Executive Overview
In his excellent 1999 book, Matthews' commented as follows:

Over the last two to three decades, randomized concurrently controlled clinical trials have become established as the method which investigators must use to assess new treatments if their claims are to find widespread acceptance. The methodology underpinning these trials is firmly based in statistical theory, and the success of randomised clinical trials perhaps constitutes the greatest achievement of statistics in the second half of the twentieth century.

Credit for conducting the first pharmaceutical randomized clinical trial (RCT) is most often given to Sir Austin Bradford Hill for his work in the late 1940s on the United Kingdom Medical Research Council’s trial of the effects of the aminoglycoside antibiotic streptomycin on tuberculosis, which became the first antibiotic treatment for this disease. While RCT methodology has increased in sophistication in the intervening years, their fundamental mission in drug development of providing compelling evidence that a new drug shows therapeutic efficacy has remained steadfast, enabling drugs that have made enormous differences to people’s quality of life across the globe to be brought to market.
Table of Contents

Introduction 04

The Ethical Underpinnings of Randomized Clinical Trials 05

Noteworthy Milestones in the Development of Randomized Clinical Trials 05

The Discipline of Statistics and the Randomized Clinical Trial 08

The Hierarchy of Evidence: Where Do Randomized Clinical Trials and Meta-analyses Stand? 11

Evaluating Safety, Efficacy, and a Drug’s Benefit-risk Profile 13

Personalized Medicine 16

Additional Clinical Trial Designs 16

The New Regulatory Landscape for Risk-based Monitoring 19

Concluding Comments 21

References 22

About the Author 26
Executive Overview (continued)
This White Paper celebrates the 50th Anniversary of the Kefauver-Harris Amendments to the Food, Drug, and Cosmetic Act, which were signed into law by President John F. Kennedy on 10th October, 1962. Since then, the well-conducted RCT has become the gold standard for generating evidence of efficacy. The Amendments can be meaningfully regarded as a powerful engine that has driven and continues to drive development of the statistical and ethical sciences that underpin the RCT. The paper therefore reflects on the role that randomized clinical trials have played in drug development across the last half century, and also discusses developments that will impact clinical trial design and operational execution in the coming years. It is hoped that the paper will be of interest to all individuals interested and involved in integrated biopharmaceutical medicine, including everyone involved in lifecycle drug development and also health professionals who prescribe, dispense, and administer pharmaceutical medicines to patients.

Introduction
The Kefauver-Harris Amendments to the Federal Food, Drug, and Cosmetics Act of 1938 (FFDCA) were signed into law by President Kennedy a half-century ago on October 10th, 1962. The signing into law of the FFDCA, which revised the Pure Food and Drug Act of 1906 and had been under congressional discussion for several years, was galvanized by the elixir sulfanilamide tragedy in which more than 100 people were fatally poisoned by an ingredient in the elixir following its ingestion. Signing of the Amendments into law was spurred on by the thalidomide tragedy in Europe. Thalidomide was first marketed in 1956 in Germany for the treatment of insomnia and vomiting in early pregnancy. In 1961 a sizeable increase in the incidence of congenital birth defects was noted. These defects were typically an absence or reduction of the long bones of the limbs with normal or rudimentary hands and feet. The association of these conditions with thalidomide was not recognized for several years after the drug was marketed, and several thousand babies in Europe suffered from this congenital condition.

While public awareness of the events in Europe was a powerful motivator for congressional action for further drug law reform, the United States did not suffer the same tragedy from the use of thalidomide. Having seen the reports coming from Europe, Dr. Frances Kelsey, a newly appointed reviewer at the Food and Drug Administration (FDA), undertook considerable research and, as a result, took a firm stance against the drug’s approval. In recognition of her diligence she was awarded the President’s Award for Distinguished Federal Civilian Service two months before the Amendments were signed.3

The RCT (more accurately but less commonly called the randomized concurrently-controlled clinical trial since a concurrent control treatment arm is a fundamental aspect of the trial’s comparative nature) became the key methodology for providing the compelling evidence of efficacy required by the Amendments. The Amendments and the RCT can therefore be regarded as very closely related in the history of modern biopharmaceutical medicine.
The Ethical Underpinnings of Randomized Clinical Trials

Derenzo and Moss\(^1\) captured the importance of ethical considerations in all aspects of clinical studies as follows:

Each study component has an ethical aspect. The ethical aspects of a clinical trial cannot be separated from the scientific objectives. Segregation of ethical issues from the full range of study design components demonstrates a flaw in understanding the fundamental nature of research involving human subjects. Compartmentalization of ethical issues is inconsistent with a well-run trial. Ethical and scientific considerations are intertwined.

From the scientific perspective, an inappropriate study design and/or a poorly conducted trial is generally incapable of answering a research question. Participants voluntarily take part in clinical research with the understanding that their participation may not benefit them directly (e.g., they may be randomized into a placebo treatment arm), but it will provide information that will be useful to a much larger group of people should the drug be approved for marketing and hence available to perhaps millions of patients. This is one of the ‘benefits’ that is weighed against the ‘risks’ of participants’ exposure to a drug under development. From the ethical perspective, therefore, three consequences of poorly designed and conducted clinical trials become clear. First, if the trial does not permit the best possible information to be obtained, the participants’ expectations have been violated. Second, a poor trial may lead to a drug failing to be approved for marketing when in reality it is safe and efficacious. Patients who would have benefited from the drug will therefore be denied the opportunity to receive it. Third, a poor trial may lead to a drug being approved when in reality it is not acceptably safe, where the term ‘acceptably safe’ refers to a favorable benefit-risk balance despite risks being present.\(^4\)\(^5\)

The Amendments played a seminal role in the development of clinical research ethics. The practice of informed consent and the creation of Institutional Review Boards, for example, are directly traceable to them, and ongoing discussions led to the Belmont Report, which addressed the “basic ethical principles” of respect for persons, beneficence, and justice.\(^6\)

Noteworthy Milestones in the Development of Randomized Clinical Trials

In his 1999 textbook, Matthews\(^1\) commented as follows:

Over the last two to three decades, randomized concurrently controlled clinical trials have become established as the method which investigators must use to assess new treatments if their claims are to find widespread acceptance. The methodology underpinning these trials is firmly based in statistical theory, and the success of randomised clinical trials perhaps constitutes the greatest achievement of statistics in the second half of the twentieth century.

