



Published in final edited form as:

*Postgrad Med.* 2011 September ; 123(5): 194–204. doi:10.3810/pgm.2011.09.2475.

## Key Concepts of Clinical Trials: A Narrative Review

**Craig A. Umscheid, MD, MSCE<sup>1,4</sup>, David J. Margolis, MD, PhD, MSCE<sup>2,5</sup>, and Craig E. Grossman, MD, PhD<sup>2,6</sup>**

<sup>1</sup>Center for Evidence-Based Practice, University of Pennsylvania, Philadelphia, Pa

<sup>2</sup>Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA

<sup>3</sup>Department of Medicine, University of Pennsylvania, Philadelphia, PA

<sup>4</sup>Leonard Davis institute of Health Economics, University of Pennsylvania, Philadelphia, PA

<sup>5</sup>Department of Dermatology, University of Pennsylvania, Philadelphia, PA

<sup>6</sup>Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA

### Abstract

The recent focus of federal funding on comparative effectiveness research underscores the importance of clinical trials in the practice of evidence-based medicine and health care reform. The impact of clinical trials not only extends to the individual patient by establishing a broader selection of effective therapies, but also to society as a whole by enhancing the value of health care provided. However, clinical trials also have the potential to pose unknown risks to their participants, and biased knowledge extracted from flawed clinical trials may lead to the inadvertent harm of patients. Although conducting a well-designed clinical trial may appear straightforward, it is founded on rigorous methodology and oversight governed by key ethical principles. In this review, we provide an overview of the ethical foundations of trial design, trial oversight, and the process of obtaining approval of a therapeutic, from its pre-clinical phase to post-marketing surveillance. This narrative review is based on a course in clinical trials developed by one of the authors (DJM), and is supplemented by a PubMed search predating January 2011 using the keywords “randomized controlled trial,” “patient/clinical research,” “ethics,” “phase IV,” “data and safety monitoring board,” and “surrogate endpoint.” With an understanding of the key principles in designing and implementing clinical trials, health care providers can partner with the pharmaceutical industry and regulatory bodies to effectively compare medical therapies and thereby meet one of the essential goals of health care reform.

### Keywords

drug development; phase I trial; phase II trial; phase III trial; phase IV trial; randomized controlled trial

---

© Postgraduate Medicine

Correspondence: Craig E. Grossman, MD, PhD, Department of Radiation Oncology, Division of Oncology Research, University of Pennsylvania, Anatomy Chemistry Building, Room B13, 3620 Hamilton Walk, Philadelphia, PA 19104., Tel: 215-573-6403, Fax: 215-898-0090, craig.grossman@uphs.upenn.edu.

#### Conflict of Interest Statement

Craig A. Umscheid, MD, MSCE, David J. Margolis, MD, PhD, MSCE, and Craig E. Grossman, MD, PhD disclose no conflicts of interest.

## Introduction

The explosion in health care costs in the United States has recently spurred large federal investments in health care to identify the medical treatments of highest value. Specifically, \$1.1 billion has been appropriated by the American Recovery and Reinvestment Act of 2009 for “comparative effectiveness” research to evaluate “...clinical outcomes, effectiveness, and appropriateness of items, services, and procedures that are used to prevent, diagnose, or treat diseases, disorders, and other health conditions.”<sup>1</sup> Although numerous study designs can address these goals, clinical trials (and specifically randomized controlled trials [RCTs]) remain the benchmark for comparing disease interventions. However, the implementation of clinical trials involves a rigorous approach founded on scientific, statistical, ethical, and legal considerations. Thus, it is crucial for health care providers to understand the precepts on which well-performed clinical trials rest in order to maintain a partnership with patients and industry in pursuit of the safest, and most effective and efficient therapies. We present key concepts as well as the dilemmas encountered in the successful design and execution of a clinical trial.

## Materials and Methods

This narrative review is based on a course in clinical trials developed by one of the authors (DJM), and is supplemented by a PubMed search predating January 2011 using the keywords “randomized controlled trial,” “patient/clinical research,” “ethics,” “phase IV,” “data and safety monitoring board,” and “surrogate endpoint.”

## The Ethical Foundation of Clinical Trials

Despite the first reported modern clinical trial described in James Lind’s “A Treatise of the Scurvy” from 1753, it was not until the mid-20th century that ethical considerations in human research were addressed. In response to the criminal medical experimentation of human subjects by the Nazis during World War II, 10 basic principles of human research were formulated as the Nuremberg Code of 1949.<sup>2</sup> This code was later extended globally as The Declaration of Helsinki and adopted by the World Medical Association in 1964.<sup>3</sup> Notably, it advanced the ethical principle of “clinical equipoise,” a phrase later coined in 1987 to describe the expert medical community’s uncertainty regarding the comparative efficacy between treatments studied in a clinical trial.<sup>4</sup> This ethical precept guides the clinical investigator in executing comparative trials without violating the Hippocratic Oath.

