Risk-Based Monitoring and Fraud Detection in Clinical Trials Using JMP® and SAS®

Richard C. Zink

Contents

1

Introduction

1.1 Overview

Pharmaceutical development is an extremely complex affair. Once a promising compound is identified, steps are taken to optimize the chemical properties and formulation, understand the pharmacokinetics, pharmacodynamics, and safety through animal testing, then introduce the drug into humans to identify an efficacious and safe dose that can address some unmet medical need. The process involves countless tests and experiments, identifying clinicians to recruit patients into clinical trials, communicating with vendors for supplies or analysis, and routinely contacting regulatory agencies to ensure standards are met or to disclose any safety signals. There are study data to retrieve, monitor, and prepare for analysis and submission; frequent reviews to identify safety and quality concerns; scores of statistical analyses to perform; and study reports to author. This effort is made even more difficult for multinational trials: Documents require translation; differing time zones and holidays affect schedules; supplies, samples, and personnel travel long distances; and new rules and requirements are applied based on local regulatory

bodies. And all of this has to happen in a timely fashion; patents are of limited duration and allow a brief opportunity for the sponsor to recoup its investment.

Given the multitude of tasks and issues just described, the numerous parties involved, and the many pitfalls that can sideline a drug (or vaccine or medical device), it is a bit surprising that the whole affair doesn't collapse on a routine basis. The success of the pharmaceutical industry is due in large part to its adherence to the processes and procedures needed for achieving various goals, as well as its commitment to detailing the resolution of problems that may occur along the way. Standard Operating Procedures (SOPs) are documents that outline the steps necessary to ensure that everything is carried out completely and accordingly, including communicating with the relevant parties both internal and external to the company. SOPs are a regular part of training for all departments and, depending on the topic, may require annual review. There are often SOPs on how to prepare and validate analyses for the final study report, how to address safety concerns that may occur at clinical sites, how to effectively incorporate and manage data monitoring committees for interim study review, and what to do should a regulatory agency perform an audit or inspection. Once in a while, you may find an SOP on SOPs!

So why include a discussion on process and procedure in a book about clinical trial software? For one, the pharmaceutical industry is at a crossroads. Productivity has dropped, clinical trials have become increasingly complex, and the costs of conducting them have skyrocketed [1–4]. Controlling these costs is essential, especially when faced with a low probability of success along the development pathway toward the marketplace. Said costs are ultimately passed on to the consumer; reining in these costs is one way to keep drugs affordable for the people who need them. According to Venet and coauthors, if costs continue to rise at the current pace, clinical trials to establish efficacy and tolerability will become impossible to conduct. This will make drugs unavailable for areas of unmet need, stifling innovation in established treatment areas, or placing an extreme price burden on consumers and health care systems [4].

Many innovations have been suggested and developed to streamline the pharmaceutical development process and improve the likelihood of clinical and regulatory success. For example, adaptive design methodologies allow for early stopping of a clinical trial in the presence of overwhelming efficacy or excess toxicity, or when the novel compound has little chance to distinguish itself from control. Extensive modeling and simulation exercises are used to suggest the most successful path forward in a clinical program based on the available data and reasonable assumptions based on past development. Patient enrichment based on genomic markers is used to select a study population more likely to receive benefit from the drug, resulting in smaller clinical trials. Some innovations have more to do with the operational aspects of clinical trials. These include electronic case report forms (eCRFs), new technologies for collecting diary data or obtaining laboratory samples, or new software that enables the efficient review of data for quality and safety purposes. And still other innovations involve the regulatory submission and review process through electronic submissions and data standards.

While pharmaceutical development is driven by process and performance, it can be slow to implement new ideas, even if they are shown to have substantial benefit. First and foremost, people are naturally resistant to change, particularly if they are comfortable in how they perform a given task. Second, the pace at a pharmaceutical company rarely slows. Individuals involved with evaluating new software, products, or techniques still have to keep the trials for which they are responsible operating smoothly while meeting or exceeding timelines. And clinical trials can last many years; finding an opportunity to implement changes may be difficult. Third, changes brought

about through innovation are rarely innocuous; they can affect the processes and performance for multiple groups of individuals. This last point is particularly important to consider, since in order to implement a new method successfully, one must anticipate the growing pains and hiccups that may occur along the way.

This book is concerned with innovating the data review process for clinical trials by introducing software and techniques to steer clear of the manual "examine every data point" methods of days past. The new approach moves the reviewer from the static paper environment into an interactive and visual one. First, I promote the centralized and programmatic review of clinical trial data for signals that would indicate safety or quality problems at the clinical sites. Second, I describe methods to help uncover patient and investigator misconduct within the clinical trial. Finally, I discuss some ways to accelerate clinical trial reviews so that reviewers do not spend precious time on previously examined data. Throughout this book, I illustrate the various concepts and techniques using JMP Clinical, which I describe in "1.4 JMP Clinical" on page 9.

Any clinical trial is the result of the efforts of a diverse team. This team includes clinicians, statisticians, data managers, programmers, regulatory associates, and monitors, to name a few. Every team member has a role to play in the review of trial data. Each individual brings a unique skill set important for understanding patient safety, protocol adherence, or data insufficiencies that can affect the final analysis, clinical study report, and subsequent regulatory review and approval. For all of the aforementioned roles, aspects of JMP Clinical can streamline day-to-day work and provide new insights. It is my sincere hope that any individual from the clinical trial team can make use of this book and the examples contained within to streamline, accelerate, and enrich their reviews of clinical trial data. The few places where I get a bit technical or present SAS code or JMP scripts can be skipped by the average reader with no loss to their understanding. Major topics of this book are described in the next section. Each chapter is relatively self-contained so that the reviewer can read sections important to the task at hand.

1.2 Topics Addressed in This Book

1.2.1 Risk-Based Monitoring

Since 1990, the International Conference on Harmonisation (ICH) has brought together the regulatory bodies of the European Union, Japan, and the United States. The mission of the ICH is to define a set of technical and reporting guidelines for clinical trials to minimize the testing required in humans and animals to what is absolutely necessary to establish efficacy and safety, reduce development times, and streamline the regulatory review process. In particular, ICH Guideline E6 outlines standards for Good Clinical Practice (GCP) in the design, conduct, and reporting of clinical trials involving human participants [5]. GCP has two primary goals: to protect the well-being of subjects involved in a clinical trial and to maintain a high level of data quality to ensure the validity and integrity of the final analysis results.