This section notes some milestones in the development of RCTs.

It is not possible to separate the ethical aspects of a clinical trial involving human subjects from the trial’s scientific objectives. Ethical considerations are of paramount importance in clinical research.
The Medical Research Council’s Trial of Streptomycin for Pulmonary Tuberculosis

Credit for conducting the first pharmaceutical RCT is often given to Sir Austin Bradford Hill and his 14 colleagues (the Streptomycin in Tuberculosis Trials Committee, chaired by Dr. Geoffrey Marshall) for his work in the late 1940s on the United Kingdom Medical Research Council’s (MRC’s) trial of the effects of the aminoglycoside antibiotic streptomycin on pulmonary tuberculosis.7 The control group treatment consisted of bed rest (the standard of care at the time) and the streptomycin group treatment consisted of bed rest plus intramuscular administration of 2 grams/day of streptomycin, given in four injections at six-hourly intervals. Compelling evidence of efficacy was provided, and streptomycin subsequently became the first antibiotic treatment for this disease.

Control groups had certainly been used in medical research prior to this trial, but often the method of allocating participants to one of two treatment groups was alternate allocation, simply placing the next individual entering the trial in the alternate treatment group to the one entered by the previous individual.8 The method of allocating (randomizing) participants to either treatment group on this occasion was as follows:7

Determination of whether a patient would be treated by streptomycin and bed-rest (S case) or by bed-rest alone (C case) was made by reference to a statistical series based on random sampling numbers drawn up for each sex at each centre by Professor Bradford Hill; the details of the series were unknown to any of the investigators or to the co-ordinator and were contained in a set of sealed envelopes, each bearing on the outside only the name of the hospital and a number. After acceptance of a patient by the panel, and before admission to the streptomycin centre, the appropriate numbered envelope was opened at the central office: the card inside told if the patient was to be an S or a C case, and this information was then given to the medical officer of the centre.

Additional informative commentary on this trial is provided by Yoskioka8 and by other authors in a 1998 issue of the British Medical Journal published 50 years following the original publication of the trial’s results.

Tuberculosis remains a global health issue of considerable proportions and a very active area of clinical research. Fortunately, as Lienhardt and colleagues9 recently observed, “a portfolio of promising new compounds for the treatment of tuberculosis is on the horizon.” In the same issue of the Journal of Infectious Diseases, authors from the MRC’s Clinical Trials Unit and colleagues noted that innovative trials designs should be considered to speed drug and combination treatment regimen development for the treatment of tuberculosis.10 The sophisticated designs they discussed bear witness to the continuing evolution of the RCT, and will be discussed further later in this Paper.

Development of the RCT in the United States

Cook and DeMets11 observed that the era of modern clinical trials in the United States can be regarded as beginning with the Coronary Drug Project (CDP), which ran from 1966 to 1975.12 The trial was sponsored by the National Heart Institute, which later became the National Heart, Lung, and Blood Institute (NHLBI, which is part of the National Institutes of Health [NIH]). The trial’s focus was secondary prevention, and 8,341 men between the ages of 30 and 64 years old who had recently survived myocardial infarction were randomized to one of six treatment arms that included five active treatments and a placebo, with all were participants also receiving the standard of care treatment at that time. The active treatments were representative of the drugs used at the time: 2.5 mg/day of conjugated estrogens, 5.0 mg/day of conjugated estrogens, 1.8 gm/day of clofibrate, 6.0 mg/day of dextrothyroxine, and 3.0 gm/day of niacin. The participants were followed for the occurrence of another cardiovascular event such as death or a second heart attack. Fifty-three
investigative sites were included, along with a coordinating center and many committees, including a Steering Committee and a Data and Safety Monitoring Committee. The two estrogen treatment arms and the dextrothyroxine treatment arm were terminated before the end of the trial because of adverse effects experienced by the participants in them. No evidence of efficacy was found for clofibrate, while niacin showed modest benefit in decreasing nonfatal recurrent myocardial infarction but did not decrease total mortality.

A 15-year follow-up study\(^1\) found that mortality from all causes for participants who had received niacin was 11% lower than those who were in placebo treatment arm (52.0% versus 58.2%), a result that attained statistical significance. The authors commented that “This late benefit of niacin, occurring after discontinuation of the drug, may be a result of a translation into a mortality benefit over subsequent years of the early favorable effect of niacin in decreasing nonfatal reinfarction or a result of the cholesterol-lowering effect of niacin, or both.”

---

**The National Institutes of Health, and particularly the National Heart, Lung, and Blood Institute, played an important role in sponsoring many ground-breaking cardiovascular randomized controlled trials. Lessons learned from these were then transferred to other institutes and centers at NIH.**

---

**Industry-sponsored Trials**

Large industry-sponsored RCTs did not become common until the late 1980s and 1990s.\(^1\) As an example, consider the Scandinavian Simvastatin Survival Study (4S).\(^6\) The trial was designed to evaluate “the effect of cholesterol lowering with simvastatin on mortality and morbidity in patients with coronary heart disease (CHD).” A total of 4,444 patients with angina pectoris or a previous heart attack who also had high serum cholesterol levels and were on a lipid-lowering diet were randomized to the simvastatin treatment group or the placebo treatment group. Compared with individuals receiving standard of care, simvastatin produced highly significant reductions in the risk of death and morbidity in individuals with CHD followed for a median of 5.4 years. Importantly, “The improvement in survival produced by simvastatin was achieved without any suggestion of an increase in non-CHD mortality... No previously unknown adverse effects were apparent in this trial. Thus the substantial and sustained reduction of total and LDL cholesterol in the simvastatin group was not associated with any serious hazard.”\(^6\)

**Demonstrating a Lack of Compelling Evidence of Efficacy**

In addition to providing compelling evidence of efficacy, RCTs can also demonstrate a lack of compelling evidence of efficacy even though previous investigations in a drug’s development program had suggested such evidence would be found. Torcetrapib was a novel cholesteryl ester transfer protein (CETP) inhibitor that was demonstrated to inhibit the development of atherosclerosis in nonclinical studies (a rabbit model), and, in early-phase clinical studies, to increase high-density lipoprotein cholesterol (HDLc) between 60% to 100% while at the same time lowering low-density lipoprotein cholesterol (LDLc) by up to 20%.\(^7\) Based on ‘conventional wisdom’ regarding these two cardiovascular safety biomarkers (higher HDLc is good, and lower LDLc is good), this evidence suggested a cardioprotective effect of torcetrapib. The Investigation of Lipid Level Management to Understand its Impact in Atherosclerotic Events (ILLUMINATE) trial therefore
tested the proposition that torcetrapib would reduce the risk of clinical cardiovascular events. However, torcetrapib was associated with an increased risk of major cardiovascular events, and also increased mortality (from both cardiovascular and noncardiovascular causes). The drug’s sponsor terminated ILLUMINATE prematurely at the recommendation of the trial’s independent Steering Committee, based on advice from the trial’s Independent Data and Safety Monitoring Board.18

The Discipline of Statistics and the Randomized Clinical Trial

RCTs provide information, knowledge, and evidence that facilitates rational decision making, both on the part of regulatory agencies at the time of marketing approval decisions and of prescribing physicians when deciding, in conjunction with their patients, whether a particular treatment option is appropriate for patients on a case-by-case basis.

