate for intended use
Attribution	Inherent	Source and individual generating and updating data are inextricably linked to data values	Percentage of data values linked to source and user ID of individual who generated and changed record

Italicized wording quoted from Wand and Wang [18]

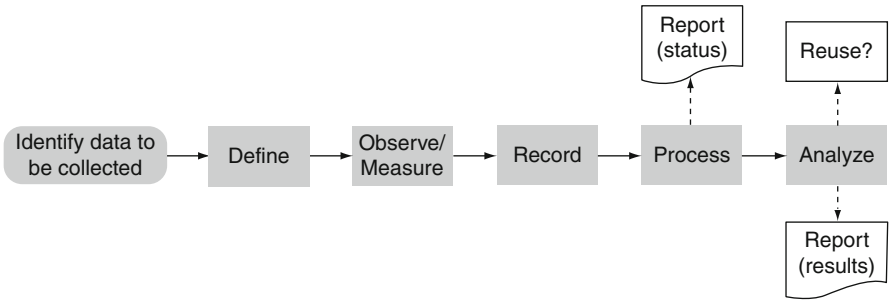

Fig. 10.2 Data-centric view of the research process. A set of general steps for choosing, defining, observing, or otherwise measuring, recording, analyzing, and using data apply to almost all research (From Data Gone Awry [8], with permission)

new environment, or using new technology). Such a framework helps guard against methodological omissions and assures that data will meet specified needs. However, data quality dimensions alone are an incomplete solution. A systematic way to assess data sources and processes on a project is necessary. Figure 10.2 shows the set of steps comprising the data-related parts of the research process. These steps are described at a general level so that they can be applied to any project. From the data-oriented point of view, the steps include: (1) identifying data to be collected, (2) defining data elements, (3) observing and measuring values, (4) recording those observations and measurements, (5) processing data to render them in electronic form and prepare them for analysis, and (6) analyzing data. While research is ongoing, data may be used to manage or oversee the project. After the analysis is completed, results are reported, and the data may be shared with others.

Identifying and Defining Data to Be Collected

Identifying and defining the data to be collected are critical aspects of clinical research. Data definition initially occurs as the protocol or research plan is developed. Too often, however, a clinical protocol reads more like a shopping list (with higher-level descriptions of things to be collected, such as *paper towels*) than a scientific document (with fully specified attributes such as *brand name*, *weight*, *size of package*, *and color of paper towels*). When writing a protocol, the investigator be as specific as possible because in multicenter trials, the research team will use the protocol to design the data collection forms. Stating in the protocol that a pregnancy test is to be done at baseline is not sufficient—the protocol writer should specify the sample type on which the test is to be conducted (e.g., pregnancy test is to be performed on women of childbearing potential).

As standards such as the Protocol Representation Standard [24] mature and supporting software becomes available, full specification of protocol elements will become the most efficient method for defining data, as metadata specified in the

protocol will be immediately available for generation of data collection forms. (See Chap. 9) Lack of specificity in data definition is the mechanism by which data identification and definition can cause serious data quality problems, for example, two sites using different measurement methods, or not measuring the same construct. The information necessary to fully specify a clinical measurement, with context sufficient to remove ambiguity, differs based on the type of data. For example, specification of the specimen (and often, the method by which the specimen is obtained) is important for some tests. For blood pressure measurements, the position, location of measurement, and device used may be important. Without careful identification and specification of this context, data collectors at clinical sites may inadvertently introduce unwanted variability.

The principle of "Occam's razor" applied to clinical research suggests that it is necessary only to collect the data needed to assure patient safety, answer the scientific question(s), and uniquely identify the collected data elements. Jacobs and Studer report that for every dollar spent to produce a data collection form, $20–$100 are required to fill each one in, process it, and store it, emphasizing that "the true cost of a form involves people not paper" [25]. When extensive data cleaning is required, this ratio becomes even more exaggerated. Eisenstein and colleagues report extensive cost savings in clinical trials by decreasing the number of data collection form pages [26, 27]. At the time of this writing, the relationship between form length and data accuracy for online forms remains unprobed [28]. Further, the evidence relating form length to decreased response rate while considered equivocal by some [28] has been demonstrated in controlled and replicated experiments [29, 30]. There is no question, however, that collecting more data increases costs and places additional burden on clinical investigational sites and data centers [26, 27].

These two principles, parsimony in the number of data elements collected, and full specification of those that are collected, are preventative data quality interventions. Parsimony, or lack thereof, may impact data accuracy and timeliness dimensions, while data definition impacts the specificity dimension and significantly impacts secondary data users.

Defining Data Collection Specifications

The previous section covered the definition and specification of data elements themselves. This section covers definition of the tools, often called data collection forms or case report forms, for acquiring data. The design of data collection forms, whether paper or electronic, directly affects data quality. Complete texts have been written on form design in clinical trials, (see Data Collection Forms in Clinical Trials by Spilker and Schoenfelder (1991) Raven Press NY). There are books on general form design principles, for example, Jacobs and Studer (1991) Forms Design II: The Complete Course for Electronic and Paper Forms. In addition, the field of usability engineering and human-computer interaction has generated many publications on screen or user interface design. A good introductory work is Shneiderman and

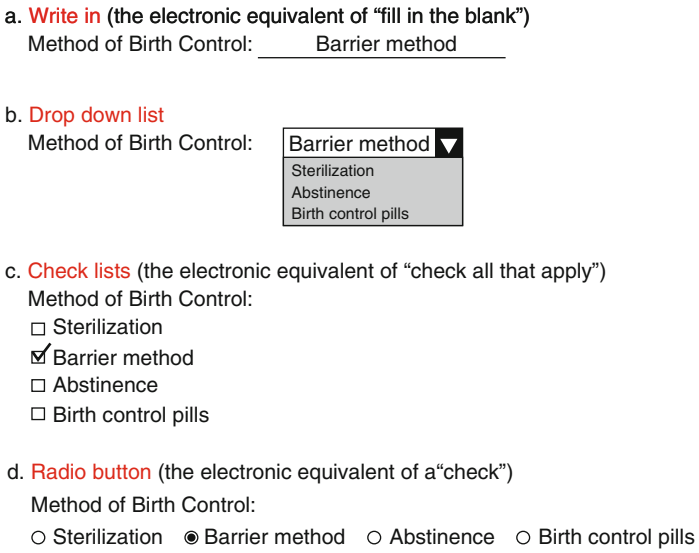

Fig. 10.3 Example data collection structures. For many data elements, more than one data collection structure exists