Guideline E6 suggests that clinical trial data should be actively monitored to ensure data quality. Despite passages that state "the sponsor should determine the appropriate extent and nature of monitoring" and "statistically controlled sampling may be an acceptable method for selecting the

data to be verified," recent practice for pharmaceutical trials has often shown a brute-force approach to source data verification (SDV) of respective CRFs through on-site monitoring [5–7]. The recent Food and Drug Administration (FDA) guidance document on risk-based monitoring (defined later) suggests a few reasons as to why this may have occurred [3]. First, the on-site monitoring model may have been (incorrectly) perceived as the preferred approach of the FDA. Second, the FDA document suggests that the agency places more emphasis on centralized monitoring than what may have been feasible at the time ICH E6 was finalized (there have been considerable technical and analytical advances in the 17 years since ICH E6 was written). While language in E6 refers to central monitoring, it does state a need for on-site monitoring "before, during, and after the trial."

However the pharmaceutical industry arrived at the current practice for clinical trial monitoring, it is now generally accepted by industry and multiple regulatory agencies that the process needs to change [1,3,8–10]. Such extensive on-site review is time consuming, expensive (up to a third of the cost of a clinical trial), and—as is true for any manual effort—limited in scope and prone to error [1,4,11–15]. In contrast to on-site monitoring, risk-based monitoring (RBM) makes use of central computerized review of clinical trial data and site metrics to determine whether clinical sites should receive more extensive quality review through on-site monitoring visits. There are many benefits to centralized review beyond cost: Statistical and graphical checks can determine the presence of outliers or unusual patterns in the data, comparisons can be made between sites to assess performance and identify potentially fraudulent data or miscalibrated or faulty equipment, and issues can be identified and resolved while the trial is ongoing.

Changing current monitoring practices to a risk-based approach will likely take time; the industry must become comfortable with a reduced presence at clinical sites and implement procedures for the remote review of clinical data, statistical sampling of data for SDV, and targeted monitoring practices. However, it is clear that the reliance on SDV, a major focus of current on-site monitoring practice, is increasingly viewed to have little to no positive impact on study conclusions [4,12]. In its position paper on RBM, TransCelerate BioPharma Inc. notes that only 7.8% of the queries generated from nine sample studies were the result of SDV, a huge investment for minimal return on data quality [1]. An example from a large international, multicenter trial found that of the issues identified, 28.4% could have been identified from the study database, and a further 66.8% could have been identified with some additional centralized edit-checks in place [16]. Further, Nielsen and coauthors illustrate that a reduced SDV monitoring approach could locate all critical queries from a pool of 30 completed clinical trials [17].

However, centralized review can only identify issues contained within the study database or other routinely collected information [4]. On-site monitoring may still be required to assess the quality of overall trial conduct, including whether appropriate regulatory documentation is available, the staff is familiar with and committed to the protocol, the staff is appropriately trained, and trial resources are adequate, well-maintained, and functional [3,5,6]. Recent literature has suggested that a diversified approach to monitoring, including centralized statistical and programmatic checks, can identify deficiencies that would otherwise go unnoticed with on-site review alone [4,6,18]. When issues are identified to a degree that may suggest a more systemic problem at the site, targeted on-site monitoring activities can be applied according to the extent of the problem and the importance of the data to the conclusions of the study [1]. The literature also stresses this important point: Data does not need to be error-free to provide reliable results from a clinical trial [1,3,6]. Finally, in addition to the risk-based methods described previously, the literature suggests

a proactive approach to quality and safety through appropriate trial and CRF design, well-defined study procedures, and sufficient training of site personnel.

Chapters 2 and 3 of this book discuss an implementation of RBM within JMP Clinical that keeps to the recommendations of TransCelerate BioPharma [1]. The current application makes use of the clinical trial database and allows the team to supplement this information with any other data captured at the site level in order to assess the performance of the sites. Making use of the study data for RBM eliminates any unnecessary redundancies for similar data tracked external to the database, as well as the need for any potential reconciliation, should discrepancies arise. Chapter 2 introduces basic concepts of RBM and how to conduct reviews within JMP Clinical. Chapter 3 describes how users can customize their analysis and review experience.

1.2.2 Fraud Detection

Fraud is an important subset of topics involving data quality, one that perhaps conjures images of Sherlock Holmes (or Scooby-Doo and the Gang) on the hunt for clues to apprehend the bad guy. Quality issues in clinical trials can be due to a number of factors, among them carelessness (such as transcription errors), contamination of samples, mechanical failures, or miscalibrated equipment, poor planning (e.g., lack of appropriate backups or contingency planning should problems occur), poor training in trial procedures, and fraud. Fraud stands out among other quality issues in that there is a “deliberate attempt to deceive” or the “intention to cheat” [18]. In this book, we consider both patient- and investigator-perpetrated fraud in clinical trials. For investigators, fraud is often viewed as fabricating, manipulating, or deleting data. Examples of this behavior include deleting data highlighting a safety concern, propagating (carrying forward) data to avoid performing additional testing, or the wholesale manufacture of one or more patients at the site. For patients, enrolling at two or more clinical trials sites (usually for additional financial compensation, or access to additional drugs or medical services) is particularly problematic. This, of course, violates assumptions of statistical independence necessary for many statistical tests. In practice, subjects with multiple enrollments become an accounting and reporting nightmare for the trial team.

Despite a bevy of statistical and graphical tools available to identify unusual data, fraud is extremely difficult to diagnose. For one, many of the methods used to identify misconduct at a center involve comparisons against other clinical trial sites. Such analyses could identify natural differences in patient population or variations in technique between the sites that wouldn’t constitute fraudulent behavior. Further, as we’ll see later on, analyses motivated by a need to identify a particular type of malfeasance can detect data anomalies that actually have perfectly reasonable explanations. In general, stating that any unusual findings are explicitly due to fraud may require evidence beyond what’s available in the clinical trial database [19].

It is believed that fraud in clinical trials is rare. Buyse and co-authors estimate the proportion of investigators engaging in misconduct below 1%, though they suggest cases may go either undiagnosed or unreported; additional cited works therein show few to no instances of fraud [18]. Two recent publications describe higher than expected rates of scholarly retractions in the life science and biomedical literature, often due to fraud or suspected fraud [20,21]. Further, in a survey of statisticians, half of 80 respondents reported awareness of fraud or deliberate deception in at least one project in the preceding 10 years, though there were some concerns about the survey’s response rate and the lack of a clear definition of fraud [22,23]. Although far from a

scientifically conducted poll, I obtained a similar response from an audience of approximately 35 statisticians during a section of talks on clinical trial fraud at the 2013 Statisticians in the Pharmaceutical Industry (PSI) Annual Meeting. In other surveys, misconduct was or was considered unlikely to occur [24,25].