Statistics can be thought of as an integrated discipline that is important in all of the following associated activities:4

- Identifying a research question that needs to be answered;
- Deciding upon the design of the clinical trial, the methodology that will be employed, and the numerical information (data) that will be collected;
- Presenting the design, methodology, and data to be collected in a Study Protocol. This study protocol specifies the manner of data collection and addresses all methodological considerations necessary to ensure the collection of optimum quality data for subsequent statistical analysis.
- Identifying the statistical techniques that will be used to describe and analyze the data in the study protocol (or an associated Statistical Analysis Plan that is written in conjunction with the study protocol);
- Describing and analyzing the data to evaluate if there is compelling evidence that the drug is safe and effective.
- Presenting the results of a clinical study to a regulatory agency in a clinical study report and presenting the results to the clinical community in conference presentations and journal publications.

All of these activities are essential for RCTs, and professional statisticians should be involved at every step.

Randomization: Why Do We Need It?

Biological considerations are at the heart of biopharmaceutical drug development in several ways. First, clinical trials investigate topics of clinical relevance, and clinical relevance is intimately related to biological relevance. The ultimate goal in this field of clinical research is to develop a new compound that is biologically active, acceptably safe, well tolerated, and useful in the treatment of biological states that are, or may become, of clinical concern. A second central importance of biological considerations is individual variation: not everyone responds to the same drug in the same way, a statement that remains true when factors such as sex, age, weight, and ethnicity are taken into account. This observation provides a direct link between biological science and statistical science in the setting of clinical research.

One reason for individual variation in drug response is individual variation in drug metabolism. The tricyclic antidepressant nortriptyline provides an illuminating example of such genetic influence. The majority of patients taking this drug require around 75-100 milligrams per day (mg/day) to reach the intended steady-state blood plasma concentration. Poor metabolizers (whose abnormal genetic variant enzymes break down the drug less efficiently than usual) require only 10-20 mg/day. At the other end of the continuum, ultra-rapid metabolizers (whose abnormal variant enzymes
break down the drug more efficiently than usual) require around 300-500 mg/day to achieve the same blood concentration.19

A second reason why individuals respond differently to the same drug is genetic differences in the structure of the target receptor, the biological structure (usually a macromolecule) with which the drug is intended to interact to produce the desired beneficial response. Additionally, there can be genetic differences in the structure of off-target receptors, biological structures with which the drug is not supposed to interact. Such interaction leads to adverse drug responses, the severity of which can vary among individuals.

Genetic variation and the resultant individual differences in drug response therefore lead to the following categorization of responders:24

- **Optimal.** These individuals show the intended therapeutic response, and do not have adverse drug reactions.
- **Suboptimal.** Such individuals show less than the intended level of therapeutic response.
- **Supra-optimal.** These individuals show a greater therapeutic response than intended.
- **Adverse.** Such individuals show relatively serious adverse drug responses.

This variation is precisely the reason that the process of randomization is needed.

---

**Biological considerations are at the heart of biopharmaceutical drug development. Genetic variation that leads to different people responding to the same drug in a different manner is precisely the reason that the process of randomization is needed in therapeutic confirmatory (comparative) clinical trials. This observation provides a direct link between biological science and statistical science in the setting of clinical research.**

---

**The Process of Randomization**

The process of randomization involves randomly assigning participants in RCTs to one of various treatment arms in a parallel-groups clinical trial (where each person participates in just one of the treatment arms: in a cross-over design each participant completes all treatment arms). The goal of randomization is to control for the many potentially confounding influences that cannot be controlled for (e.g., height and weight) or cannot be determined by simple visual observation (e.g., the nature of an individual’s metabolic pathways and on-target and off-target drug receptors). That is, the goal is to minimize to the greatest extent possible any selection bias in participant assignment to the treatment groups. In statistical language, participants in the trial have an independent chance of receiving the different treatment arms. The simplest design is one in which half of the participants are randomized into the drug treatment arm and the other half are randomized into the control treatment arm. The randomization procedure occurs after a potential participant’s eligibility for a clinical trial has been determined (i.e., the study protocol inclusion and exclusion criteria have been satisfied by that individual) and before the individual provides any trial data.
A randomization schedule does not have to place 50% of participants into each of two arms, but there is a sound statistical reason to do so: the power of statistical tests performed at the end of the study to determine if there is a statistically significant mean difference in drug response between the two treatment groups is greatest when the number of participants in the two arms is equal. However, there are certainly circumstances in which a sponsor may decide to use unequal randomization, e.g., using a 2:1 schedule in which two-thirds of the participants are randomized into the drug treatment group, and one third into the control group. Such a schedule enables the sponsor to collect safety data on 67% of the participants rather than 50% of the participants, while still maintaining a certain degree of statistical power to look for a potential statistically significant mean difference between the groups with regard to the primary efficacy endpoint of interest.