Plaisant (2004) Designing the User Interface: Strategies for Effective Human-computer Interaction. While this topic is too broad to discuss in depth here, two principles that are directly relevant to clinical research informatics, and not covered in more general texts, warrant attention here. The first is the match between the type of data and data collection structure; the second is the *compatibility-proximity principle* [31]. A general assumption is that the more structured the data, the higher the degree of accuracy and ease of processing. We will see, however, that this can be counterbalanced by considerations related to ease of use.

As a general principle, the data collection structure should match the type data. Data elements can be classified according to Stevens' scales (nominal, ordinal, interval, and ratio) [32], or as categorical versus continuous. Likewise, classification can also be applied to data collection structures describing how the field is represented on a form, including: verbatim text fill in the blank, drop-down lists, check boxes ("check all that apply"), radio buttons ("check one"), and image maps. Examples of data collection structures are shown in Fig. 10.3.

Mismatches between data type and collection structure, for example, collecting data in a structure more or less granular than reality, can cause data quality problems. Collecting data at a more granular structure than exists or than can be discerned in reality, for example, 20 categories of hair color, invites variability in classification. Collecting data at a less granular structure, *data reduction*, than can be discerned in reality also invites variability and results in information loss. The real granularity cannot be resolved once the data are lumped together into the categories. For example, if height is collected in three categories, short, medium and

tall, the data cannot be used to answer the question, "how many subjects are over 6 feet tall?" Another way to think about data reduction is in terms of Steven's scales. Data are reduced through collection at a lower scale, for example, collecting a yes or no indicator for high cholesterol. When the definition of high cholesterol changed, data sets that collected the numerical test result continued to be useful, while the data sets that contained an indicator, yes or no to high cholesterol, became less useful. There are many cases such as high-volume data collected through devices where reduction in the number of data values collected or retained or stored is necessary and desirable. The amount of information loss is dependent on the method employed. Reduction of CRF data is through both data collection at a lower scale than the actual data and through decision not to collect certain data values. Because data reduction results in information loss, it limits reuse of the data and should only be employed after careful deliberation.

Data collection structure can cause quality problems in capturing categorical data in other ways. When the desired response for a field is to mark a single item, the available choices should be exhaustive (i.e., comprehensive) and mutually exclusive [33–35]. Lack of comprehensiveness causes confusion when completing the form, leading to unwanted variability. Similarly, overlapping categories cause confusion and limit reuse of the data.

The *compatibility-proximity principle* was first recognized in the field of cognitive science. When applied to the design of data collection forms, it means that the representation on the form should as closely as possible the cognitive task of the person completing the form. For example, if body mass index (BMI) is a required measurement, but the medical record captures height and weight, the form should capture height and weight, and the BMI should be calculated by a computer. Sometimes, this principle is stated as "collect raw data." Values on the form should allow data to be captured using multiple units so that the person completing the form is not required to convert units. Importantly, the flow of the form should follow as closely as possible the flow of the source document [33–35]. An additional application of the compatibility-proximity principle is that all items that the person completing the form needs to complete his or her task should be immediately apparent on the form itself (separate form completion instruction booklets are less effective) [34]. There is evidence that data elements with higher cognitive load on the abstractor or form completer also have higher error rates [35–47]. Adhering to the compatibility-proximity principle, by decreasing cognitive load, may help prevent this.

There are, however, four countervailing factors that must be weighed against the compatibility-proximity principle: (1) for projects involving multiple sites, matching aspects of each site's medical record in the data collection form, representation may not be possible; (2) there may be reasons for using a more structured data collection form that outweigh the benefits of precisely matching the medical record; (3) in circumstances where a calculated or transformed value is necessary for immediate decision-making at the site, a real-time solution or tool to support the additional cognitive tasks is needed; such a tool may require raw data as input; and (4) it may not be possible to design forms that match clinical workflow, for example,

some electronic systems limit data collection structure to one question-answer pair per line, precluding collection of data using tabular formats.

Defining data collection is not limited to the data collection structure. It also includes the source and means by which the data will be obtained. For example, will data be abstracted from medical records, collected de novo from patients directly, or collected electronically through measuring devices? The identification of possibilities and selection of one over the alternatives is a design decision requiring knowledge of the advantages and disadvantages of each option and how they impact costs and the dimensions of data quality. Thus, ability to characterize data sources and processes in these terms is a critical competency of clinical research informaticists.

Like parsimony and full specification, defining the data collection mechanism is a preventative data quality intervention. The chosen data sources and mechanisms of collection and processing may impact data accuracy, precision, and timeliness dimensions, while the definition itself may impact the specificity dimension and the utility of data for secondary uses.

Observing and Measuring Data

The different types of measurement and observations used in clinical research are too many and too various to enumerate here. Clinical data may be reported by the patient, observed by a physician or other healthcare provider, or measured directly via instrumentation. Some measurements return a concrete number (e.g., temperature) or answer, while others require interpretation (e.g., the trace output of an electrocardiogram).