Though the preceding paragraph paints an inconsistent picture, clinical trial fraud is likely underestimated for several aforementioned reasons. First, instances have conceivably gone undiagnosed due to a lack of available tools and training for uncovering fraud. Part of this may be due to the past overreliance on manual on-site monitoring techniques, which made comparisons across CRF pages, subjects, time, and clinical sites difficult. Further, even if unusual data are identified, going that additional step to confirm any misdoing may prove unsuccessful. Second, even if suspected fraud is detected, it may go unreported over fears that the negative publicity can severely damage the perception of an organization among regulatory agencies, patients, and the general public. Even if the sponsor has behaved entirely appropriately, such attention can bring increased scrutiny and pressure to the clinical trial and larger development program [26].

Recommendations to prevent clinical trial misconduct include simplifying study entry criteria, minimizing the amount of data collected, and ensuring sufficient and varied trial monitoring [4,6,18,26]. Even in the presence of incorrect data due to manipulation or other quality issues, trial integrity will be preserved in most cases, most often due to randomization and blinding of study medication, or because the anomalies are limited to few sites [1,3,6,18,26]. In general, clinical trial data are highly structured, and human beings are bad at fabricating realistic data, particularly in the many dimensions that would be required for it to appear plausible [4,18,19,26–30]. So this begs the question: If clinical trial fraud is so uncommon, with seemingly limited potential to seriously compromise the results of the trial, then why bother looking for it at all?

The simplest answer is that an ounce of prevention is worth a pound of cure [18]. It is much easier to identify problems as they occur while the trial is ongoing, with the opportunity to resolve the issue or modify the trial as needed. Compare this to the scenario of finding a systemic problem once the trial has been unblinded and the final study results have been prepared. At this point, there are fewer options available to the study team to find an appropriate solution (particularly when their every action will be scrutinized due to the availability of randomization codes). Defining a series of statistical and graphical checks to be implemented on a regular basis is a minimal investment for the team to make to prevent potential catastrophe.

Most important, however, we look for quality issues and misconduct because of GCP—to protect the rights and well-being of the patients enrolled in our clinical trials. Monitoring ensures that trial participants receive the best possible care and are protected from any potential wrongdoing. This is equally true for future patients who hope to use the new treatment to improve their quality of life. The rights and safety of the patient are exactly why methodologies in data quality and fraud detection should be a regular part of our statistical training. Evans states that “a perfect analysis on the wrong data can be much more dangerous than an imperfect analysis of correct data”; he suggests this is reason enough for discussing these methodologies even when others can potentially use this knowledge to avoid detection [19]. The safety and well-being of trial participants obligates us to share, collaborate, and improve methods for detecting misconduct in clinical trials. With a greater emphasis on centralized monitoring and less visibility at the clinical trial site, it is important to have a set of robust methods in place to identify fraud. Chapter 4 describes various analyses available in JMP Clinical to identify potentially fraudulent data at clinical sites. Chapter 5 presents methods to detect fraud committed by study participants.

1.2.3 Snapshot Comparisons

While Chapters 2–5 focus on assessing data quality in an efficient manner, Chapter 6 focuses on the practical considerations of review that are brought about as a result of how clinical trial data are collected and reviewed. To perform the final analysis as early as possible after the trial ends, data are collected and cleaned as they become available. Depending on the size of the trial, the number of centers, and whether enrollment is currently ongoing, new data may become available on a daily basis. This creates a constantly updating and changing database. In general, it isn't practical to update needed data sets and regenerate review reports daily—it would be difficult for reviewers to cope with this constantly moving target! Instead, an intermittent "snapshot" of the study database is taken that reflects the currently collected data and any changes since the previous snapshot. The snapshot is reviewed and necessary queries are generated to address any inconsistencies in the data or gaps in the information provided.

After a sufficient period of time, a new snapshot is generated that incorporates new data collected since the previous snapshot and any changes that were made to previously available data due to sponsor query and/or correction at the trial site. The frequency of study snapshots may depend on how much new data becomes available as well as the current lifespan of the trial. Snapshots may initially be infrequent until all study centers are operating and enrolling subjects, though snapshots occur with regularity once a sufficient number of patients are participating. Once the trial approaches last-patient-last-visit (LPLV) status, snapshots may occur very frequently, perhaps even daily, to review the final subjects' data and ensure that all needed corrections have been addressed before "locking" the trial database. Once a trial is locked, treatment codes become unblinded, and the final analysis is performed. At this point, it is expected that no further changes to the database will occur; doing so would raise suspicions that changes were made based on available treatment assignments.

As part of the review process, any number of analyses or listings may be generated to assess the quality of the data and issues pertaining to patient safety. These may include analyses to identify whether any subjects fail study eligibility criteria based on the available data, or listings to review serious adverse events or clinically significant laboratory abnormalities, dosing compliance, inconsistencies between CRF pages pertaining to visit dates or study phases, particularly noteworthy concomitant medications, and so on. During my days in the pharmaceutical industry, the biostatistics and programming group would regularly supply these listings to our clinical and regulatory team members. Inevitably, as these reviews continued, the question from our colleagues became, "Is it possible to just provide the new data so I don't have to review what I've already seen?" This seems like a perfectly reasonable request. However, the review reports were often slight modifications of the analyses that would be performed for the final clinical study report at the end of the trial. The programs rarely were written to highlight changes from one snapshot to the next, and there rarely were sufficient resources to write additional programs.

Creating informative reviews from one snapshot to the next is always more complex than just identifying and subsetting to the new data. For example, some data change during the course of the trial, so naturally there is an interest in reviewing the previous values. And to further complicate things, I found that people often wanted the new data as well as the previous data so that they could remind themselves of what led to a patient's current state of affairs, or to see if previous trends were continuing with the additional records. In other words, reviewers wanted the

ability to subset or filter to new data at will. In a paper or static environment (e.g., PDF tables or listings), this request had the potential to generate twice as many analyses. Database software can track changes over time, but these tools are limited to a few individuals and are rarely available for review of the data sets that ultimately will be used for analysis and submission to regulatory agencies.

Chapter 6 of this book describes how JMP Clinical performs comparisons between current and previous data snapshots to identify new or modified values so that reviewers do not waste time on previously examined data. Further, I'll illustrate how the JMP data filter and generated review flags allow the analyst to easily switch between review summaries, including all or only newly available records. Finally, I'll examine how the extensive notes facility allows the user to create and save notes at the analysis, patient, or record level. These features provide more efficient and accurate reviews to identify and manage potential safety concerns, and to meet or exceed demanding timelines.