The fundamental statistical question of interest in a clinical trial involving participants randomized into a drug treatment arm and a control arm becomes: Is the variation in drug response between the two groups statistically significantly greater than the variation within the two groups? This question can be answered by employing the statistical methodology of Analysis of Variance (ANOVA), which partitions the total variation within the overall data set into ‘between-groups variance’ and ‘within-groups variance.’ Considering within-groups variance to remain equal, the greater the between-groups variance, the greater the likelihood of obtaining a statistically significant difference between the mean responses of the two groups. Conversely, considering between-groups variance to remain equal, the greater the within-groups variance the less the likelihood of obtaining a statistically significant difference in the mean responses of the two groups. The test statistic in ANOVA is called $F$, which pays respect to the visionary statistician Sir Ronald Fisher who developed this approach.\textsuperscript{20,21} $F$ is calculated as a ratio as follows:

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$

As for other test statistics, $F$ has to reach a certain size for the result of the analysis to attain statistical significance. This size varies from trial to trial, and is dictated by the total number of participants included. However, it can be stated for all cases that, to attain statistical significance, an $F$ value must be greater than 1 (unity). If the difference between the mean response in the drug treatment group and the mean response in the control treatment group, i.e., the drug’s treatment effect, attains statistical significance, the process of randomization has facilitated the provision of compelling evidence that administration of either the test drug or the control drug to each participant, i.e., a systematic influence, has influenced drug (biological) responses to a degree that is over and above any difference that could be ascribed to chance factors alone. That is, compelling evidence of the drug’s efficacy has been provided.

Fisher’s work on scientific experimental methodology was initiated in the field of plant breeding and agriculture, another domain in which natural variation is evident and randomization is therefore necessary. References to more of his writings\textsuperscript{22,23} and to introductory conceptual and computational statistics books discussing experimental research methodology\textsuperscript{24,25} are provided.
Attributes of a Good Clinical Trial

Piantadosi defined a clinical trial as an experiment that tests a medical treatment on human participants. The use of human participants requires ethical and biological considerations to be given consideration, as already discussed. Other considerations include study design, experimental methodology, operational execution, and statistical analysis. If asked to summarize the purposes of these factors in one sentence each, the following might be suitable:

- **Study design**: Designing a clinical trial to facilitate the collection of data, i.e., unbiased and precise numerical representations of biologically important information that best answers the study’s research question, i.e., the question the clinical trial is being conducted to answer.
- **Experimental methodology**: Considering and implementing all necessary procedures that, if executed correctly, allow the acquisition of optimum quality data.
- **Operational execution**: Conducting all operational and experimental tasks correctly and therefore successfully acquiring optimum quality data.
- **Statistical analysis**: Describing, summarizing, analyzing, and interpreting the data collected to answer the study’s research question.

With regard to the nature of a clinical trial, Piantadosi commented as follows: “The most critical and difficult prerequisite for a good study is to select an important feasible question to answer. Accomplishing this is a consequence primarily of biological knowledge. Conceptual simplicity in design and analysis is a very important feature of good trials. Good trials are usually simple to analyze correctly.” This sentiment is also reflected in ICH Guidance E8: “Clinical trials should be designed, conducted, and analyzed according to sound scientific principles to achieve their objectives, and should be reported appropriately. The essence of rational drug development is to ask important questions and answer them with appropriate studies. The primary objectives of any study should be clear and explicitly stated.”

The Hierarchy of Evidence: Where Do Randomized Clinical Trials and Meta-analyses Stand?

In recent years, meta-analyses have attained increasing prominence in the evidence-based medicine literature. Meta-analysis facilitates a quantitative evaluation of the evidence provided by two or more individual trials that have addressed the same research question. It commonly involves the statistical combination of summary statistics from various trials (study-level data, i.e., the study’s treatment effect and the variance associated with it are extracted from the published report of each trial), but it also refers to analyses performed on the combination of participant-level data (a more powerful strategy when possible). The conceptual basis of meta-analysis is straightforward: more data provide a better opportunity to get an optimum-quality answer to a research question. However, Turner and Durham commented as follows:
If all of the components involved in conducting a meta-analysis are performed appropriately, and the extent to which the results are helpful is not overstated (that is, any limitations are appropriately acknowledged and shared whenever and wherever communicating the results), the results can be informative and instructive. Unfortunately, however, it is easier than one might suspect to conduct a meta-analysis inappropriately and then to overstate the results in a variety of circumstances.

Kay expressed his concerns more colorfully, noting that to ensure that a meta-analysis is scientifically valid, the analysis must be planned and conducted in an appropriate way, and that “It is not sufficient to retrospectively go to a bunch of studies that you like the look of and stick them together!” Therefore, while an analysis is certainly conducted, the term meta-methodology usefully captures everything else required to provide a reliable answer.30

Even though conducting a meta-analysis does not require a new trial to be conducted, it is still a research method in its own right. Given this acknowledgment, the term meta-analysis, while at first appearing appropriately descriptive, does not adequately capture the need for methodological rigor in the full array of required actions. Turner therefore suggested that the term meta-methodology can be meaningfully employed to convey the need for paramount methodological rigor when conducting the full array of actions required for its meaningful execution. Perhaps both terms can be meaningfully used, with meta-analysis referring to the computational and interpretive process involved in executing the central analysis itself, and with the term meta-methodology being an all-encompassing name that refers to all necessary meta-methodological aspects of conducting this form of investigation, i.e., the preparation of the dataset to be analyzed, all appropriate analyses that must be conducted (including tests for heterogeneity and robustness), and the appropriate presentation of results and interpretations in all venues and circumstances.

Different authors have differing views on the relative strength of evidence provided by individual clinical trials and meta-analyses, and there is not space enough here to debate this issue. Suffice it currently to make two observations. First, for evidence-based medicine to be practiced, sound evidence is required. Well-conducted individual trials and well-conducted meta-methodology and meta-analysis can both provide sound evidence. Second, the way in which the results of any research investigation employing either experimental methodology or meta-methodology are communicated to the scientific community, practicing physicians, and, increasingly, the general public is of great importance. Turner observed that “following the publication of their article in a journal, some meta-analysts disseminate their findings in the mass media with a bravado that markedly departs from calm, scientific, and clinical discourse, and seemingly with the expectation that the nation’s physicians will change their practice of medicine immediately.” Fortunately, many others are more judicious.