It is difficult (and sometimes impossible) to correct values that are measured incorrectly, biased, or gathered or derived under problematic circumstances. Recorded data can be checked to ascertain whether they fall within valid values or ranges and can be compared with other values to assess consistency, but doing so after the data have been collected and recorded eliminates the possibility to correct errors in observation. For this reason, error checking processes should be built into measurement and observation whenever feasible. This can be accomplished by building redundancy in to data collection processes [48, 49]. Some examples include: (1) measurement of more than one value (e.g., taking three serial blood pressures), (2) drawing an extra vial of blood and running a redundant assay for important measurements, (3) asking a different question to measure the same construct, and (4) measuring the same parameter via two independent methods. Immediate independent measurement with immediate feedback can be used to identify and correct discrepancies at the point of measurement. Independent measurement alone can also provide a replacement value if needed (e.g., the second vial of blood that saves the day when the first vial hemolyzes). Independent assessment with immediate feedback should be distinguished from error checking with immediate feedback. Error checking is a comparison of a recorded value

against a known standard, for example, valid ranges, or relative comparison to another value. While error checking can identify some errors, it will miss those within the valid value set. Errors within the valid value set can only be identified through redundancy. Secondly, error checking may occur at the point of measurement or recording, but is usually not built in to measurement processes, and thus occurs after the fact, and only serves as an identification mechanism, rather than as a correction mechanism. In summary, measurement discrepancies can be mitigated through careful procedures and training; however, errors are nonetheless inevitable. While error checking near or after measurement can identify errors, immediate independent verification with contemporaneous feedback remains the safest option.

Another important aspect of measurement and observation, one that has a critical effect on data quality, is ensuring consistency between or among clinical investigational sites. The "albuterol" example given at the beginning of the chapter reflects an all-too-common problem rooted in the fact that clinical investigational sites each practice medicine and research differently and institutional policies vary from location to location. In addition, equipment may vary from site to site, and there is usually at least some degree of staff turnover during studies, meaning that levels of available skill, knowledge, and experience at a given site will fluctuate over time. These and other factors contribute to variations in procedures governing observation and measurement, adding unwanted variability to clinical data.

For these reasons, clear, unambiguous, and uniform procedures that all study personnel can follow are essential to maintaining data quality. Consistency can often be improved by providing sites with critical study-related equipment or devices (so that all study data are being gathered with the same devices), training site personnel in study procedures and the administration of tests and questionnaires, using central reading centers where rating or interpretation of data is required, and requiring all sites to follow equipment calibration schedules that offer preventative methods to improve data quality from measurement and observation.

Measurement and observation should also be subject to ongoing assessment and control. Some methods directly assess the measurement or observation; examples include assessing interrater reliability, reviewing recorded interviews, and monitoring investigational sites for adherence to procedure are all ways of providing ongoing assessment and control. While other assessment and control methods are indirect, examples include counts of data inconsistencies, instances of noncompliance to protocol specified time windows, and statistical methods of checking for aberrant by site. These indirect methods may identify sites or study staff that may be performing aspects of the study differently from other sites. However, these indirect measures are only surrogates for data quality, i.e., measures of inconsistency, rather than direct assessment of accuracy. With such indirect assessments, care must be taken to respect natural variations (including those caused by variations in population) among sites. Assessment and control methods are usually targeted at the accuracy, timeliness, or completeness dimensions.

Recording Data

Recording data is the process of writing down (e.g., as from a visual readout or display) or directly capturing electronically data that have been measured, thereby creating a permanent record. The first time a data value is recorded—whether by electronic means or handwritten, on an official medical record form, or a piece of scratch paper, by a principal investigator or anyone else—*is considered the source* [7]. If questions about a study's results arise, the researcher (and ultimately, the public) must rely upon the source to reconstruct the research results. Several key principles are applicable: (1) the source should always be clearly identified; (2) the source should be protected from untoward alteration, loss, and destruction; and (3) good documentation practices, as described by US Food and Drug Administration regulations codified in 21 CFR Part 58 [50], should be followed. These practices include principles such as data should be legible, changes should not obscure the original value, the reason for change should be indicated, and changes should be attributable (to a particular person). While it seems obvious that the *source* is foundational, even sacred to the research process, cases where the source is not clearly identified or varies across sites have been reported and are common [51, 52]. Data quality is also affected at the recording step by differences such as the recorder's degree of fidelity to procedures regarding number of significant figures and rounding; such issues can be checked on monitoring visits or subjected to assessment and control methods discussed in the previous section. Data recording usually impacts the accuracy, timeliness, or completeness dimensions. However, where recording is not adequately specified, precision may also be impacted.

Processing Data

In a recent literature review and pooled analysis that characterized common data collection and processing methods with respect to accuracy, data quality was seen to vary widely according to the processing method used [53]. Further, it appears that the process most associated with accuracy-related quality problems, medical record abstraction, is the most ubiquitous, as well as the least likely to be measured and controlled within research projects [53].

Although not as significant in terms of impact on quality as abstraction, the method of data entry and cleaning can also affect the accuracy of data. On average, double data entry is associated with the highest accuracy and lowest variability, followed by single data entry (Table 10.2). While optical scanning methods could provide accuracy comparable to key-entry methods, they were associated with higher variability. Other factors such as on-screen checks with single data entry, local versus centralized data entry and cleaning, and batch data cleaning checks may act as substantial mediators with the potential to mitigate differences between

Table 10.2 Accuracy associated with common data processing methods

	Min.	Median	Mean	Max.	Std. Dev.	LCL	UCL
Abstraction	70	647	960	5,019	1,018	510	818
Optical	2	81	207	1,106	338	4	220
Single entry	4	26	80	650	150	21	36
Double entry	4	15	16	33	10	6	24
No batch data cleaning	2	270	648	5,019	946	200	475
Batch data cleaning	2	36	306	1,351	428	23	287

methods [53]. Additionally, other factors have been hypothesized in the literature, but an association has yet to be established, for example, staff experience [53], number of manual steps [54], and complexity of data [51]. For these reasons, measurement of data quality is listed as a minimum standard in the Good Clinical Data Management Practices document [54]. Because of the potentially significant impact that variations in data quality can have on the overall reliability and validity of conclusions drawn from research findings [55], publication of data accuracy with clinical research results should be required.