Further advancement of the principles of respect for persons, beneficence (to act in the best interest of the patient), and justice emerged in the 1979 Belmont Report,<sup>5</sup> which was commissioned by the US government in reaction to the Tuskegee syphilis experiment.<sup>6</sup> This report applied these concepts to the processes of informed consent, assessment of risks and benefits, and equitable selection of subjects for research. Importantly, the boundaries between clinical practice and research were clarified, distinguishing activities between “physicians and their patients” from those of “investigators and their subjects.” Here, research was clearly defined as “an activity designed to test a hypothesis...to develop or contribute to generalizable knowledge.”<sup>5</sup>

It was the Belmont Report that finally explicated the principle of informed consent proposed 30 years prior in the Nuremberg Code. Informed consent, now a mandatory component of clinical trials that must be signed by all study participants (with few exceptions), must clearly state:<sup>7</sup>

1. This is a research study (including an explanation of the purpose and duration; and the risks, benefits, and alternatives of the intervention)

2. Participation is voluntary
3. The extent to which confidentiality will be maintained
4. Contact information for questions or concerns

Interestingly, this elemental safeguard in patient research is not without flaws. In reality, the investigator has limited information regarding the risks and benefits of an intervention because this is paradoxically the objective of performing the study. Challenges still remain with exercising informed consent, as illustrated by study participant comprehension deficiencies and self-reported dissatisfaction with the process.<sup>8</sup> This has prompted explorations to improve participant understanding of consent documents and procedures.<sup>9</sup>

In 1991, the ethical principles from these seminal works were culminated into Title 45, Part 46 of the Code of Federal Regulations, titled “Protection of Human Subjects.”<sup>7</sup> Referred to as the “Common Rule,” it regulates all federally supported or conducted research, with additional protections for prisoners, pregnant women, children, neonates, and fetuses.

## Overview of Trial Design

Clinical trials, in their purest form, are designed to observe outcomes of human subjects under “experimental” conditions controlled by the scientist. This is contrasted to noninterventional study designs (ie, cohort and case-control studies), in which the investigator measures but does not influence the exposure of interest. A clinical trial design is often favored because it permits randomization of the intervention, thereby effectively removing the selection bias that results from the imbalance of unknown/immeasurable confounders. Within this inherent strength is the capacity to unveil causality in an RCT. Randomized clinical trials, however, still remain subject to limitations such as misclassification or information bias of the outcome or exposure, co-interventions (where one arm receives an additional intervention more frequently than another), and contamination (where a proportion of subjects assigned to the control arm receive the intervention outside of the study).

Execution of a robust clinical trial requires the selection of an appropriate study population. Despite all participants voluntarily consenting for the intervention, the enrolled cohort may potentially differ from the general population from which they were drawn. This type of selection bias, called “volunteer bias,” may arise from such factors as study eligibility criteria, inherent subject attributes (eg, geographic distance from the study site, health status, attitudes and beliefs, education, and socioeconomic status), or subjective exclusion by the investigator because of poor anticipated enrollee compliance or overall prognosis.<sup>10</sup> Although RCTs seek to achieve internal validity by enrolling a relatively homogeneous population according to predefined characteristics, narrow inclusion and exclusion criteria may limit their external validity (or “generalizability”) to a broader population of patients with highly prevalent comorbidities that may not be included in the sample cohort. This theme underscores why an experimental treatment’s “efficacy” (ie, a measure of the success of an intervention in an artificial setting) may not translate into its “effectiveness” (ie, a measure of its value applied in the “real world”). Attempts to improve patient recruitment and generalizability using free medical care, financial payments,<sup>11</sup> and improved communication techniques<sup>12</sup> are considered ethical as long as the incentives are not unduly coercive.<sup>13</sup>

In order to assess the efficacy of an intervention within the context of a clinical trial, there must be deliberate control of all known confounding variables (including comorbidities), thereby requiring a homogeneous group of participants. However, the evidence provided by a well-designed and executed clinical trial will have no value if it cannot be applied to the

general population. Thus, designers of clinical trials must use subjective judgment (including clinical, epidemiological, and biostatistical reasoning) to determine at the outset how much trade-off they are willing to make between the internal validity and generalizability of a clinical trial.

A “surrogate endpoint” is often chosen in place of a primary endpoint to enhance study efficiency (ie, less cost and time, improved measurability, and smaller sample size requirement). Ideally, the surrogate should completely capture the effect of the intervention on the clinical endpoint, as formally proposed by Prentice.<sup>14</sup> Blood pressure is a well established surrogate for cardiovascular-related mortality because its normalization has been associated with clinically beneficial outcomes, such as fewer strokes, and less renal and cardiac complications.<sup>15</sup> However, one must use caution when relying on surrogates, as they may be erroneously implicated in the direct causal pathway between intervention and true outcome.<sup>16,17</sup> A frequently described, clinically logical, but flawed use of a surrogate endpoint was premature ventricular contractions (PVCs) to assess whether antiarrhythmic drugs reduced the incidence of sudden death after a myocardial infarction in the Cardiac Arrhythmia Suppression Trial (CAST). Despite evidence of the association between PVCs and early arrhythmic mortality, pharmacologic suppression of PVCs unexpectedly increased the very event (mortality) that it was supposed to remedy.<sup>18</sup> As surrogates are commonly employed in phase I–II trials, it is highly likely that a high proportion of clinically effective therapeutics are discarded because of false-negative results using such endpoints. This is exemplified in the trial by the International Chronic Granulomatous Disease (CGD) study group, in which the surrogate markers of superoxide production and bactericidal efficiency were initially applied to assess the efficacy of interferon- $\gamma$  for treatment of CGD.<sup>19</sup> For reasons outside the scope of this review, the authors decided *a priori* to extend the study duration in order to adequately detect the clinical endpoint of interest (recurrent serious infections) instead of the originally proposed surrogate markers (superoxide production and bactericidal efficiency). Treatment with interferon- $\gamma$  was incredibly successful, as the rate of recurrent serious infections was highly reduced. However, there was no observable effect on superoxide production and bactericidal activity. Had the primary endpoint not been changed, the originally proposed surrogate biomarkers would have masked the clinically relevant efficacy of this treatment. These examples illustrate the importance of validating surrogates as reliable predictors of clinical endpoints using meta-analyses and/or natural history studies of large population cohorts, in conjunction with ensuring biological plausibility.<sup>20</sup>