Clinical Practice and Evidence-based Medicine
It is possible for a practicing physician to think initially that RCTs are the province of clinical research scientists and physicians who act as Principal Investigators (hence being in charge of clinical trial execution at investigational sites at which trials are conducted), and not the province of those exclusively engaged in providing patient care. Given a few moments’ thought, however, an alternative viewpoint recommends itself. The importance of the RCT to all physicians is that clinical research informs clinical practice and evidence-based medicine, and practicing physicians can benefit from having sufficient knowledge about RCTs to understand their role in placing new drugs within their treatment armamentaria and generating the evidence contained in treatment practice guidelines.
Once drugs have been approved by regulators, and particularly when there are multiple classes of drugs available for prescription, evidence-based practice guidelines issued by professional societies and reputable organizations can be of considerable assistance to physicians in clinical practice. These can theoretically contain recommendations based on individual RCTs and on meta-analyses. However, whatever the evidence base, it ultimately falls upon individual physicians to make treatment decisions that are in their patients' best interests. Katz captured this sentiment as follows:31

If our patient is older than, younger than, sicker than, healthier than, ethnically different from, taller, shorter, simply different from the subjects of a study, do the results pertain?...All of the art and all of the science of medicine depend on how artfully and scientifically we as practitioners reach our decisions. The art of clinical decision-making is judgment, an even more difficult concept to grapple with than evidence.

Sackett et al. defined evidence-based medicine as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.”32 There are two components to evidence-based medicine, and two related sets of responsibilities. The first component is clinical research. Clinical research is a scientific endeavor that provides evidence concerning potential therapeutic interventions (with current discussions focusing on one particular therapeutic intervention, biopharmaceutical drugs). Once clinical trials have been conducted, the evidence obtained is published in clinical and medical journals. Everyone involved in clinical research has the responsibility to provide the best possible evidence in this manner. This includes all aspects of clinical research, including study design, experimental methodology, operational execution, data management, analysis and interpretation of data, and accurate and complete representation and communication of study findings as published in journals.

The second component of evidence-based medicine is clinical practice. Clinicians have the responsibility of providing the best possible care to each of their individual patients. One component of being able to provide this care is remaining aware of pertinent evidence published in journals, and also being able to decide for themselves whether the evidence presented is good evidence, and if the message conveyed by a meta-analysis is justified based on the quality of its conduct.

Evaluating Safety, Efficacy, and a Drug’s Benefit-risk Profile
While evaluating a drug’s efficacy is a critical component of drug development (and has been the focus of this paper to date), evaluating a drug’s safety profile must be of primary importance. However, assessing safety can be much more complex than assessing efficacy. Data relevant to the assessment of a drug’s safety are collected in all phases of drug development, and a portfolio of safety data is accumulated across the entire drugs development program (both nonclinical and clinical).

As noted earlier in this paper, regulatory guidance was largely driven by safety concerns in the previous century: the elixir sulfanilamide and the thalidomide tragedies bear witness to this. Given this occurrence, Senn33 observed a paradoxical trend in the development of statistical theory relating to clinical trials:

It is a curious fact that whereas the original inspiration for much of the legislation covering drug development has its origin in concerns about the safety of pharmaceuticals... much of the statistical theory of planning clinical trials has to do with investigating efficacy rather than safety.

There are several explanations for this that, in themselves, are very reasonable. First, drug development programs are designed with therapeutic needs in mind. Second, while some adverse
events (side effects) may be expected based on the mechanism of action of the therapeutic benefit, others that are unexpected can certainly occur. Third, unexpected adverse events are typically rare (unexpected serious adverse events can be very rare), which makes them hard to detect in the best designed and conducted preapproval trials. Nonetheless, despite the statistical challenges associated with identifying and assessing a very wide range of potential safety issues, we must do everything we can. As Senn continued, “The subject is extremely important…if there are [safety issues] with a drug, the sooner they are discovered the better.”

Recently, specific safety assessments have been crystallized, and instructive examples are found within the domain of cardiovascular drug safety. In 2005, the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) released its E14 Guidance that addressed the cardiac safety of noncardiac drugs. This guidance, which focuses on the provision of compelling evidence that a noncardiac drug does not unacceptably impact the heart’s normal electrical activity, has been adopted in North America, Europe, and Japan: a more recent “Questions & Answers” document associated with this guidance document provides updated thoughts on this issue. In 2008, the FDA released a guidance designed to prospectively exclude unacceptable cardiovascular risks associated with new antidiabetic agents for the treatment of type 2 diabetes mellitus (T2DM), and similar guidance has been released by the European Medicines Agency (EMA). While no regulatory guidance has yet been issued, scientists and regulators are currently discussing how best address blood pressure responses to noncardiovascular (non-antihypertensive) drugs. This ‘return to a focus on safety’ fits in directly with the topic of benefit-risk estimation.

**Estimating a Drug’s Benefit-risk Profile**

Benefit-risk assessment can be represented as follows:

\[
\text{Benefit-risk estimate} = \frac{\text{Estimate (probability and degree) of benefit}}{\text{Estimate (probability and degree) of risk}}
\]

Consider benefit first, which is operationalized in clinical trials as efficacy (as discussed throughout this paper) and, once the drug has been marketed, as effectiveness. The inclusion of both probability and degree addresses the fact that the likelihood of the drug working in a particular patient or patient population is just as important as the degree to which it works when it does work. The same is true for risk, which captures both the probability and degree of harm. If the probability of a relatively serious unwanted drug response is 1 in 1,000,000, the estimation of the benefit-risk balance for a given estimate of benefit is quite different than if the probability of the side effect is 1 in 100.

Benefit-risk assessment is a key facet of decision-making at both the regulatory (public health) level and the level of individual patient pharmacotherapy. It requires consideration of both benefit and risk. For a drug to receive regulatory approval, the regulatory agency must find a drug to have a favorable benefit-risk profile: that is, the benefit to the population as a whole must outweigh any potential risk to certain individuals. These individuals will be protected by appropriate drug labeling, which will advise physicians against prescribing the drug to individuals with certain characteristics, and other risk mitigation strategies such as the use of companion diagnostics (discussed shortly). At the individual patient level, prescribing physicians must assess the benefit-risk balance for every patient on a case-by-case basis, and, for the drug to prescribed, the physician must believe that the potential benefit to the patient outweighs any potential risk.
It is important to acknowledge that a drug’s benefit-risk profile can change over time, for several reasons. First, regulatory decisions to grant marketing approval are by definition made using data available at that time. However, as noted by the Institute of Medicine,38 “The approval decision does not represent a singular moment of clarity about the risks and benefits associated with a drug—preapproval clinical trials do not obviate continuing formal evaluation after approval.” As Turner4 observed, “The simple statistical truth is that (very) rare adverse drug reactions are probabilistically (very) unlikely to be detected in preapproval trials.” The ‘rules of threes’40 addresses the sample size in preapproval trials that would be needed to be 95% confident that a single case of an identified adverse of interest would be seen: the size is roughly three times the reciprocal of the frequency of the event in the general population. For example, the sample size that would therefore be needed to observe at least one event when the frequency of the event is 1 in 10,000 would be around 30,000 participants. (Frequencies of more rare events can be 1 in 100,000 or even 1 in a million.) A trial of that size is not feasible at the preapproval stage of lifecycle drug development, but once a drug has been approved and prescribed to a large number of patients, events of this frequency can be detected. Postmarketing surveillance (both passive and active) is therefore extremely important. If, over time, the risks are seen to outweigh the benefit, market withdrawal is likely.