While our focus thus far has been on the accuracy dimension, data processing methods and execution can also impact timeliness and completeness dimensions. Impact on timeliness can be mitigated by using well-designed data status reports to actively manage data receipt and processing throughout the project or even prevented by designing processes that minimize delays. The impact of data processing on completeness can be mitigated in the design stages through collecting data that are likely to be captured in routine care or through providing special capture mechanisms, for example, measuring devices, capture directly from participants, or use of worksheets. Additionally, throughout the study, completeness rates for data elements can be measured and actively managed.

Analyzing Data, Reporting Status, and Reporting Results

Analyzing and reporting data differ fundamentally from other steps discussed in the preceding sections, as they lack the capacity to introduce error into the data values themselves. Errors in analysis and reporting programming or data presentation, while potentially costly, do not change underlying data. Analysis and reporting programming is typically applied to a copy of the database. However, analysis and reporting do have the potential to misrepresent the data. Assuring and controlling quality at the analysis and reporting stage is achieved through choice of appropriate methods, through validation of programming, and through applying the compatibility-proximity principle to data displays through matching the scale of represented data and representing display.

Using the Framework to Plan for Data Quality

When starting a new project, the clinical data manager and/or clinical research informaticist is faced with a design task: match the data collection scenario for the project to the most appropriate data sources and processing methods. The framework presented here can be used to structure this task to increase the transparency of decisions to the research team and lessen the likelihood that anything is missed. The first step is to group the data to be collected into categories depending on source, for example, medical history and medications will be abstracted from the medical record, blood pressures will come from a study provided device, lab values will be transferred electronically from a central lab, and so on. Where data sources within the medical record are varied, a more granular treatment may be required. The data sources and process by which the data are obtained can then be diagrammed, and alternative sources, methods, and processes can be considered. For example, some data sources may have undesirable preprocessing steps or known higher variability and may be excluded from consideration. Once the data sources have been chosen and the data gathering process has been specified, the steps in Fig. 10.4 can be applied to identify known error sources, to consider the possibility or desirability of preventing or mitigating the error, and to evaluate the methods for accomplishing the change. Data quality dimensions that are important to the research study are assessed for each type of data and each processing step. The output of this process is then discussed with the research team and incorporated into the plan for data collection and management. Importantly, application of this framework is a tool and mental exercise to use in planning and a tool to promote discussion and informed decision-making by the research team. Use of such a framework should impact the data collection and management plan, ultimately optimizing data quality. Use of this framework to produce an additional written document is explicitly not the intent (Fig. 10.4).

Infrastructure for Assuring Data Quality

Whenever organizations depend solely upon the skill, availability and integrity of individuals to assure data quality, they place themselves at risk. Levels of skill, ability, and knowledge not only differ from one person to another, but may even differ in the same person depending on circumstances (e.g., fatigue can degrade the performance of a skilled operator). Further, in the absence of clear and uniform procedures and standards, different persons will perform tasks in different ways; and while free expression is honored in artistic pursuits, it is dangerous when operationalizing research. A data quality assurance infrastructure provides crucial guidance and structure for humans who work with research data. Simply put, it assures that an organization will consistently produce the required level of data quality.

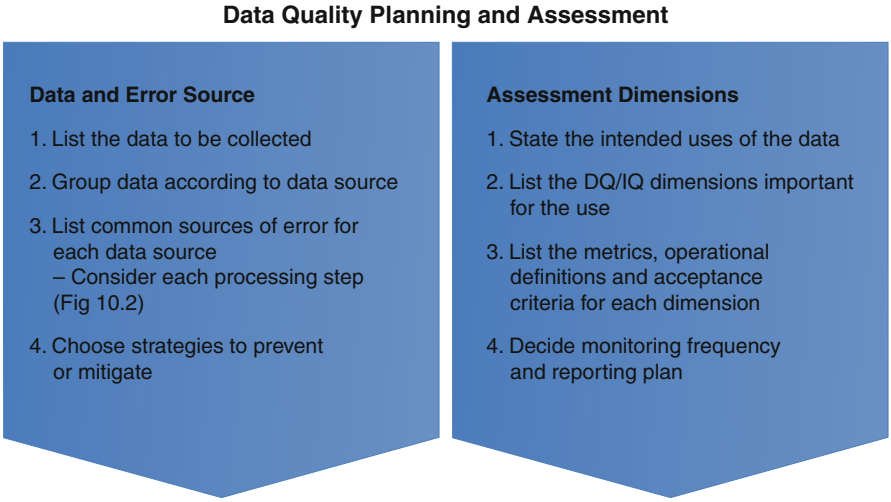

Fig. 10.4 Data Quality Planning and Assessment. This framework links data quality planning and assessment with the decisions about which data elements to collect. During planning, data to be collected are listed and grouped by type and or source of data. Known error sources for each are considered and deliberate decisions are made about prevention, mitigation, or doing nothing. At the same time, the data quality dimensions important to the intended use are identified. Metrics, acceptance or action criteria and operational definitions for each are developed as well as reporting plans. Some mitigation strategies may prompt inclusion of metrics and monitoring for known error types

The following criteria are commonly assessed in preaward site visits and audits. It is no surprise that they comprise a system for assuring data quality.

1. *Organizational consensus regarding the required level of data quality, informed by an understanding of the cost of achieving it and the consequences of failing to achieve it*
   Because the leaders of organizations or clinical trials are not typically data quality professionals, informaticists, or statisticians, data quality-related information, i.e., needs and impacts of not meeting them, may need to be communicated to leadership in a manner that can be acted upon, for example, a draft policy for approval. Where organizations exhibit inadequate support data quality, it may be because this critical information has not been conveyed to leadership in a compelling way that demonstrates the need, the associated costs, and the benefits.