1.3 The Importance of Data Standards

Since its inception in 1997, the Clinical Data Interchange Standards Consortium (CDISC) has developed standards for data models, study designs, and supporting clinical trial documents. Standards for data models include the Study Data Tabulation Model (SDTM) for clinical trial data obtained from CRFs, questionnaires, and/or diaries; and the Analysis Data Model (ADaM) for derived data that support the analysis tables, listings, and figures common in drug applications [31,32]. At their core, CDISC standards are a means to streamline communications across the various parties involved in drug development, allowing for quicker review of drug submissions [33].

The SDTM model divides clinical trial data domains, a collection of logically related observations with a common topic, into a number of classes: events, findings, interventions, special purpose, and trial design. Events domains include adverse events, medical history, disposition, and protocol deviations. Examples of the findings domains are physical examinations, vital signs, ECG test results, laboratory tests, and questionnaires. Interventions domains describe concomitant medications, exposures to study treatment, and the use of substances such as alcohol, tobacco, and caffeine. Special purpose domains include demographic characteristics, general comments, and study visit attendance. Trial design domains include data describing the treatment arms, visits, and inclusion/exclusion criteria. Within each domain, variable names, labels, formats, and terms are provided, and the SDTM model states whether the inclusion of each variable in a domain is required, expected, or permissible.

The goal of ADaM is to clearly and unambiguously communicate the data behind the statistical analyses so that minimal additional manipulation is needed to generate study results (i.e., "one PROC away"). Furthermore, while SDTM typically comprises observed data separated into specific domains, ADaM can include complex derivations, such as imputation and windowing of measurements, on data spanning several domains. An additional goal of the ADaM model is traceability: providing sufficient details to allow the user to go from a tabulation to the ADaM data set to the corresponding SDTM data sets. The most important ADaM data set is ADSL, the

subject-level data set. This data set provides definitions for study population flags and includes subgrouping and stratification variables, important dates such as first and last dose, and the planned and actual treatments used for each period of the study.

For analyses of clinical trial data, JMP Clinical takes advantage of CDISC standards. Perhaps more correctly, CDISC standards are a requirement! This may initially appear limiting, particularly for companies that have yet to incorporate these standards into their daily operations. However, the FDA not only provides its preferred interpretations of the standards, but it also has announced its intention to make these standards a requirement for regulatory submissions [34–36]. What are the practical benefits of requiring CDISC standards for our software? Most importantly, JMP Clinical ships with out-of-the-box functionality. JMP Clinical isn't a set of tools that requires mapping to your particular data standards, or that necessitates a team of individuals to develop and support a set of reports. Once you register your CDISC-formatted study within JMP Clinical, you're able to generate patient profiles, automate adverse event narratives, and create a host of other reports and analyses. An additional benefit of working directly from CDISC data sets is that you spend your time reviewing and analyzing the very data sets that will be submitted to the FDA or other regulatory agencies. Throughout the book, CDISC variables will be written as domain.domain-variable to be explicit. For example, USUBJID from the demography (DM) domain will be written as DM.USUBJID. When a term can apply to multiple domains, "xx" will be used to imply multiple two-letter domain codes. Interested readers can review what else SAS has available to support CDISC data standards [37].

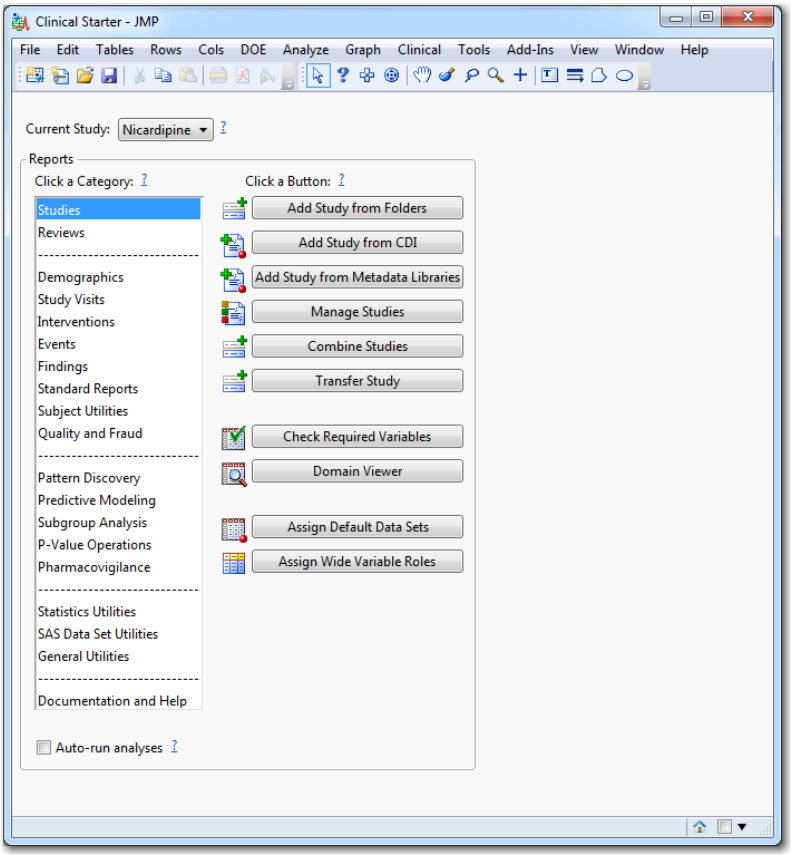
1.4 JMP Clinical

Though I have alluded to the software, I have yet to answer the question: What exactly is JMP Clinical? Like its companion product JMP Genomics, JMP Clinical is a package that combines the analytical power of SAS with the elegant user interface and dynamic graphics of JMP. The primary purpose of this software is to simplify data discovery, analysis, and reporting for clinical trials. JMP Clinical will transform how your team currently conducts its clinical trial reviews. With its straightforward user interface and primary reliance on graphical summaries of results, everyone from the clinical trial team—including clinicians, statisticians, data managers, programmers, regulatory associates, and monitors—can explore data to identify outliers or safety or quality concerns. In this way, everyone on the team can speak the same language by using a single review tool without the hassles of importing data or results into other programs, or maintaining multiple overlapping systems.

JMP Clinical 5.0, the software version that is the focus of this book, includes JMP 11.1 and SAS 9.4 M1 for BASE SAS, SAS/GRAPH, and SAS/ACCESS to PC files, as well as SAS analytics releases 12.3 for SAS/IML, SAS/STAT, and SAS/Genetics. Also included are macros written in SAS and the JMP Scripting Language (JSL) that generate the various analyses and reports that are specific to JMP Clinical. Users familiar with JMP can expect a similar experience; analyses and reports are available from a menu-driven, point-and-click interface (Figure 1.1 on page 10). Generating reports often involves compiling a SAS program under the hood that uses the options selected from JMP dialogs to perform the various data manipulations and analyses. The results of these analyses are surfaced to the user using JSL to make them interactive and dynamic for

further exploration of interesting signals. Though this product makes use of SAS, it is neither expected nor required that the user have any SAS programming experience.