A second reason for a drug’s benefit-risk profile changing is related not to new information gathered about the drug, but the future availability of other drugs. If a drug that is approved some time after the drug in question provides equal benefit but is associated with less risk, the first drug becomes less attractive. The same is true if the new drug is associated with both equal benefit and risk, but is more convenient to take (e.g., weekly or monthly rather than daily).

Benefit-risk assessment play a very important role in both pharmacotherapy and drug development, which is why the Sentinel Initiative noted that the FDA regards benefit-risk analysis to be “one of the important facets of the science of safety that urgently requires additional development.”41 Recently, this topic has been receiving a lot of attention in the literature. One aspect of interest is moving from a qualitative approach (it is easy to understand that benefit needs to be greater than risk) to a more quantitative approach, where both benefit and risk are more quantitatively assessed, allowing a more quantitative assessment of the benefit-risk balance. A selection of papers on the topic of benefit-risk is provided.42-48

“If we want to benefit from medicine, we must accept some risks. We first need to consider the risks when deciding whether or not to use the medicine. When we have decided to take the medicine, because the likely benefit sufficiently outweighs the risks, we have to understand how to minimize these risks. The user thus needs two quite separate kinds of information about possible harm: first, a realistic assessment of benefits and risks when the drug is properly used; second, what precautions and circumspections ‘proper use’ requires.”49
Personalized Medicine
As the Sentinel Initiative also noted, an increasing focus on safety is in part “a result of an emerging science of safety.” The science of safety includes a growing understanding of diseases and their origins at the molecular biological level, which facilitates an increasing knowledge of adverse reactions to, and also therapeutic benefit from, pharmacotherapy. Molecular-level investigations are revealing information about “the unique genetic and biological features of each person that someday will help determine how he or she responds to treatment.” Progress toward this goal is being made.

Consider first the case of abacavir, an antiretroviral used against infection with the human immunodeficiency virus (HIV). Approximately 6% of individuals carry the HLA-B*5701 allele, an abnormal genetic variant that is strongly associated with hypersensitivity to abacavir. This hypersensitivity is a multi-organ systemic illness that can have life-threatening complications if the drug is continued while symptoms progress, or if it is given again following termination of treatment once the symptoms have dissipated (re-challenge). Screening potential recipients of the drug for the presence of the HLA-B*5701 allele and not prescribing the drug for those who have it has proved to be a successful strategy in reducing hypersensitivity reactions, while also allowing the large majority of patients to take the drug without fear of a serious adverse drug reaction.

Moving now to the role of companion diagnostics with regard to benefit, consider the case of trastuzumab, a drug used to treat early-stage breast cancer that is human epidermal growth factor receptor 2-positive (HER2+). Two types of test are available to determine HER2 status. One is a fluorescence in situ hybridization (FISH) test, which assesses whether or not a patient’s cancer cells have a normal number of HER2 genes. The other is an immunohistochemistry (IHC) test to assess how much HER2 protein there is on the surface of the cancer cells. Potential recipients of this drug must have a HER2 test to determine that their cancer is HER2+ before taking trastuzumab, as benefit has only been demonstrated in patients whose tumors are HER2+. Since there is always some risk associated with any drug, taking a drug that does not have a chance of being therapeutically beneficial is not an option since such action would immediately result in an unfavorable benefit-risk balance.

Two drugs approved by the FDA during 2011 in combination with an FDA-approved companion diagnostic test provide interesting case studies. Crizotinib was approved to treat certain patients with late-stage (locally advanced or metastatic) non-small cell lung cancers who express an abnormal variant anaplastic lymphoma kinase (ALK) gene in conjunction with a companion diagnostic test that will help determine if a patient for whom the drug is being considered has the abnormal variant ALK gene. Similarly, vemurafenib, indicated for melanoma, can only be prescribed for patients with a certain abnormal variant of the BRAF gene, BRAF\(^{V600E}\), as identified by an FDA-approved test. For a more detailed discussion, see a review by Turner.

Additional Clinical Trial Designs
The study design that has been discussed to date can be called a fixed design or a fixed sample design, one in which there is no latitude to deviate from the precise plans detailed in the study protocol and the statistical analysis plan. The study design is specified at the beginning of the trial, the number of participants that will be enrolled is clearly stated, and the data analysis plan is laid out in detail before the RCT starts. Once the trial has commenced it progresses as planned until its conclusion, at which time the statistical analyses are conducted. However, alternate approaches have also been developed.
Group Sequential and Adaptive Study Designs

In contrast to fixed designs, group sequential and adaptive designs incorporate interim analyses, analyses that are performed during the execution of the trial. These designs are considered here in turn.

The purpose of interim analysis in group sequential trials is to determine whether the trial should be terminated at that point for one of several reasons. One reason would be that the interim analyses provided compelling evidence (even at a relatively early point in the trial) that the investigative drug was effective. Linking back to our earlier discussions of ethical considerations, if the trial were to be continued at this time, future participants randomized into a placebo control group would be receiving the control drug for no justifiable reason. As noted previously, a central tenet of participant participation in clinical trial is clinical equipoise. Once it is known that a drug is effective, clinical equipoise no longer exists, and it is therefore ethically suspect from this point forward for some participants to receive a treatment known to be less effective. A second reason to stop the trial is that the interim analysis provided compelling evidence of toxicity. A third is provision of compelling evidence that the trial would not be able to achieve its intended purpose even if continuing on to the maximum number of participants specified in its protocol: this scenario is termed stopping for futility.