2. *Appropriate tools for supporting the collection and management of data*
   Although specialized devices and software are of themselves neither necessary nor sufficient for producing quality data, their presence is often perceived as representing rigor or important capabilities. Specialized tools often automate workflow and enforce controls on the collection and processing of data. Controls

built into software are referred to as technical controls. These features can potentially increase efficiency, accuracy, and adherence to procedures by eliminating the variance associated with manual steps and options; for these reasons, data managed using automated systems are often perceived to be of higher quality. Where specialized software with these technical controls is not available, custom programming can be done to create them in available software. Other types of controls are managerial and procedural controls. These use policies, manuals of operations, and work procedures to assure consistency and quality. It is worth emphasizing that high-quality data can be achieved without specialized systems through the use of managerial and procedural controls; however, doing so often entails more highly qualified staff and additional costly manual checking and review. Where specialized technical controls are not in place, depending on the quality needed, their function may need to be developed or addressed through procedural controls.

3. *Design of processes capable of assuring data quality*
Likened to mass customization, in clinical research, scientific differences in studies and circumstances of management by independent research groups drive variation in data collection and processing. Because each study may use different data collection and management processes, the design and assessment of such processes is an important skill in applied clinical research informatics. The first step in matching a process to a project is to understand how the planned processes, including any facilitative software, perform with respect to data quality dimensions. For example, it is common practice for some companies to send a clinical trial monitor to sites to review data prior to data processing; thus, data may wait for a month or more prior to further processing. Where data are needed for interim safety monitoring, processes with such delays are most likely not capable of meeting timeliness requirements.

   Designing and using capable processes is a main component of error prevention. For this reason, clinical research informaticists must be able to anticipate error sources and types and ascertain which errors are preventable, detectable, and correctable and the best methods for doing so. Processes should then be designed to include error mitigation, detection, and correction. Process control with respect to data quality involves ongoing measurement of data quality dimensions such as accuracy, completeness, and timeliness, plus taking corrective action when actionable issues are identified. A very good series of statistical process control books has been published by Donald Wheeler. Several articles have been published on SPC applications in clinical research [55–61].

4. *Documented standard operating procedures (SOPs) are required by FDA regulation and in most research contracts.*
The complete data collection and management process should be documented prior to system development and data collection. The importance of SOPs is underscored by the fact that documented work procedures are mandated by

International Standards Organization (ISO) quality system standards. Variation in approaches to documenting procedures are common, but the essential requirement is that each process through which data pass should be documented in such a way that the published data tables and listings can be traced back to the raw data. Differences between the scientific and operational aspects of clinical research projects often necessitate multiple levels of documentation; for example, a standard procedure level that applies across studies, coupled with a project-specific level that pertains to individual studies or groups of similar studies. Further, because organizations, regulations, and practices change, process documentation should be maintained in the context of a regular review and approval cycle.

5. *Personnel management infrastructure; job descriptions, review of and feedback on employee performance, and procedures for managing performance*.
Written job descriptions generally include minimum qualifications and experience, a detailed list of job responsibilities, and reporting structure. These descriptions help the candidate as well as the hiring manager(s) assess a person's suitability for a job. In addition, they help organizations communicate expectations and maintain performance standards for a given position. Appropriate data quality assurance infrastructure also includes regular review of employees' work and a means of providing meaningful and actionable feedback to employees. If management is nonexistent or incapable of reviewing employees' work and providing oversight and technical guidance, a key component of the quality assurance infrastructure is absent. Managers should also identify and define both good and inadequate performance, and there should be organizational procedures for encouraging the former and correcting the latter. While these concerns may sound more appropriate for a business office, personnel management infrastructure is crucial to data quality in clinical research because even with continuing technological development, humans still perform all of the design, and much of the data collection and processing, and human performance directly affects data quality.

6. *Project management in clinical research informatics begins with understanding the basic data-related requirements of a study, i.e., the data deliverables, associated costs, the necessary levels of quality, and the amount of time required or available*.
Project management also includes planning to meet requirements as well as ongoing tracking, assessment, and reporting of status with respect to targets. Project management profoundly affects data quality; for example, good planning and forecasting make the necessary resources and time for a given project transparent. Keeping a project on schedule eliminates (or at least mitigates) pressure to rush or cut corners and often results in employees who feel less harassed or fatigued.

Together, these six structural components form a quality system for the collection and management of data in clinical research (Fig. 10.5).

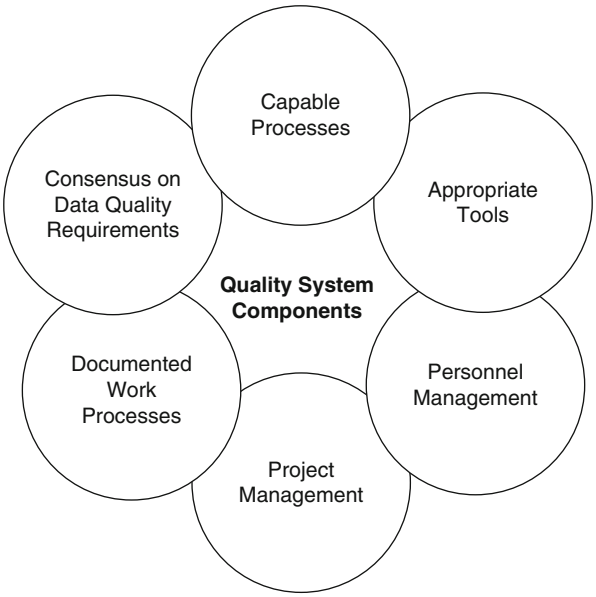

Fig. 10.5 Components of a data quality system. The environment in which data are collected and processed impacts data quality. Thus, achieving and controlling data quality usually requires action from entities in the broader environment

Impact of Data Quality on Research Results

In most clinical research, the goal is to answer a scientific question. This is often done through inferential statistics. Unfortunately, a "one size fits all" data quality acceptance criterion is not possible because statistical tests vary in their robustness to data errors. Further, the impact on the statistical test depends on the variable in which the errors occur and the extent of the errors. Further still, data that are of acceptable quality for one use may not be acceptable for another, i.e., the "fitness for use" aspect addressed earlier. It is for these reasons that regulators cannot set a data quality minimum standard or an "error rate threshold."