For a trial to adequately address the “primary question(s)” of interest, a sufficient sample size is required to have enough power to detect a potential statistical difference. Traditionally, power is defined as having at least an 80% chance of finding a statistically significant difference between the outcomes of 2 interventions when a clinically meaningful difference exists. The outcomes or endpoints of the investigation, whether objective (eg, death) or subjective (eg, quality of life), must always be reliable and meaningful measures. Statistical analyses commonly used to analyze outcomes include logistic regression for dichotomous endpoints (eg, event occurred/did not occur), Poisson regression for rates (eg, number of events per person-years), Cox regression for time-to-events (eg, survival analysis), and linear regression for continuous measures (eg, weight).

## Overview of Drug Development

The general road to drug development and approval has been defined and regulated by the US Food and Drug Administration (FDA) for decades. Safety has historically been its primary focus, followed by efficacy. If a drug appears promising in pre-clinical studies, a drug sponsor or sponsor-investigator can submit an investigational new drug (IND) application. This detailed proposal contains investigator qualifications and all pre-clinical

drug information and data, and a request for exemption from the federal statutes that prohibit interstate transport of unapproved drugs. After approval, the drug is studied (phase I–III trials, described below) and if demonstrated safe and efficacious in the intended population, the drug sponsor can then submit a New Drug Application (NDA) to the FDA. After an extensive review by the FDA that often involves a recommendation by an external committee, the FDA determines whether the therapeutic can be granted an indication and marketed. After final approval, the drug can continue to be studied in phase IV trials, in which safety and effectiveness for the indicated population is monitored. To facilitate evaluation and endorsement of foreign drug data, efforts have been made to harmonize this approval process across the United States, Europe, and Japan through the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).<sup>21</sup>

## Pre-Clinical, Phase I, and Phase II Trials

Pre-clinical investigations include animal studies and evaluations of drug production and purity. Animal studies explore: 1) the drug's safety in doses equivalent to approximated human exposures, 2) pharmacodynamics (ie, mechanisms of action, and the relationship between drug levels and clinical response), and 3) pharmacokinetics (ie, drug absorption, distribution, metabolism, excretion, and potential drug–drug interactions). This data must be submitted for IND approval if the drug is to be further studied in human subjects.

Because the FDA emphasizes “safety first,” it is logical that the first of 4 stages (known as “phases”) of a clinical trial is designed to test the safety and maximum tolerated dose (MTD) of a drug, human pharmacokinetics and pharmacodynamics, and drug–drug interactions. These phase I trials (synonymous with “dose-escalation” or “human pharmacology” studies) are the first instance in which the new investigational agent is studied in humans, and are usually performed open label and in a small number of “healthy” and/or “diseased” volunteers. The MTD, or the drug dose before a dose-limiting toxicity, can be determined using various statistical designs. Dose escalation is based on very strict criteria, and subjects are closely followed for evidence of drug toxicity over a sufficient period. There is a risk that subjects who volunteer (or the actual physicians who enroll patients) for phase I studies will misinterpret its objective as therapeutic. For example, despite strong evidence that objective response rates in phase I trials of chemotherapeutic drugs is exceedingly low (as low as 2.5%),<sup>22</sup> patients may still have a “therapeutic misconception” of potentially receiving a direct medical benefit from trial participation.<sup>23</sup> Improvements to the process of informed consent<sup>9</sup> could help dispel some of these misconceptions while still maintaining adequate enrollment numbers.

Phase II trials, also referred to as “therapeutic exploratory” trials, are usually larger than phase I studies, and are conducted in a small number of volunteers who have the disease of interest. They are designed to test safety, pharmacokinetics, and pharmacodynamics, but may also be designed to answer questions essential to the planning of phase III trials, including determination of optimal doses, dose frequencies, administration routes, and endpoints. In addition, they may offer preliminary evidence of drug efficacy by: 1) comparing the study drug with “historical controls” from published case series or trials that established the efficacy of standard therapies, 2) examining different dosing arms within the trial, or 3) randomizing subjects to different arms (such as a control arm). However, the small number of participants and primary safety concerns within a phase II trial usually limit its power to establish efficacy, and thereby supports the necessity of a subsequent phase III trial.

At the conclusion of the initial trial phases, a meeting between the sponsor(s), investigator(s), and FDA may occur to review the preliminary data, IND, and ascertain the viability of progressing further to a phase III trial (including plans for trial design, size, outcomes, safety concerns, analyses, data collection, and case report forms). Manufacturing concerns may also be discussed at this time.