Figure 1.1 *JMP Clinical Starter Menu*



While JMP Clinical is a desktop product, it has the capability to access data and run analyses using a SAS metadata server, or access data and run analyses for studies defined using the SAS Clinical Data Integration (CDI) product. Though the review experience is generally the same, there are some important differences when operating in server mode that I will make note of throughout this book. In general, however, server profiles will be created by an experienced IT professional within your organization (see **Add SAS Server Profile** in the documentation). In addition, this individual may have the responsibility to add and manage studies as new studies or snapshots become available. When connecting to a server (**File > SAS Server Profiles > Select Profile**), studies for which you have been granted access will appear in your **Current Study** drop-down menu. All metadata (including notes and RBM files) will be shared among those individuals with access to the study. When operating locally, each user must register studies available on the local network to his or her particular instance of JMP Clinical.

Because the audience for this book is particularly diverse, many readers may be more familiar with SAS than with JMP. When I was in the pharmaceutical industry, I would have counted myself

among this group of individuals; I wrote everything in SAS and had no experience whatsoever with JMP. In retrospect, this was unfortunate. The power of JMP lies in its ability to quickly and easily explore data in a graphical and interactive environment. Want a regression plot? Drag and drop two variables into the **Graph Builder** platform. See an outlier? Point and click to highlight the offending observation in the data table. It's much easier to identify and further explore anomalies from a picture than it is from a listing or table full of carefully constructed summary statistics (though JMP provides statistics as well). By the end of this book, I hope to convince other SAS users of the benefits of JMP.

Some SAS-only users may have concerns about reproducing a particular visual display when working in an interactive environment. However, every analysis or graph in JMP produces the underlying script that generates the result, and this script can be used to regenerate results at any time. For example, go to **Help > Sample Data** and open the **Big Class** data table from the alphabetical list of data files. Now go to **Analyze > Distribution** and select age, sex, height, and weight as **Y, Columns** and click **OK** (Figure 1.2 on page 11). Feel free to modify the output by closing panels of summary statistics or adding additional plots as I have done (Figure 1.3 on page 12). From the output, click the red triangle by the text Distributions, then go to **Script > Save Script to Script Window**. The new window displays the underlying script for the summary results (Figure 1.4 on page 13). If I click the **Run Script** button in the toolbar (the red running man), the results in Figure 1.3 on page 12 will be regenerated.

Figure 1.2 *JMP Distribution Dialog*

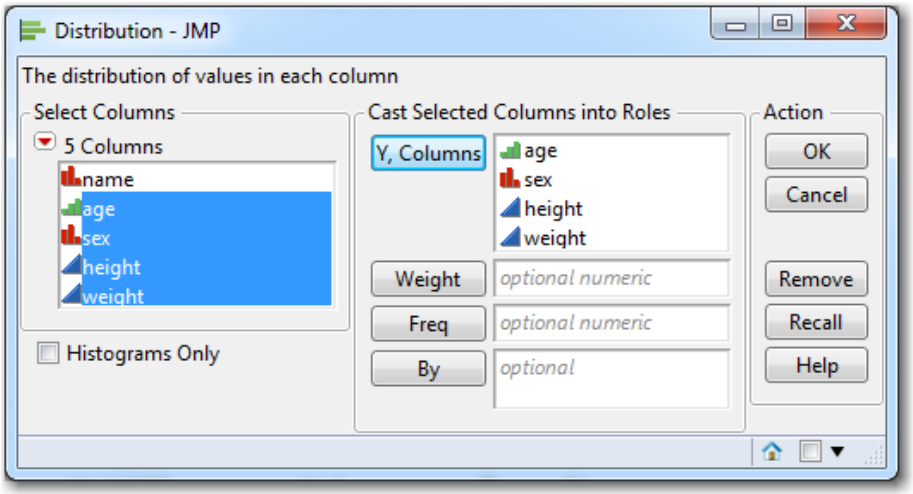
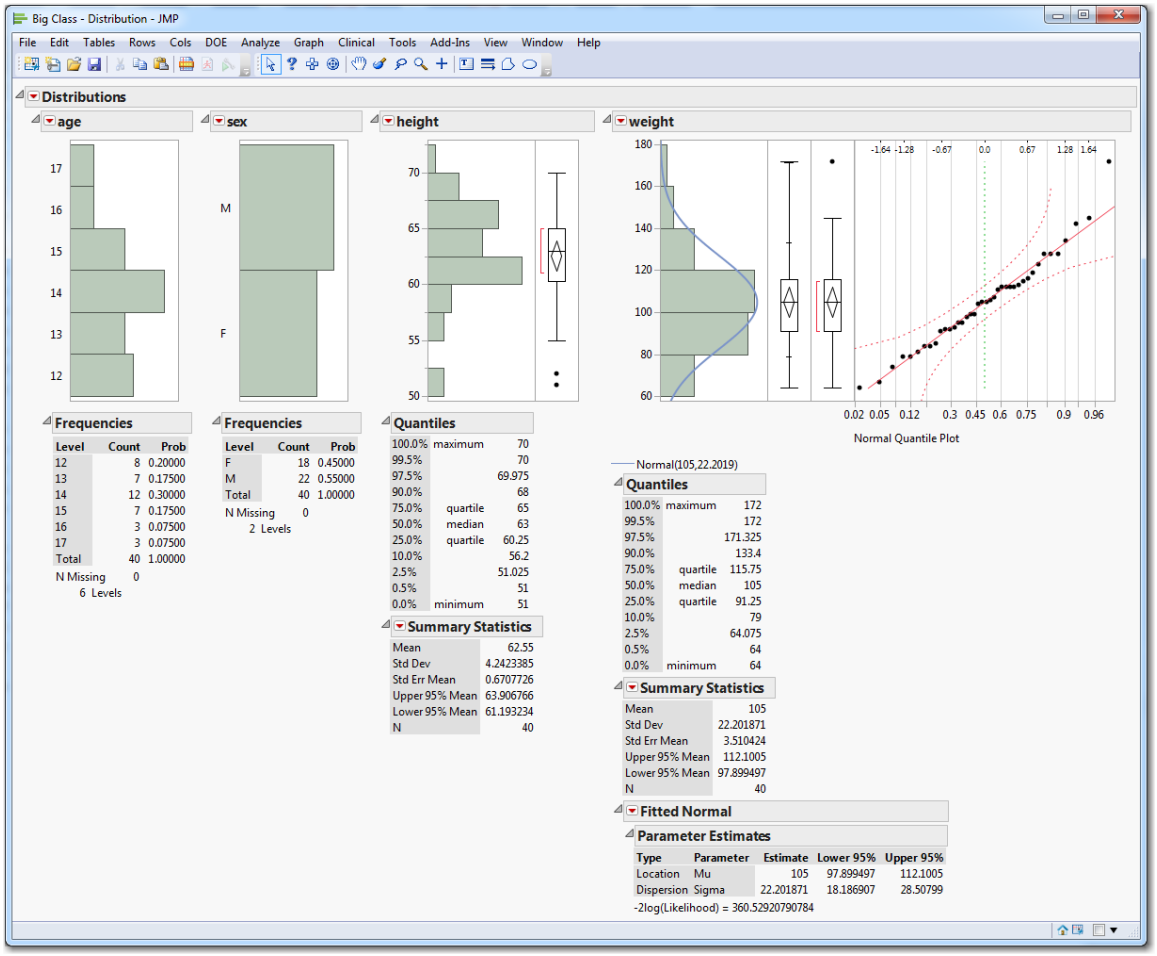