What we can say is that data errors, measurement variability, incompleteness, and delays directly impact the statistical tests through adding variability, potentially decreasing power. As shown conceptually in Fig. 10.6, added variability makes it more difficult to tell if two distributions (i.e., a treatment and a control group) are different. Data error rates reported in the literature are well within ranges shown to cause power drops or necessitate increases in sample size in order to preserve statistical power [62, 63]. While it is true that sample size estimates are based on data that also have errors, i.e., the sample size accounts for some base level of variability, data errors have been shown to change p values [26] and attenuate correlation coefficients to the null [64–66] (i.e., for trials that fail to reject the null hypothesis, data errors rather than a true lack of effect could be responsible) [67]. In the context

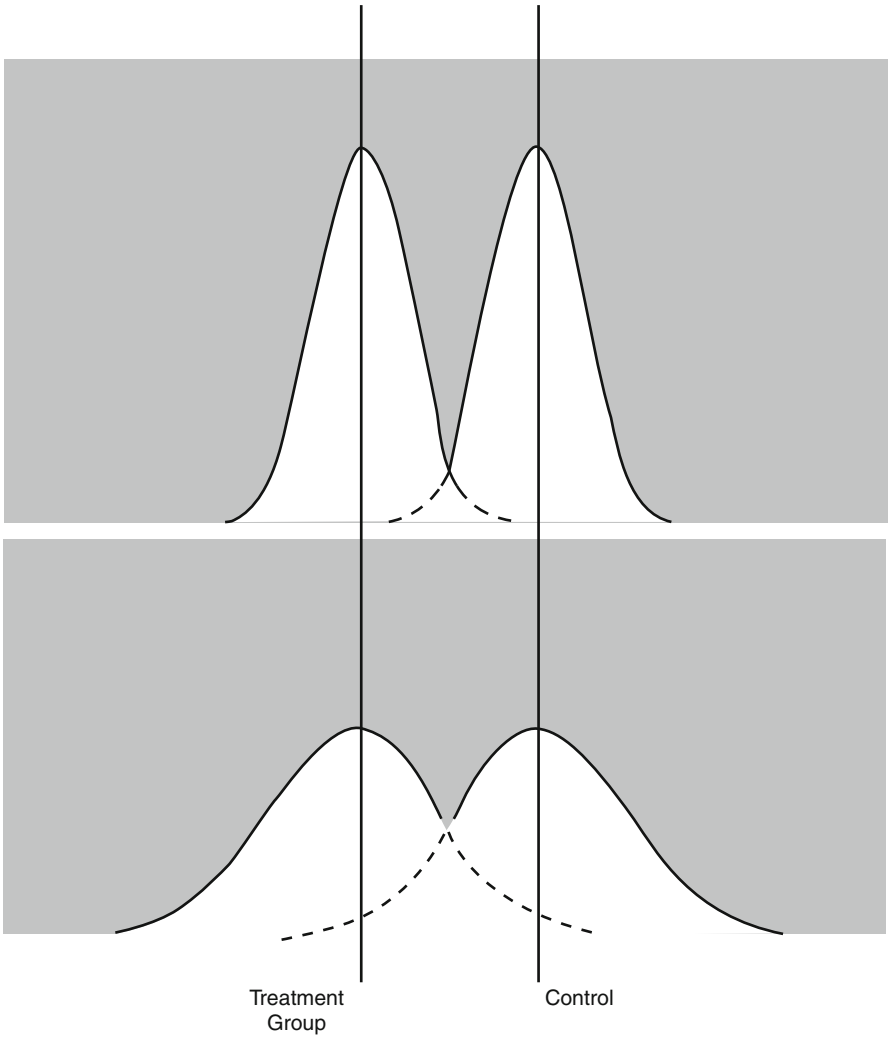

Fig. 10.6 Effect of adding variability. The top two distributions have less variability (are narrower) than the bottom two, making it easier to tell them apart both visually and statistically

of large data error rates, a researcher must choose either to: (1) accept power loss, risking an incorrect indication toward the null hypothesis due to data error, or (2) undertake the expense of measuring the error rate and possibly also the expense of increasing the sample size accordingly to maintain the original desired statistical power [55, 63, 66]. The adverse impact of data errors has also been demonstrated in other secondary data uses such as registries and performance measures [68–74]. Thus, whether or not data are of acceptable quality for a given analysis is a question to be assessed by the study statistician. The assessment should be based on measured error and completness rates.

Summary

The following important points apply to data and information collected and managed in clinical research: (1) errors occur naturally, (2) sources of error are numerous, (3) some errors can be prevented, (4) some errors can be detected, and (5) some errors can be corrected. The sets in 3–5 do not completely overlap. At the same time, there are errors that cannot be prevented, detected, or corrected (e.g., a study subject who deliberately provides an inaccurate answer on a questionnaire). Errors exist in all data sets, and it is foolish to assume that any collection of data is error-free. While higher quality data are often associated with overall savings, preventing, detecting, and correcting errors are associated with additional or redistributed costs.

The skilled practitioner possesses knowledge of error sources and matching methods for prevention, mitigation, detection, and correction where they exist. Further, the skilled practitioner applies this knowledge to design clinical research data collection and management processes that provide the needed quality at an acceptable cost. Achieving and maintaining data quality in clinical research is a complex undertaking. If data quality is to be maintained, it must also be measured and acted upon throughout the course of the research project.

There is widespread agreement that the validity of clinical research rests on a foundation of data. However, there is limited research to guide data collection and processing practice. The many unanswered questions, if thoughtfully addressed, can help investigators and research teams balance costs, time, and quality while assuring scientific validity.