## Phase III Trials

Based on prior studies demonstrating drug safety and potential efficacy, a phase III trial (also referred to as a “therapeutic confirmatory,” “comparative efficacy,” or “pivotal trial”) may be pursued. This stage of drug assessment is conducted in a larger and often more diverse target population in order to demonstrate and/or confirm efficacy and to identify and estimate the incidence of common adverse reactions. However, given that phase III trials are usually no larger than 300 to 3000 subjects, they consequently have the statistical power to establish an adverse event rate of no less than 1 in 100 persons (based on Hanley’s “Rule of 3”).<sup>24</sup> This highlights the significance of phase IV trials in identifying less-common adverse drug reactions, and is one reason why the FDA usually requires more than one phase III trial to establish drug safety and efficacy.

The most common type of phase III trials, comparative efficacy trials (often referred to as “superiority” or “placebo-controlled trials”), compare the intervention of interest with either a standard therapy or a placebo. Even in the best-designed placebo-controlled studies, it is not uncommon to demonstrate a placebo effect, in which subjects exposed to the inert substance exhibit an unexpected improvement in outcomes when compared with historical controls. While some attribute the placebo effect to a general improvement in care imparted to subjects in a trial, others argue that those who volunteer for a study are acutely symptomatic and will naturally improve or “regress to the mean” as the trial progresses. This further highlights the uniqueness of study participants and why a trial may lack external validity. The application of placebos, including surgical placebos (“sham procedures”),<sup>25-27</sup> has ignited some debate; the revised Declaration of Helsinki supports comparative efficacy trials by discouraging the use of drug placebos in favor of “best current” treatment controls.<sup>28,29</sup>

Another type of phase III trial, the equivalency trial (or “positive-control study”), is designed to ascertain whether the experimental treatment is similar to the chosen comparator within some margin prespecified by the investigator. Hence, a placebo is almost never included in this study design. As long as the differences between the intervention and the comparator remain within the prespecified margin, the intervention will be deemed equivalent to the comparator.<sup>30</sup> Although the prespecified margin is often based on external evidence, statistical foundations, and clinical experience, there remains little guidance for setting acceptable margins. A variant of the equivalency trial, the noninferiority study, is conducted with the goal of excluding the possibility that the experimental intervention is less effective than the standard treatment by some prespecified magnitude. One must be cautious when interpreting the results of all types of equivalency trials because they are often incorrectly designed and analyzed as if they were comparative efficacy studies. Such flaws can result in a bias towards the null, which would translate into a false-negative result in a comparative efficacy study, but a false-positive result in an equivalency trial. Of note, the noninferiority trial is more susceptible to false-positive results than other study designs.<sup>31</sup>

A hallmark of the phase III trial design is the balance in treatment allocation for comparison of treatment efficacy. Implemented through randomization, this modern clinical trial practice attempts to eliminate imbalance of confounders and/or any systematic differences (or biases) between treatment groups. The statistical tool of randomization, first introduced



into clinical trials by Sir Austin Bradford Hill,<sup>32</sup> was born out of the necessity (and ethical justification) of rationing limited supplies of streptomycin in a British trial of pulmonary tuberculosis.<sup>33</sup> The most basic randomization model, simple randomization, randomly allocates each subject to a trial arm regardless of those already assigned (ie, a “coin flip” for each subject). Although easy to perform, major imbalances in treatment assignments or distribution of covariates can ensue, making this strategy less than ideal. To improve on this method, a constraint can be placed on randomization that forces the number of subjects randomly assigned per arm to be equal and balanced after a specified block size (“block randomization”). For example, in a trial with 2 arms, a block size of 4 subjects would be designated as 2 positions in arm A and 2 positions in arm B. Even though the positions would be randomly assigned within the block of 4 subjects, it would be guaranteed that, after randomization of 4 subjects, 2 subjects would be in arm A and 2 subjects would be in arm B (Table 1). The main drawback of applying a fixed-block allocation is that small block sizes can allow investigators to predict the treatment of the next patient, resulting in “unblinding.” For example, if a trial has a block size of 2, and the first subject in the block was randomized to treatment “A,” then the investigator will know that the next subject will be randomized to “the other” treatment. Variable block sizes can help prevent this unblinding (eg, a block size of 4 followed by a block size of 8 followed by a block size of 6).

Another feature of phase III trial design is stratification, which is commonly employed in combination with randomization to further balance study arms based on prespecified characteristics (rather than size in the case of blocking). Stratification facilitates analysis by ensuring that specific prognostic factors of presumed clinical importance are properly balanced in the arms of a clinical trial. Stratification of a relatively small sample size that has also undergone block randomization may result in loss of the originally intended balance, thereby supporting the merits of alternative techniques such as minimization or dynamic allocation, designed to reduce imbalances among multiple strata and study arms.<sup>34</sup>

Often, the phase III trial design dictates that the interventions be “blinded” (or masked) in an effort to minimize assessment bias of subjective outcomes. Specific blinding strategies to curtail this “information bias” include “single blinding” (subject only), “double blinding” (both subject and investigator), or “triple blinding” (data analyst, subject, and investigator). Unfortunately, not all trials can be blinded (eg, method of drug delivery cannot be blinded), and the development of established drug toxicities may lead to inadvertent unmasking and raise ethical and safety issues. When appropriate, additional strategies can be applied to enhance study efficiency, such as assigning each subject to serve as his/her own control (crossover study) or evaluating more than one treatment simultaneously (factorial design).