The number of possible interim analyses to be conducted must be specified in a group sequential study’s protocol. Consider an example in which the efficacy an investigational drug is being compared with a control drug, and a maximum of 1,000 participants will be recruited, with 500 being randomized into the drug treatment arm and 500 into the control treatment arm. The protocol allows for a maximum of four interim analyses, and, if the trial is not terminated at any of these occasions, a final analysis will be conducted. The first interim analysis will occur after 200 participants (100 in each treatment arm) have completed their participation in the trial. If there is compelling evidence of efficacy, the trial will be stopped at that point. If not, the trial process until a second cohort of 200 participants has completed their participation, when the second interim analysis is completed. The data from all 400 participants is used for this analysis. If there is compelling evidence of efficacy, the trial will be stopped at that point, and if not it will proceed until a third cohort of participants has completed their participation in the trial. The third interim analysis then occurs, using data from all 600 participants who have completed their participation. If the trial continues, the fourth interim analysis occurs when a total of 800 participants have completed their participation. If not stopped at that point, the trial then proceeds to its protocol-defined endpoint, by which time 1,000 participants have completed their participation, and the final analysis is conducted.

Two statistical points should be made here. First, for a given degree of efficacy, the greater the number of participants whose data are analyzed, the greater the likelihood of it obtaining statistical significance, a prerequisite of the drug receiving marketing approval (the efficacy must also be of sufficient magnitude that it is considered clinically significant). When the trial in the previous example was being designed, researchers thought that it might be necessary to enroll 1,000 participants to demonstrate statistically significant efficacy. However, interim analysis facilitated the opportunity to demonstrate such efficacy after smaller numbers of participants had completed their participation. Second, the statistical analyses have to take into account the fact that a total of five analyses (four interim plus the final analysis) may be conducted, and hence multiplicity corrections are made. The reason for this is that the greater the number of analyses performed, the greater the likelihood that a statistically significant result will be obtained purely by chance, i.e., a false-positive result will be obtained. It is quite legitimate to conduct multiple analyses in a variety of circumstances, but the issue of multiplicity must always be addressed.
The purpose of interim analysis in adaptive trials is to determine how best to modify the remainder of the trial to increase the amount of useful information that can be gained from the remaining participants and hence the overall trial. This can include changing the endpoints of interest, eliminating certain treatment groups, changing the sample size, and modifying the statistical analysis plan. It should be recognized that the first time adaptive design is encountered, it can appear to run counter to fundamental tenets of the classical RCT employing a fixed design. However, while such RCTs remain extremely useful and powerful, as do group sequential designs, there is growing interest in the continued evolution of adaptive design methodology and analysis. However, the complexity of some of these designs currently poses challenges for both clinical scientists and regulators, and continued development of appropriate statistical methodologies is a high priority.

Coming Full Circle: New Designs for Tuberculosis Trials

Earlier in this paper the MRC’s trial of streptomycin for pulmonary tuberculosis, a ground-breaking RCT, was discussed. Sixty years later, tuberculosis remains a serious threat to global public health. As noted by Phillips et al., “The emergence and rapid spread of multidrug-resistant and extensively drug-resistant tuberculosis have given an increased priority to the development and evaluation of novel drug regimens that are expected to be more effective, less toxic, and increase adherence.”10 The words ‘drug regimens’ in this quote are important: pharmacotherapy for this disease requires combinations of drugs rather than a single drug. The traditional approach for testing combination therapy has been to conduct multiple Phase 2 RCTs for every potential new drug combination, with promising combinations progressing to Phase 3 trials. Given the expense of Phase 3 trials, there is a very large cost when a false-positive Phase 2 result is obtained, i.e., a drug combination that looked promising in Phase 2 progressing to Phase 3 but not being found to be effective in that trial. Additionally, investigational sites, patient populations from which to recruit participants for trials, and funds for conducting tuberculosis trials are all relatively restricted. A different approach is therefore needed.

The multi-arm multi-stage (MAMS) design offers considerable advantages in this context. Phillips et al. also provided an example of this.10 They discussed a three-stage trial that, at commencement, contains five treatment arms, a control drug regimen and four novel drug regimens. At commencement, therefore, participants are randomized to one of the five treatment arms. Two interim analyses are detailed in the study protocol. When the first interim analysis is conducted at the end of Stage 1, it is determined that novel regimen 4 does not display sufficient efficacy compared with the control regimen to warrant its continuation. That treatment arm is therefore terminated, and all participants subsequently entering the trial are randomized to one of the remaining treatment arms. The second interim analysis is conducted at the end of Stage 2. At this point, it is determined that neither novel regimen 1 nor novel regimen 3 display sufficient efficacy compared with the control regimen to warrant continuation. These two treatment arms are therefore terminated, which means that only the control regimen and novel regimen 2 continue until the end of the trial.

This design, which involves the dropping of poorly performing novel regimens, offers two benefits. First, drug regimens that were relatively ineffective do not progress to Phase 3 trials. This likely reduces the expense and time costs of false-positives. Second, participant safety is enhanced from the commencement of Stage 2 onwards, since all participants entering the trial once a treatment arm has been terminated for being relatively ineffective are not randomized to receive that treatment. As Phillips et al. noted, “In the next few years at least 4 new drug classes will need evaluation in combination with each other and in combination with the standard drugs to define the best possible treatment.” MAMS designs therefore offer significant advantages in this therapeutic area.
The New Regulatory Landscape for Risk-based Monitoring

Discussion to this point has largely focused on statistical (study design and analysis) and methodological (randomization) aspects of RCTs. However, operational execution is also critically important. One aspect of this execution is monitoring, and documents recently released by both the EMA and FDA are likely to have a large influence on how monitoring is conducted for future trials.

The EMA document, released in August 2011, is entitled “Reflection Paper on Risk Based Quality Management in Clinical Trials.”61 The paper expresses the view that avoidable quality problems arise in too many trials, and that the costs of monitoring them are very high. This observation suggests that the current approach to clinical quality management is in need of review and reorientation.62 It is important to develop better ways to ensure that limited resources are best targeted to address the most important issues and priorities, especially those associated with predictable or identifiable risks to the wellbeing of trial participants and to the quality of trial data. The paper notes that determination of the extent and nature of monitoring should be based on considerations such as the objective, purpose, design, complexity, blinding, size, and endpoints of the trial. Thus, the purpose of the EMA’s reflection paper is to facilitate the development of a more systematic, prioritized, risk-based approach to quality management of clinical trials, to support the principles of Good Clinical Practice, and to complement existing quality practices, requirements and standards.62

The guidance proposes an operational definition of the optimal combination of monitoring strategies for a given trial as the approach that maximizes participant protection, the quality and integrity of clinical trial data obtained from them, and compliance with all applicable regulations. The FDA is aware that preceding documents may not clearly reflect their current recommendations regarding monitoring practices, and that they recognize that they “must clearly articulate our recognition of the value of alternative approaches to facilitate change in industry’s monitoring practices.”63

For purposes of [the FDA’s risk-based] guidance, monitoring generally refers to the methods used by sponsors of investigational studies, or CROs delegated responsibilities for the conduct of such studies, to oversee the conduct of and reporting of data from clinical investigations, including appropriate investigator supervision of study site staff and third party contractors. The findings should be used to correct investigator and site practices that could result in inadequate human subject protection and/or poor data quality.