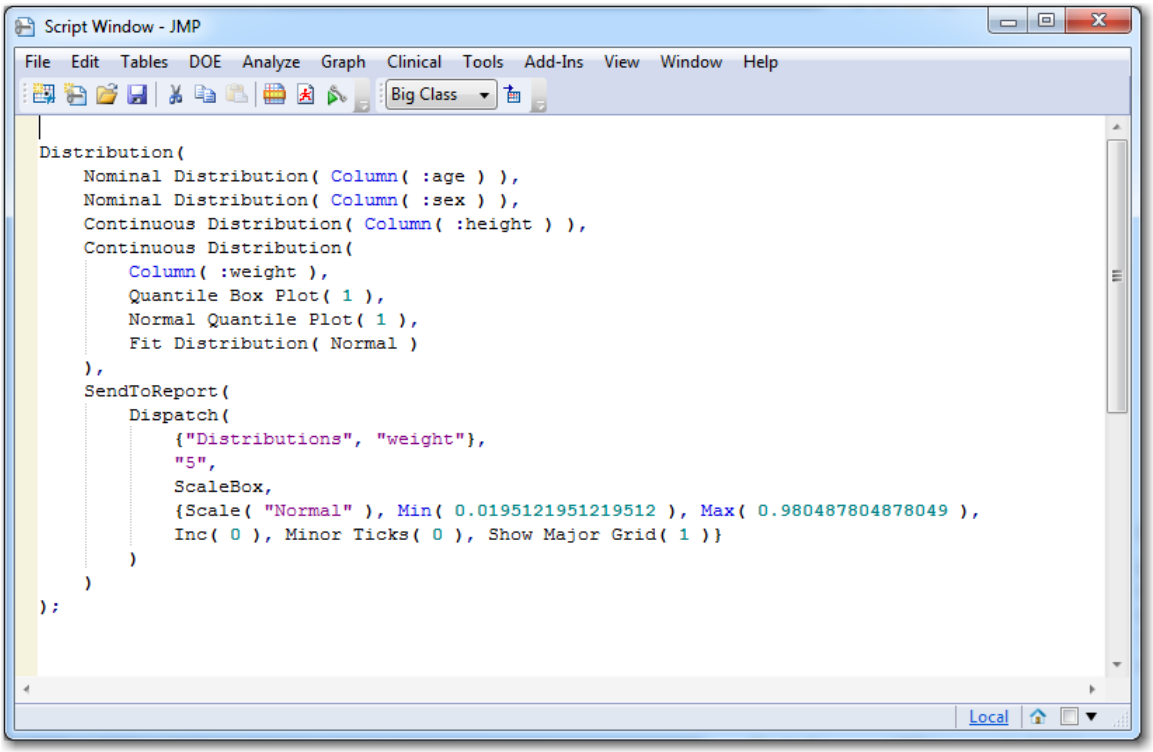
Figure 1.3 *Big Class Distribution Output*



As mentioned previously, JMP Clinical makes use of CDISC variables that are required within each domain. However, in order for some reports within JMP Clinical to function appropriately or provide the greatest detail, there are often additional variable requirements. To assess what important CDISC variables may be missing from your study, run **Check Required Variables** under the **Studies** menu from the Clinical Starter. This report will summarize any analyses that may not run as a result of certain missing variables.

Users new to JMP can review *JMP Essentials, Second Edition* for help navigating the software and performing basic analyses and graphing [38]. Individuals wanting a better understanding of JMP for statistical analysis can read *JMP Start Statistics: A Guide to Statistics and Data Analysis Using JMP, Fifth Edition* [39]. *Jump into JMP Scripting* provides numerous examples and details for those users looking to master JSL [40]. If these books aren't readily available, the **Help** menu provides access to additional books, tutorials, and sample data. Assistance for JMP Clinical is available from **Help > Books > JMP Life Sciences User Guide**.