The most common approach to analyzing phase III trials is the intention-to-treat analysis, in which subjects are assessed based on the intervention arm to which they were randomized, regardless of what treatment they actually received. This is commonly known as the “analyzed as randomized” rule. A complementary or secondary analysis is an “as-treated” or “per-protocol” analysis, in which subjects are evaluated based on the treatment they actually received, regardless of whether they were randomized to that treatment arm. Intention-to-treat analyses are preferable for the primary analysis of RCTs,<sup>35</sup> as they eliminate selection bias by preserving randomization; any difference in outcomes can therefore be attributed to the treatment alone and not confounders. In contrast, an “as-treated” or “per-protocol” approach may eliminate any benefit of random treatment selection in an interventional trial, as it estimates the effect of treatment received. The study thereby becomes similar to an interventional cohort study with the potential for treatment selection bias. If adherence in the treatment arm is poor and contamination in the control group is high, an intention-to-treat

analysis may fail to show a difference in outcomes. This is in contrast to a per-protocol analysis that takes into account these protocol violations.

Based on the vast combination of strategies applicable to the design of a phase III study, the Consolidated Standards of Reporting Trials (CONSORT) guideline was established to improve the quality of trial reporting and assist with evaluating the conduct and validity of trials and their results.<sup>36</sup> Employing a flow diagram (Figure 1) and a 22-item checklist (Table 2),<sup>37</sup> readers can easily identify the stages in which subjects withdraw from a study (eg, found to be ineligible, lost to follow-up, cannot be evaluated for the primary endpoint). Because exclusion of such missing data can reduce study power and lead to bias, the best way to avoid these challenges is to adhere to the CONSORT checklist, thereby enrolling only eligible patients and ensuring that they remain on-study.

## Phase IV Trials

Once a drug is approved, the FDA may require that a sponsor conduct a phase IV trial as a stipulation for drug approval, although the literature suggests that less than half of such studies are actually completed or even initiated by sponsors.<sup>38</sup> Phase IV trials, also referred to as “therapeutic use” or “post-marketing” studies, are observational studies performed on FDA-approved drugs to: 1) identify less common adverse reactions, and 2) evaluate cost and/or drug effectiveness in diseases, populations, or doses similar to or markedly different from the original study population. Limitations of pre-marketing (eg, phase III) studies become apparent with the statistic that roughly 20% of drugs acquire new black box warnings post-marketing, and approximately 4% of drugs are ultimately withdrawn for safety reasons.<sup>39,40</sup> As described by one pharmacoepidemiologist, “this reflects a deliberate societal decision to balance delays in access to new drugs with delays in information about rare adverse reactions.”<sup>41</sup>

Over the past decade, there has been a steady rise in voluntarily and spontaneously reported serious adverse drug reactions submitted to the FDA’s MedWatch program, from 150 000 in 2000 to 370 000 in 2009.<sup>42</sup> Reports are submitted directly by physicians and consumers, or indirectly via drug manufacturers (the most common route). Weaknesses of this post-marketing surveillance are illustrated by recent failures to quickly detect serious cardiovascular events resulting from the use of the anti-inflammatory medication Vioxx® and prescription diet drug Meridia®. It was only after the European SCOUT (Sibutramine Cardiovascular OUTcome Trial) study, driven by anecdotal case reports concerning cardiovascular safety, that the FDA withdrew Meridia® from the market in late 2010.<sup>43</sup> The most common criticisms of the FDA’s post-marketing surveillance are: 1) the reliance on voluntary reporting of adverse events, resulting in difficulty calculating adverse event rates because of incomplete data on total events and unreliable information on the true extent of exposures; 2) the trust in drug manufacturers to collect, evaluate, and report drug safety data that may risk their financial interests; and 3) the dependence on one government body to approve a drug and then actively seek evidence that might lead to its withdrawal.<sup>38,41</sup> Proposed solutions include the establishment of a national health data network to oversee post-marketing surveillance independent of the FDA-approval process,<sup>44</sup> preplanned meta-analyses of a series of related trials to assess less-common adverse events,<sup>45</sup> and large-scale simple RCTs with few eligibility and treatment criteria (ie, Peto studies).<sup>46</sup>

## Clinical Trial Oversight

Historic abuses and modern day tragedies highlight the importance of Institutional Review Boards (IRBs) and Data and Safety Monitoring Boards (DSMBs) in ensuring that human research conforms to local and national standards of safety and ethics.<sup>47,48</sup> Under the Department of Health and Human Services Title 45 Part 46 of the Code of Federal



Regulations (CFR), IRBs are charged with protecting the rights and welfare of human subjects involved in research conducted or supported by any federal department or agency.<sup>7</sup> In order to ensure compliance with the strict and detailed guidelines of the CFR, members of IRBs (one of whom must be a non-scientist, and one of whom must be independent of the board's home institution) are authorized under the "Common Rule" to approve, require modification to, or reject a research activity. Based on the perceived risk of the study, IRBs have a number of levels of review from exempt for "minimal risk" studies (defined by the "Common Rule" as risks that are no greater than those encountered in daily life or routine clinical examinations or tests) to the more lengthy and involved full board reviews for higher-risk studies. General criteria for IRB approval include: 1) risks to subjects are minimized, and are reasonable in relation to benefits; 2) selection of subjects is equitable; 3) informed consent is sought; 4) sufficient provisions for data monitoring exist to maintain subjects' safety; 5) adequate mechanisms are in place to ensure subject confidentiality; and 6) rights and welfare of vulnerable populations are protected.<sup>7</sup>