The FDA’s guidance is intended to clarify that risk-based monitoring, including the appropriate use of centralized monitoring and technological advances (e.g., e-mail, webcasts, and online training modules), can meet statutory and regulatory requirements under appropriate circumstances.
The guidance therefore describes “strategies for monitoring activities that reflect a modern, risk-based approach that focuses on critical study parameters and relies on a combination of monitoring activities to oversee a study effectively.” These strategies provide sponsors with the ability to provide the required oversight of participants' rights, welfare, and safety, and to ensure optimal quality and integrity of clinical trial data.

Planning for risk-based monitoring must start at the time a study protocol is being prepared. As the FDA guidance comments, “The most important tool for ensuring human subject protection and high-quality data is a well-designed and articulated protocol.” At that point, sponsors should perform a risk assessment that incorporates consideration of the types of data to be collected in the trial, the methodologies required to collect them, and “the range of potential safety and other human subject protection concerns that are inherent to the clinical investigation.” While the protocol must address all details of the study's design and execution in detail, it has become acknowledged that certain procedures and data are critical to the integrity of the trial's results, and that risk-based monitoring should therefore pay particular attention to them. These include:

- Data that are critical to the reliability of the study findings, specifically those data that support primary and secondary endpoints;
- Other data that are critical to subject safety, such as serious adverse events and events leading to discontinuation of treatment;
- Processes that underpin subject safety and ethical treatment, such as seeking appropriate medical consultation or scheduling extra visits in the event of specified clinical or laboratory findings;
- Processes that underpin the integrity of these data, such as blinding or referring specified events for adjudication.

Types of monitoring include on-site monitoring and centralized monitoring. On-site monitoring refers to in-person evaluations carried out by sponsor personnel or their representatives (typically from a contact research organization) at the site(s) at which a trial is being conducted.

Centralized monitoring refers to a remote evaluation carried out by sponsor personnel or representatives at a location other than the sites at which the clinical investigation is being conducted. Electronic data capture (EDC) systems are making it possible to implement centralized monitoring methods enabling decreased reliance on on-site monitoring: activities traditionally performed by on-site monitoring, e.g., source data verification, can often be accomplished remotely. Risk-based monitoring is a blend of both types of monitoring: the ability to combine on-site and centralized monitoring in the optimal balance facilitates risk-based monitoring, and a key facet of the guidance is the FDA's emphasis that such a strategy is acceptable from a regulatory perspective, and, moreover, likely better than on-site monitoring alone.

In addition to reducing the costs of monitoring, centralized monitoring offers unique advantages. It has become apparent that centralized statistical assessment provides a powerful means of identifying data trends not easily detected by on-site monitoring and also various data anomalies, including fabrication of data at a site that are ‘too good to be true’ in comparison with data from other sites. Additionally, it usefully identifies sites that require additional monitoring and/or training: that is, higher-risk sites receive more monitoring attention in person.
The guidance emphasizes that “no single approach to monitoring is appropriate or necessary for every clinical trial,” and encourages sponsors to tailor monitoring plans to all trials on a case-by-case basis, focusing on participant protection and data integrity risks in the specific context of a trial’s protocol. See Sax et al. for a more extensive discussion of risk-based monitoring.

Concluding Comments
This White Paper is published to celebrate the 50th Anniversary of the Kefauver-Harris Amendments, a landmark driving force in the development of the RCT. While its continuing evolution is of great importance, the statistical underpinnings of the RCT have withstood the test of time. All individuals interested and involved in integrated biopharmaceutical medicine, including everyone involved in lifecycle drug development and also health professionals who prescribe, dispense, and administer pharmaceutical medicines to patients, are well served by an understanding and appreciation of their history and importance.
References
1 Matthews JNS. *Introduction to randomized controlled clinical trials*. Edward Arnold.


J. Rick Turner, PhD
Senior Director, Clinical Communications

Dr. Turner joined Quintiles after serving as the chairman of the Department of Clinical Research at Campbell University School of Pharmacy. Prior to that he was a principal clinical submissions scientist at GlaxoSmithKline, where he received awards for his work on the GlaxoSmithKline Clinical Trial Registry and in new product development. He is also the President and Chief Scientific Officer, Turner Medical Communications LLC.

Dr. Turner received his PhD in Cardiovascular Behavioral Medicine from the University of Birmingham, UK. His program of experimental clinical research led to 50 peer-reviewed publications and the receipt of two international research awards, from the Society for Psychophysiological Research and the American Psychosomatic Society. He is very interested in the integration of behavioral medicine and biopharmaceutical medicine approaches for the continued enhancement of patient health and well-being.

Since entering the biopharmaceutical industry he has published extensively in peer-reviewed and professional journals, and authored six books addressing statistical and methodological aspects of randomized concurrently-controlled clinical trials. He was an invited participant in the 2010 National Heart, Lung, and Blood Institute Clinical Trials Symposium, giving a presentation entitled “The Power of the Randomized, Concurrently-controlled Clinical Trial.”

Dr. Turner is particularly interested in the development of drugs for type 2 diabetes mellitus, a global public health concern of staggering proportions. He has testified before two US Food and Drug Administration Advisory Committees, and is working with many stakeholders to expedite the development of drugs for this disease. He is also working with various stakeholders to increase adherence to drugs for all chronic diseases, including diabetes, by making greater use of knowledge and strategies from behavioral medicine.

Acknowledgments
1. Parts of this paper are included from the following Commentary article: Turner JR. The 50th Anniversary of the Kefauver-Harris Amendments: Efficacy Assessment and the Randomized Clinical Trial. Journal of Clinical Hypertension. doi: 10.1111/jch.12012.
2. The author wishes to thank Sarah Weitzel and Jim Chestnut, Quintiles’ Library & Information Services, for providing copies of many of the articles referenced in this White Paper.