Figure 1.4 *Big Class Distribution Script*

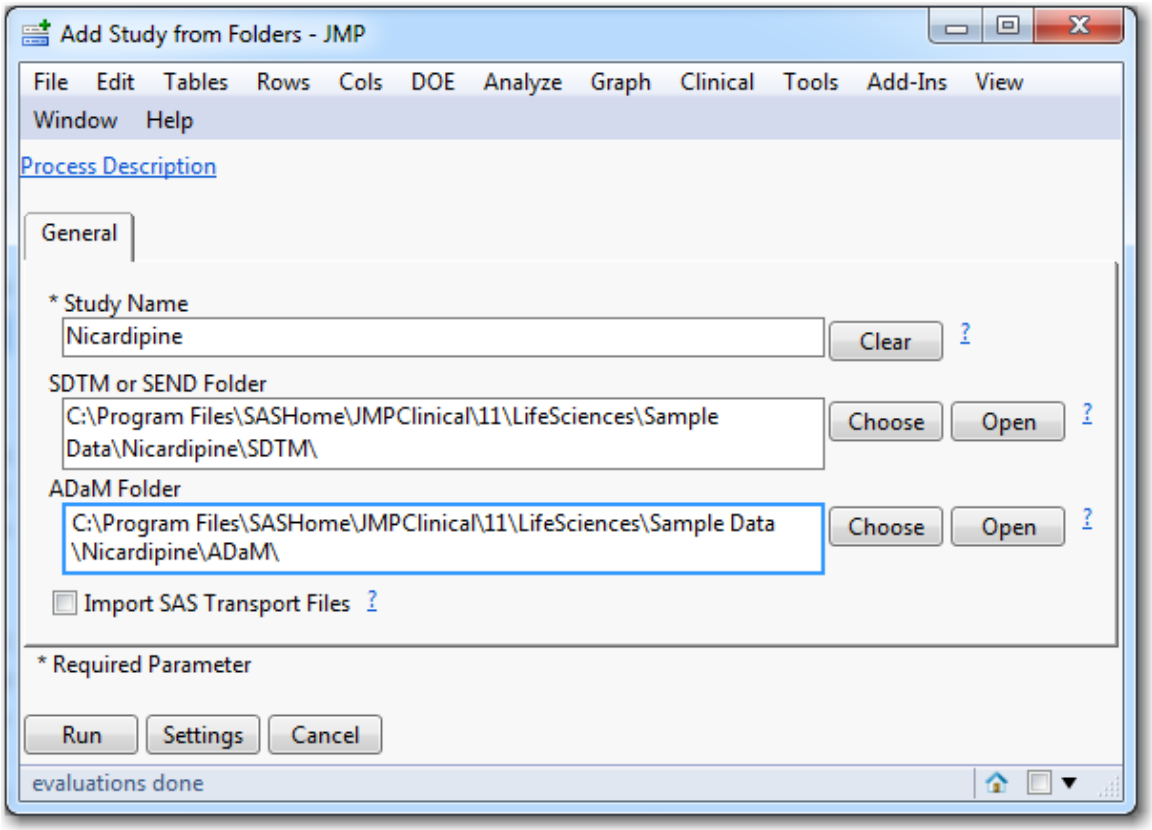


1.5 Clinical Trial Example: Nicardipine

JMP Clinical ships with sample data from a randomized, placebo-controlled clinical trial of nicardipine hydrochloride, a calcium channel blocker used to treat high blood pressure, angina, and congestive heart failure [41]. The drug comes in both oral and intravenous formulations. Because nicardipine also shows activity on blood vessels in the brain, this clinical trial was designed to ascertain whether there was any benefit in using nicardipine in the treatment of patients who experienced a subarachnoid hemorrhage (SAH, bleeding between the brain and the tissues that cover the brain). Patients were dosed with either intravenous nicardipine (up to 15 mg/kg/hr) or placebo for up to two weeks with the goal of reducing the incidence of delayed cerebral vasospasm, a leading cause of death for individuals experiencing an SAH. The trial randomized 906 subjects at 41 centers in the United States and Canada; 902 participants at 40 sites received treatment and constitute the Safety Population. The remaining four subjects are labeled as screen failures in the sample data. Throughout the book, nicardipine will refer to the drug, while **Nicardipine** will refer to the clinical trial registered within JMP Clinical.

When JMP Clinical first launches, the **Nicardipine** study will automatically be registered for use on the desktop using **Add Study From Folders**. Two folders of data sets are supplied, one for ADaM, in this case only the subject-level data set ADSL; and an SDTM folder containing the following domains: AE (adverse events), CM (concomitant medications), DM (demography), DS (Disposition), EG (ECG test results), EX (exposure), LB (laboratory test results), MH (medical history), SV (subject visits), and VS (vital signs). In case Nicardipine is not available on the desktop within the **Current Study** list, it can be added by going to **Studies > Add Study From Folders > Settings > Load > Nicardipine > OK > Run** (Figure 1.5 on page 14). To make **Nicardipine** available to everyone on a server, a **SAS Server Profile** first needs to be defined under the **File** menu (see **Add SAS Server Profile** in the documentation). Next, the server can be selected and **Nicardipine** can be added using settings similar to those previously described, though the data will first need to be copied to a location on the server (with separate folders for SDTM and ADaM files). In general, however, when operating on a server, a single individual will likely have the responsibility for adding and managing studies within the software. Studies that you have access to should be available in your **Current Study** drop-down list when JMP Clinical is launched. To work through the examples in this book, I recommend that you work with a local copy of **Nicardipine** to minimize any confusion.

Figure 1.5 *Add Study from Folders Nicardipine Sample Setting*



1.6 Organization of This Book

While this chapter has served as an introduction and brief literature review, the remaining chapters of this book describe analyses in more specific detail using the **Nicardipine** study to illustrate the various methodologies. Chapters 2 and 3 discuss an implementation of RBM based on the recommendations of TransCelerate BioPharma [1]. Here, an artificial example is created for Nicardipine using simulated data for several site metrics typically not found within the clinical trial database. Locations of trial sites were modified to include countries in Europe and Asia, and random cities were selected to represent site locations. Chapters 4 and 5 present analyses to identify potentially fraudulent data in your clinical trial. Chapter 6 describes snapshot comparison tools to highlight new and modified data and briefly summarizes the notes facilities of JMP Clinical. Finally, Chapter 7 serves as a brief epilogue. At the conclusion of each chapter, I suggest areas for possible future development. Your feedback is important to prioritize these potential improvements for future versions of the software.

References

1 TransCelerate BioPharma Inc. (2013). Position paper: Risk-based monitoring methodology. Available at: http://transceleratebiopharmainc.com/.

2 Scannell JW, Blanckley A, Boldon H & Warrington B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nature Reviews Drug Discovery* 11: 191–200.

3 US Food & Drug Administration. (2013). Guidance for industry: Oversight of clinical investigations—a risk-based approach to monitoring. Available at: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM269919.pdf.

4 Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, Genevois-Marlin E, Becker U, Bee V, Wilson V, Legrand C & Buyse M. (2012). A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials* 9: 705–713.

5 International Conference of Harmonisation. (1996). E6: Guideline for Good Clinical Practice. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf.

6 Baigent C, Harrell FE, Buyse M, Emberson JR & Altman DG. (2008). Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical Trials* 5: 49–55.

7 Morrison BW, Cochran CJ, White JG, Harley J, Kleppinger CF, Liu A, Mitchel JT, Nickerson DF, Zacharias CR, Kramer JM, Neaton JD. (2011). Monitoring the quality of conduct of clinical trials: a survey of current practices. *Clinical Trials* 8: 342–349.

8 Clinical Trials Transformation Initiative. Effective and efficient monitoring as a component of quality assurance in the conduct of clinical trials. Available at: http://www.ctti-clinicaltrials.org/what-we-do/study-conduct/monitoring. Accessed 23May2014.

9 European Medicines Agency. (2011). Reflection paper on risk based quality management in clinical trials. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/11/WC500155491.pdf.

10 Medicines and Healthcare Products Regulatory Agency. (2011). Risk-adapted approaches to the management of clinical trials of investigational medicinal products. Available at: http://www.mhra.gov.uk/home/groups/l-ctu/documents/websiteresources/con111784.pdf.

11 Eisenstein EL, Lemons PW, Tardiff BE, Schulman KA, Jolly MK & Califf RM. (2005). Reducing the costs of phase III cardiovascular clinical trials. *American Heart Journal* 149: 482–488.