Data and Safety Monitoring Boards, also referred to as "data safety committees" or "data monitoring committees," are often required by IRBs for study approval, and are charged with: 1) safeguarding the interests of study subjects; 2) preserving the integrity and credibility of the trial in order that future patients may be treated optimally; and 3) ensuring that definitive and reliable trial results be made available in a timely fashion to the medical community.<sup>49</sup> Specific responsibilities include monitoring data quality, study conduct (including recruitment rates, retention rates, and treatment compliance), drug safety, and drug efficacy. Data and Safety Monitoring Boards are usually organized by the trial sponsor and principal investigator, and are often comprised of biostatisticians, ethicists, and physicians from relevant specialties, among others. Outcomes from DSMB activities include: 1) extension of recruitment strategies if the study is not meeting enrollment goals; 2) changes in study entry criteria, procedures, treatments or study design; and 3) early closure of the study because of safety issues (external or internal), slow recruitment rates, poor compliance with the study protocol, or clinically significant differences in drug efficacy or toxicity between trial arms. To highlight the important charge of DSMBs and their relevance outside of the scientific community, there have been egregious breaches of confidentiality by DSMB members who have leaked confidential drug information to Wall Street firms for self-profit.<sup>48</sup> Hence, members of DSMBs ideally should be free of significant conflicts of interest, and should be the only individuals to whom the data analysis center provides real-time results of treatment efficacy and safety.

The complexity and expense of monitoring human research has prompted the establishment of Contract Research Organizations (CROs) to oversee clinical trials. They are commonly commercial or academic organizations hired by the study sponsor "to perform one or more of a sponsor's trial-related duties and functions," such as organizing and managing a DSMB, or managing and auditing trial data to maintain data quality.<sup>50</sup>

## Conclusion

To offer patients the most effective and safest therapies possible, it is important to understand the key concepts involved in performing clinical trials. The attention by the mass media to safety-based drug withdrawal (amounting to approximately 1.5 drugs per year since 1993<sup>51</sup>) emphasizes this point. Understanding the ethical precepts and regulations behind trial designs may also help key stakeholders respond to future research dilemmas at home and abroad. Moreover, well-designed and executed clinical trials can contribute significantly to the national effort to improve the effectiveness and efficiency of health care in the United States. Through rigorous practices applied to novel drug development and approval, physicians and patients can maintain confidence in the therapies prescribed.

## Take-Home Points

1. To ensure the safety of subjects who volunteer for clinical trials as well as preserving the integrity and credibility of the data reported, numerous regulatory boards including IRBs and DSMBs under the auspices of the federal government are involved with all studies conducted in the United States.
2. The rigorous methodology of executing a clinical trial, most significantly through the controlled and random intervention of human volunteers by the investigator, makes this epidemiologic study design one of the most powerful approaches to demonstrating causal associations in the practice of evidence-based medicine.
3. The internal validity that results from the narrowly selective enrollment criteria and artificial setting within a clinical trial must be balanced with the intent of translating the study findings to the “real world” in clinical practice (known as generalizability or external validity).
4. Enrollment and treatment allocation techniques, selected endpoints, methods of comparison, and statistical analyses must be carefully chosen in order to plausibly achieve the intended goals of the study.
5. Modern clinical trials are founded on numerous and continually evolving ethical principles and practices that guide the investigator in performing human research without violation of the Hippocratic Oath.
6. Emphasizing safety first, the most common route of studying a new therapeutic is from the establishment of the maximum tolerated dose in humans (phase I), to pharmacodynamic and pharmacokinetic studies, and exploration of therapeutic benefit (phase II), followed by comparing its efficacy to an established therapeutic or control in a larger population of volunteers (phase III), and ultimately post-market evaluation of adverse reactions and effectiveness when administered to the general population (phase IV).

## Acknowledgments

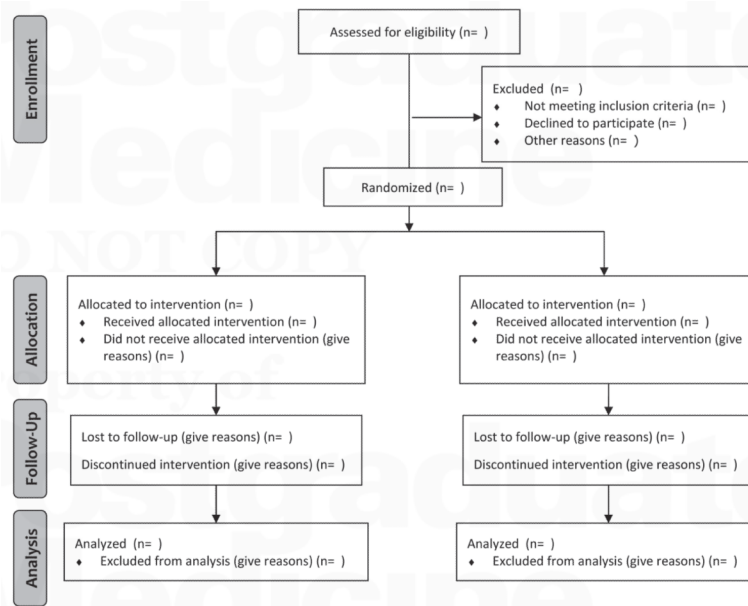
We gratefully acknowledge Rosemarie Mick, MS, for reviewing the statistical content of this manuscript. This work was supported in part by the National Institutes of Health (T32-HP-010026 [CAU] and T32-CA-009677 [CEG]). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