References

1. Davis JR, Nolan VP, Woodcock J, Estabrook EW, editors. Assuring data quality and validity in clinical trials for regulatory decision making. Institute of Medicine Workshop report. Roundtable on research and development of drugs, biologics, and medical devices. Washington, DC: National Academy Press; 1999. http://books.nap.edu/openbook.php?record_id=9623&page=R1. Accessed 6 July 2009.
2. Deming WE, Geoffrey L. On sample inspection in the processing of census returns. J Am Stat Assoc. 1941;36:351–60.
3. Deming WE, Tepping BJ, Geoffrey L. Errors in card punching. J Am Stat Assoc. 1942;37:525–36.
4. Donabedian A. A guide to medical care administration, vol. II: medical care appraisal – quality and utilization. New York: American Public Health Association; 1969. 176.
5. Arndt S, Tyrell G, Woolson RF, Flaum M, Andreasen NC. Effects of errors in a multicenter medical study: preventing misinterpreted data. J Psychiatr Res. 1994;28:447–59.
6. Juran JM, Gryna FM. Juran's quality control handbook. 4th ed. New York: McGraw-Hill; 1988.
7. Guidance for industry E6 good clinical practice: consolidated guidance, ICH E6. April 1996. Available from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073122.pdf. Accessed Aug 2011.
8. Reprinted with permission from Data Gone Awry, DataBasics, vol 13, no 3, Fall. 2007. Society for Clinical Data Management. Available from http://www.scdm.org

9. Nagurney JT, Brown DF, Sane S, Weiner JB, Wang AC, Chang Y. The accuracy and completeness of data collected by prospective and retrospective methods. Acad Emerg Med. 2005;12:884–95.
10. Feinstein AR, Pritchett JA, Schimpff CR. The epidemiology of cancer therapy. 3. The management of imperfect data. Arch Intern Med. 1969;123:448–61.
11. Reason J. Human error. Cambridge: Cambridge University Press; 1990.
12. Nahm M, Dziem G, Fendt K, Freeman L, Masi J, Ponce Z. Data quality survey results. Data Basics. 2004;10:7.
13. Schuyl ML, Engel T. A review of the source document verification process in clinical trials. Drug Info J. 1999;33:789–97.
14. Batini C, Catarci T, Scannapieco M. A survey of data quality issues in cooperative information systems. In: 23rd international conference on conceptual modeling (ER 2004), Shanghai; 2004.
15. Pipino L, Lee Y, Wang R. Data quality assessment. Commun ACM. 2002;45:8.
16. Tayi GK, Ballou DP. Examining data quality. Commun ACM. 1998;41:4.
17. Redman TC. Data quality for the information age. Boston: Artech House; 1996.
18. Wand Y, Wang R. Anchoring data quality dimensions in ontological foundations. Commun ACM. 1996;39:10.
19. Wang R, Strong D. Beyond accuracy: what data quality means to data consumers. J Manage Inform Syst. 1996;12:30.
20. Batini C, Scannapieco M. Data quality concepts, methodologies and techniques. Berlin: Springer; 2006.
21. Wyatt J. Acquisition and use of clinical data for audit and research. J Eval Clin Pract. 1995;1:15–27.
22. U.S. Food and Drug Administration. Guidance for industry. Computerized systems used in clinical trials. In: Services DoHaH, editor. Rockville: U.S. Food and Drug Administration; 2007.
23. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002;9:600–11.
24. (CDISC) CDISC. The Protocol Representation Model version 1.0 draft for public comment: CDISC; 2009. p. 96. Available from http://www.cdisc.org
25. Jacobs M, Studer L. Forms design II: the course for paper and electronic forms. Cleveland: Ameritype & Art Inc.; 1991.
26. Eisenstein EL, Lemons PW, Tardiff BE, Schulman KA, Jolly MK, Califf RM. Reducing the costs of phase III cardiovascular clinical trials. Am Heart J. 2005;9:482–8.
27. Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. Clin Trials. 2008;5:75–84.
28. Galešic M. Effects of questionnaire length on response rates: review of findings and guidelines for future research. 2002. http://mrav.ffzg.hr/mirta/Galesic_handout_GOR2002.pdf. Accessed 29 Dec 2009.
29. Roszkowski MJ, Bean AG. Believe it or not! Longer questionnaires have lower response rates. J Bus Psych. 1990;4:495–509.
30. Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, Kwan I. Increasing response rates to postal questionnaires systematic review. Br Med J. 2002;324:1183.
31. Wickens CD, Hollands JG. Engineering psychology and human performance. 3rd ed. Upper Saddle River: Prentice Hall; 2000.
32. Stevens SS. On the theory of scales of measurement. Science. 1946;103:677–80.
33. Allison JJ, Wall TC, Spettell CM, et al. The art and science of chart review. Jt Comm J Qual Improv. 2000;26:115–36.
34. Banks NJ. Designing medical record abstraction forms. Int J Qual Health Care. 1998;10:163–7.
35. Engel L, Henderson C, Fergenbaum J, Interrater A. Reliability of abstracting medical-related information medical record review conduction model for improving. Eval Health Prof. 2009;32:281.