12 Tantsyura V, Grimes I, Mitchel J, Fendt K, Sirichenko S, Waters J, Crowe J & Tardiff B. (2010). Risk-based source data verification approaches: pros and cons. *Drug Information Journal* 44: 745–756.

13 Funning S, Grahnén A, Eriksson K & Kettis-Linblad A. (2009). Quality assurance within the scope of good clinical practice (GCP): what is the cost of GCP-related activities? A survey with the Swedish association of the pharmaceutical industry (LIF)'s members. *The Quality Assurance Journal* 12: 3–7.

14 Grieve AP. (2012). Source data verification by statistical sampling: issues in implementation. *Drug Information Journal* 46: 368–377.

15 Usher RW. (2010). PhRMA bioresearch monitoring committee perspective on acceptable approaches for clinical trial monitoring. *Drug Information Journal* 44: 477–483.

16 Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S & Meredith S. (2012). The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi centre clinical trial. *Clinical Trials* 9: 257–264.

17 Nielsen E, Hyder D & Deng C. (2014). A data-driven approach to risk-based source data verification. *Therapeutic Innovation and Regulatory Science* 48: 173–180.

18 Buyse M, George SL, Evans S, Geller NL, Ranstam J, Scherrer B, LeSaffre E, Murray G, Elder L, Hutton J, Colton T, Lachenbruch P & Verma BL. (1999) The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine* 18: 3435–3451.

19 Evans S. (2001). Statistical aspects of the detection of fraud. In: Lock S & Wells F, eds. *Fraud and Misconduct in Biomedical Research, Third Edition*. London: BMJ Books.

20 Fang FC, Steen RG & Casadevall A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings from the National Academy of Sciences* 109: 17028–17033.

21 Grieneisen ML & Zhang M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *Public Library of Science (PLoS) ONE* 7(10): 1–15.

22 Ranstam J, Buyse M, George SL, Evans S, Geller NL, Scherrer B, Lesaffre E, Murray G, Edler L, Hutton JL, Colton T & Lachenbruch P. (2000). Fraud in medical research: An international survey of biostatisticians. *Controlled Clinical Trials* 21: 415–427.

23 Ellenberg S. (2000). Fraud is bad, studying fraud is hard. *Controlled Clinical Trials* 21: 498–500.

24 Al-Marzouki SA, Roberts I, Marshall T & Evans S. (2005). The effect of scientific misconduct on the results of clinical trials: A Delphi survey. *Contemporary Clinical Trials* 26: 331–337.

25 Baerlocher MO, O'Brien J, Newton M, Gautam T & Noble J. (2010). Data integrity, reliability and fraud in medical research. *European Journal of Internal Medicine* 21: 40–45.

26 Weir C. & Murray G. (2011). Fraud in clinical trials: detecting it and preventing it. *Significance* 8: 164–168.

27 Akhtar-Danesh A & Dehghan-Kooshkghazi M. (2003). How does correlation structure differ between real and fabricated data-sets? *BMC Medical Research Methodology* 3: 1–9.

28 Al-Marzouki SA, Evans S, Marshall T & Roberts I. (2005). Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *British Medical Journal* 331: 267–270.

29 Taylor RN, McEntegart DJ & Stillman EC. (2002). Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. *Drug Information Journal* 36: 115–125.

30 Wu X & Carlsson M. (2011). Detecting data fabrication in clinical trials from cluster analysis perspective. *Pharmaceutical Statistics* 10: 257–264.

31 CDISC Submission Data Standards Team. (2013). Study data Tabulation Model Implementation Guide: Human Clinical Trials, Version 3.2. Round Rock, TX: Clinical Data Interchange Standards Consortium.

32 CDISC Analysis Data Model Team. (2009). Analysis Data Model (ADaM), Version 2.1. Round Rock, TX: Clinical Data Interchange Standards Consortium. (UPDATE)

33 Zink RC & Mann G. (2012). On the importance of a single data standard. *Drug Information Journal* 46: 362–367.

34 US Food and Drug Administration, Center for Drug Evaluation and Research. (2011). CDER Common Data Standards Issues Document, Version 1.1. http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM254113.pdf.

35 US Food and Drug Administration. (2012). PDUFA reauthorization performance goals and procedures fiscal years 2013 through 2017. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM270412.pdf.

36 US Food and Drug Administration. (2013). Position Statement for Study Data Standards for Regulatory Submissions. http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm368613.htm.

37 Holland C & Shostak J. (2012). *Implementing CDISC Using SAS: An End-to-End Guide*. Cary, NC: SAS Institute Inc.

38 Hinrichs C & Boiler C. (2014). *JMP Essentials, Second Edition*. Cary, NC: SAS Institute Inc.

39 Sall J, Lehman A, Stephens M & Creighton L. (2012). *JMP Start Statistics: A Guide to Statistics and Data Analysis Using JMP, Fifth Edition*. Cary, NC: SAS Institute Inc.

40 Murphrey W & Lucas R. (2009). *Jump into JMP Scripting*. Cary, NC: SAS Institute Inc.

41 Haley EC, Kassell NF & Torner JC. (1993). A randomized controlled trial of high-dose intravenous nicardipine in aneurysmal subarachnoid hemorrhage. *Journal of Neurosurgery* 78: 537–547.

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Y

Z

About the Author

Richard C. Zink is a Principal Research Statistician Developer in the JMP Life Sciences division at SAS Institute. He is currently a developer for JMP Clinical, an innovative software package designed to streamline the review of clinical trial data. He joined SAS in 2011 after eight years in the pharmaceutical industry, where he designed and analyzed clinical trials for patients diagnosed with chronic hepatitis B infection, chronic myeloid leukemia, glaucoma, dry eye disease, blepharitis, or cystic fibrosis; he also participated in U.S. and European drug submissions and in two FDA advisory committee hearings. When not actively engaged in clinical development responsibilities, he supported non-clinical development, pharmaceutical sciences, and sales and marketing activities.

Richard is an active member of the Biopharmaceutical Section of the American Statistical Association, the Drug Information Association, and Statisticians in the Pharmaceutical Industry. He is currently the Statistics Section Editor for *Therapeutic Innovation & Regulatory Science* (formerly *Drug Information Journal*). He is a frequent speaker at workshops and scientific meetings and has lectured for courses in statistics and clinical trials. His research interests include the analysis of pre- and post-market adverse events, subgroup identification for patients with enhanced treatment response, and risk-based monitoring and fraud detection in clinical trials.

Richard holds a Ph.D. in Biostatistics from the University of North Carolina at Chapel Hill and has more than 20 years of SAS programming experience. This is his first book.