## References

1. R., H. American Recovery and Reinvestment Act of 2009. 1st ed. 2009.
2. Trials of War Criminals Before the Nuremberg Military Tribunals Under Control Council Law. Vol. 2. US Government Printing Office; Washington, DC: 1949. p. 181-182.
3. World Medical Organization. Declaration of Helsinki. December 7, 1996. *BMJ*. 1996;1448–1449. [PubMed: 8664621]
4. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med*. 1987; 317(3):141–145. [PubMed: 3600702]
5. Department of Health and Welfare. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. OPRR Reports; Washington, DC: 1979.
6. Brandt AM. Racism and research: the case of the Tuskegee Syphilis Study. *Hastings Cent Rep*. 1978; 8(6):21–29. [PubMed: 721302]
7. Services DoHaH. , editor. Code of Federal Regulations–The Common Rule: Protection of Human Subjects. Vol. 45. p. 2009

8. Lansimies-Antikainen H, Laitinen T, Rauramaa R, Pietilä AM. Evaluation of informed consent in health research: a questionnaire survey. *Scand J Caring Sci.* 2010; 24(1):56–64. [PubMed: 19552792]
9. Ridpath JR, Wiese CJ, Greene SM. Looking at research consent forms through a participant-centered lens: the PRISM readability toolkit. *Am J Health Promot.* 2009; 23(6):371–375. [PubMed: 19601476]
10. Gravetter, FJ.; Forzano, LAB. *Research Methods for the Behavioral Sciences.* 3rd ed. Wadsworth Cengage Learning; Belmont, CA: 2009.
11. Mapstone J, Elbourne D, Roberts I. Strategies to improve recruitment to research studies. *Cochrane Database Syst Rev.* 2007; (2) MR000013.
12. Treweek S, Mitchell E, Pitkethly M, et al. Strategies to improve recruitment to randomised controlled trials. *Cochrane Database Syst Rev.* 2010; (1) MR000013.
13. Dickert N, Grady C. What's the price of a research subject? Approaches to payment for research participation. *N Engl J Med.* 1999; 341(3):198–203. [PubMed: 10403861]
14. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.* 1989; 8(4):431–440. [PubMed: 2727467]
15. Chobanian, AV.; Bakris, GL.; Black, HR., et al. Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension; Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; 2003; p. 1206-1252.
16. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996; 125(7):605–613. [PubMed: 8815760]
17. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA.* 1999; 282(8):790–795. [PubMed: 10463719]
18. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991; 324(12):781–788. [PubMed: 1900101]
19. The International Chronic Granulomatous Disease Cooperative Study Group. A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. *N Engl J Med.* 1991; 324(8):509–516. [PubMed: 1846940]
20. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points—the challenge of statistical validation. *Nat Rev Clin Oncol.* 2010; 7(6):309–317. [PubMed: 20368727]
21. Good clinical practice guidelines for essential documents for the conduct of a clinical trial; International Conference on Harmonisation; Geneva, Switzerland: ICH Secretariat c/o IFPMA. 1994;
22. Roberts TG Jr, Goulart BH, Squitieri L, et al. Trends in the risks and benefits to patients with cancer participating in phase I clinical trials. *JAMA.* 2004; 292(17):2130–2140. [PubMed: 15523074]
23. Joffe S, Cook EF, Cleary PD, Clark JW, Weeks JC. Quality of informed consent in cancer clinical trials: a cross-sectional survey. *Lancet.* 2001; 358(9295):1772–1777. [PubMed: 11734235]
24. Eypasch E, Lefering R, Kum CK, Troidl H. Probability of adverse events that have not yet occurred: a statistical reminder. *BMJ.* 1995; 311(7005):619–620. [PubMed: 7663258]
25. Cahana A, Romagnoli S. Not all placebos are the same: a debate on the ethics of placebo use in clinical trials versus clinical practice. *J Anesth.* 2007; 21(1):102–105. [PubMed: 17285427]
26. Foddy B. A duty to deceive: placebos in clinical practice. *Am J Bioeth.* 2009; 9(12):4–12. [PubMed: 20013484]
27. Wilcox CM. Exploring the use of the sham design for interventional trials: implications for endoscopic research. *Gastrointest Endosc.* 2008; 67(1):123–127. [PubMed: 18155434]
28. Levine RJ. The need to revise the Declaration of Helsinki. *N Engl J Med.* 1999; 341(7):531–534. [PubMed: 10441613]
29. Vastag B. Helsinki discord? A controversial declaration. *JAMA.* 2000; 284(23):2983–2985. [PubMed: 11122566]

30. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. *J Gen Intern Med*. 2011; 26(2):192–196. [PubMed: 20857339]
31. Fleming TR. Current issues in non-inferiority trials. *Stat Med*. 2008; 27(3):317–332. [PubMed: 17340597]
32. Doll R. Sir Austin Bradford Hill and the progress of medical science. *BMJ*. 1992; 305(6868): 1521–1526. [PubMed: 1286370]
33. Hill AB. Medical ethics and controlled trials. *Br Med J*. 1963; 1(5337):1043–1049. [PubMed: 13954493]
34. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. *Control Clin Trials*. 2002; 23(6):662–674. [PubMed: 12505244]
35. Tsiatis A. Methodological issues in AIDS clinical trials. Intent-to-treat analysis. *J Acquir Immune Defic Syndr*. 1990; 3(suppl 2):S120–S123. [PubMed: 2231292]
36. Moher D, Schulz KF, Altman D, CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001; 285(15):1987–1991. [PubMed: 11308435]
37. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; 340:c332. [PubMed: 20332509]
38. Fontanarosa PB, Rennie D, DeAngelis CD. Postmarketing surveillance—lack of vigilance, lack of trust. *JAMA*. 2004; 292(21):2647–2650. [PubMed: 15572723]
39. Bakke OM, Manocchia M, de Abajo F, Kaitin KI, Lasagna L. Drug safety discontinuations in the United Kingdom, the United States, and Spain from 1974 through 1993: a regulatory perspective. *Clin Pharmacol Ther*. 1995; 58(1):108–117. [PubMed: 7628177]
40. Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DU, Wolfe SM, Bor DH. Timing of new black box warnings and withdrawals for prescription medications. *JAMA*. 2002; 287(17):2215–2220. [PubMed: 11980521]
41. Strom BL. Potential for conflict of interest in the evaluation of suspected adverse drug reactions: a counterpoint. *JAMA*. 2004; 292(21):2643–2646. [PubMed: 15572722]
42. Administration USFaD. , editor. [Accessed June 29, 2011] AERS patient outcomes by year. 2010. <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070461.htm>. Updated March 31, 2011
43. James WP, Caterson ID, Coutinho W, et al. Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects. *N Engl J Med*. 2010; 363(10):905–917. [PubMed: 20818901]
44. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med*. 2009; 151(5):341–344. [PubMed: 19638403]
45. Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. *JAMA*. 1999; 281(9):830–834. [PubMed: 10071005]
46. Hennekens CH, Demets D. The need for large-scale randomized evidence without undue emphasis on small trials, meta-analyses, or subgroup analyses. *JAMA*. 2009; 302(21):2361–2362. [PubMed: 19952322]
47. Mello MM, Studdert DM, Brennan TA. The rise of litigation in human subjects research. *Ann Intern Med*. 2003; 139(1):40–45. [PubMed: 12834317]
48. Steinbrook R. Protecting research subjects—the crisis at Johns Hopkins. *N Engl J Med*. 2002; 346(9):716–720. [PubMed: 11870258]
49. Ellenberg, S.; Fleming, TR.; DeMets, DL. *Data Monitoring Committees in Clinical Trials: A Practical Perspective*: West Sussex. John Wiley & Sons; England: 2002.
50. Guidance for Industry. E6. Good clinical practice—consolidated guidance. US Department of Health and Human Services; Bethesda, MD: 1996.
51. Issa AM, Phillips KA, Van Bebber S, et al. Drug withdrawals in the United States: a systematic review of the evidence and analysis of trends. *Curr Drug Saf*. 2007; 2(3):177–185. [PubMed: 18690965]



**Figure 1.**  
 CONSORT 2010 flow diagram.  
 Reproduced with permission from *BMJ*.<sup>37</sup>

**Table 1**

Permutations of a 4-Block Randomization Scheme in a 2-Arm Study with 24 Subjects

Within-Block Assignment Group	Schedule					
	1	2	3	4	5	6
1	A	A	B	B	A	B
2	B	B	A	A	A	B
3	A	B	B	A	B	A
4	B	A	A	B	B	A



**Table 2**

**CONSORT 2010 Checklist of Information to Include When Reporting a Randomized Trial**

Section/Topic	Item No	Checklist Item	Reported on Page No
<b>Title and abstract</b>			
	1a	Identification as a randomized trial in the title	---
	1b	structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	---
<b>Introduction</b>			
Background and objectives	2a	scientific background and explanation of rationale	---
	2b	specific objectives or hypotheses	---
<b>Methods</b>			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	---
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	---
Participants	4a	Eligibility criteria for participants	---
	4b	Settings and locations where the data were collected	---
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	---
Outcomes	6a	Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed	---
	6b	Any changes to trial outcomes after the trial commenced, with reasons	---
Sample size	7a	How sample size was determined	---
	7b	When applicable, explanation of any interim analyses and stopping guidelines	---
<b>Randomization</b>			
Sequence generation	8a	Method used to generate the random allocation sequence	---
	8b	Type of randomization; details of any restriction (such as blocking and block size)	---
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	---
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	---
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	---
	11b	If relevant, description of the similarity of interventions	---
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	---
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	---

Section/Topic	Item No	Checklist Item	Reported on Page No
<b>Results</b>			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome	---
	13b	For each group, losses and exclusions after randomization, together with reasons	---
Recruitment	14a	Dates defining the periods of recruitment and follow-up	---
	14b	Why the trial ended or was stopped	---
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	---
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	---
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	---
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	---
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	---
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	---
<b>Discussion</b>			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	---
Generalizability	21	Generalizability (external validity, applicability) of the trial findings	---
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	---
<b>Other information</b>			
Registration	23	Registration number and name of trial registry	---
Protocol	24	Where the full trial protocol can be accessed, if available	---
Funding	25	sources of funding and other support (such as supply of drugs), role of funders	---

Reproduced with permission from *BMJ*.<sup>37</sup>