36. Cunningham R, Sarfati D, Hill S, Kenwright D. An audit of colon cancer data on the New Zealand Cancer Registry. N Z Med J. 2008;121(1279):46–56.
37. Fritz A. The SEER program's commitment to data quality. J Registry Manag. 2001;28(1): 35–40.
38. German RR, Wike JM, Wolf HJ, et al. Quality of cancer registry data: findings from CDC-NPCR's breast, colon, and prostate cancer data quality and patterns of care study. J Registry Manag. 2008;35(2):67–74.
39. Herrmann N, Cayten CG, Senior J, Staroscik R, Walsh S, Woll M. Interobserver and intraobserver reliability in the collection of emergency medical services data. Health Serv Res. 1980;15(2):127–43.
40. Pan L, Fergusson D, Schweitzer I, Hebert PC. Ensuring high accuracy of data abstracted from patient charts: the use of a standardized medical record as a training tool. J Clin Epidemiol. 2005;58(9):918–23.
41. Reeves MJ, Mullard AJ, Wehner S. Inter-rater reliability of data elements from a prototype of the Paul Coverdell National Acute Stroke Registry. BMC Neurol. 2008;8:19.
42. Scherer R, Zhu Q, Langenberg P, Feldon S, Kelman S, Dickersin K. Comparison of information obtained by operative note abstraction with that recorded on a standardized data collection form. Surgery. 2003;133(3):324–30.
43. Stange KC, Zyzanski SJ, Smith TF, et al. How valid are medical records and patient questionnaires for physician profiling and health services research? A comparison with direct observation of patients visits. Med Care. 1998;36(6):851–67.
44. Thoburn KK, German RR, Lewis M, Nichols PJ, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. Cancer. 2007;109(8):1607–16.
45. To T, Estrabillo E, Wang C, Cicutto L. Examining intra-rater and inter-rater response agreement: a medical chart abstraction study of a community-based asthma care program. BMC Med Res Methodol. 2008;8:29.
46. Yawn BP, Wollan P. Interrater reliability: completing the methods description in medical records review studies. Am J Epidemiol. 2005;161(10):974–7.
47. La France BH, Heisel AD, Beatty MJ. A test of the cognitive load hypothesis: investigating the impact of number of nonverbal cues coded and length of coding session on observer accuracy. Communication Reports. 1 Apr 2007.
48. Helms Ron. Redundancy: an important data forms/design data collection principle. In: Proceedings Stat computing section, Alexandria; 1981. p. 233–237.
49. Helms R. Data quality issues in electronic data capture. Drug Inf J. 2001;35:827–37.
50. U.S. Food and Drug Administration regulations Title 21 CFR Part 58. 2011. Available from http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=58. Accessed Aug 2011.
51. Nahm ML, Pieper CF, Cunningham MM. Quantifying data quality for clinical trials using electronic data capture. PLoS One. 2008;3(8):e3049.
52. Winchell T. The mystery of source documentation. SOCRA Source 62. 2009. Available from http://www.socra.org/.
53. Nahm, M. Data Accuracy in Medical Record Abstraction. Doctoral Dissertation, University of Texas at Houston, School of Biomedical Informatics, Houston Texas, May 6, 2010.
54. SCDM. Good clinical data management practices. www.scdm.org. Society for Clinical Data Management; 2010. Available from http://www.scdm.org
55. Rostami R, Nahm M, Pieper CF. What can we learn from a decade of database audits? The Duke Clinical Research Institute experience, 1997–2006. Clin Trials. 2009;6(2):141–50.
56. Svolba G, Bauer P. Statistical quality control in clinical trials. Control Clin Trials. 1999;20(6): 519–30.
57. Chilappagari S, Kulkarni A, Bolick-Aldrich S, Huang Y, Aldrich TE. A statistical process control method to monitor completeness of central cancer registry reporting data. J Registry Manag. 2002;29(4):121–7.

58. Chiu D, Guillaud M, Cox D, Follen M, MacAulay C. Quality assurance system using statistical process control: an implementation for image cytometry. Cell Oncol. 2004;26(3):101–17.
59. McNees P, Dow KH, Loerzel VW. Application of the CuSum technique to evaluate changes in recruitment strategies. Nurs Res. 2005;54(6):399–405.
60. Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials. 2008;5(1):49–55.
61. Matheny ME, Morrow DA, Ohno-Machado L, Cannon CP, Sabatine MS, Resnic FS. Validation of an automated safety surveillance system with prospective, randomized trial data. Med Decis Making. 2009;29(2):247–56.
62. Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement error on planning sample size required in a cohort study. Am J Epidemiol. 1990;132:1185–95.
63. Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. Biol Psychiatry. 2007;47:762–6.
64. Mullooly JP. The effects of data entry error: an analysis of partial verification. Comput Biomed Res. 1990;23:259–67.
65. Liu K. Measurement error and its impact on partial correlation and multiple linear regression analyses. Am J Epidemiol. 1988;127:864–74.
66. Stepnowsky Jr CJ, Berry C, Dimsdale JE. The effect of measurement unreliability on sleep and respiratory variables. Sleep. 2004;27:990–5.
67. Myer L, Morroni C, Link BG. Impact of measurement error in the study of sexually transmitted infections. Sex Transm Infect. 2004;80(318–323):328.
68. Williams SC, Watt A, Schmaltz SP, Koss RG, Loeb JM. Assessing the reliability of standardized performance indicators. Int J Qual Health Care. 2006;18:246–55.
69. Watt A, Williams S, Lee K, Robertson J, Koss RG, Loeb JM. Keen eye on core measures. Joint Commission data quality study offers insights into data collection, abstracting processes. J AHIMA. 2003;74:20–5; quiz 27–8.
70. US Government Accountability Office. Hospital quality data: CMS needs more rigorous methods to ensure reliability of publicly released data. In: Office UGA, editor. Washington, DC; 2006. www.gao.gov/new.items/d0654.pdf
71. Braun BI, Kritchevsky SB, Kusek L, et al. Comparing bloodstream infection rates: the effect of indicator specifications in the evaluation of processes and indicators in infection control (EPIC) study. Infect Control Hosp Epidemiol. 2006;27:14–22.
72. Jacobs R, Goddard M, Smith PC. How robust are hospital ranks based on composite performance measures? Med Care. 2005;43:1177–84.
73. Pagel C, Gallivan S. Exploring consequences on mortality estimates of errors in clinical databases. IMA J Manage Math. 2008;20(4):385–93. http://imaman.oxfordjournals.org/content/20/4/385.abstract
74. Goldhill DR, Sumner A. APACHE II, data accuracy and outcome prediction. Anaesthesia. 1998;53:937–43.