

Appendix A

Clinical Trials: Overview and Terminology

Prior to the adoption of a new treatment for use in a population, it is important to assess the impact that the use of the treatment will have on the general health of the population. That is, one wants to know how the general health of the population after adoption of the treatment compares with what it would have been if the treatment had not been adopted. In practice, this can never be known exactly (since it is a counterfactual). But the governmental agencies that regulate approval of new treatments are charged with judging the treatment's impact to the extent possible. This appendix presents an overview of the purposes and various aspects of clinical trials and definitions of some of the key terms used in our study.

BASIC TERMS

An *effective* treatment is one that provides improvement in the general health of the population viewed as a whole. An *efficacious* treatment is one that in some identifiable subpopulation results in an outcome judged more beneficial than that which would exist without treatment. An efficacious treatment may not be effective owing either to its inability to be administered safely in a broad population or to its effect on other aspects of patient treatment and behaviors beyond the outcome used to evaluate efficacy.

It is also useful to differentiate among the concepts of a simple *treatment*, which would usually consist of a prescribed dose of given frequency and duration; a *treatment regimen*, which would usually involve rules for dose escalation or reduction in order to obtain greater effect while avoiding intolerable adverse experiences; and a *treatment strategy*, which would include plans for ancillary treatments and progression to other treatments in the face of disease progression.

In a phase III confirmatory study (see below), the ideal is typically an effectiveness study of a treatment strategy: effectiveness because it is the impact of a treatment on the population and a treatment strategy because the initial prescribed treatment may greatly affect the concomitant treatments and follow-on treatments administered to patients. However, true effectiveness can never be tested in an unbiased fashion because the trial setting itself is artificial and because observational studies are always subject to unmeasured bias. Phase III studies should be much closer to an effectiveness study than would be the phase II studies that might use surrogate biomarkers as a primary outcome in a subpopulation of the patients that might ultimately receive an approved treatment.

Whether the primary goal of a clinical trial is effectiveness or efficacy, the scientific validity of the comparison of the new treatment to some standard depends on the comparability of the groups that receive the experimental and control treatments. Randomization of patients to two or more treatment groups is the primary tool to ensure the comparability of samples, at least on average. Hence, it is of utmost importance that the data from each clinical trial be analyzed consistent with the *intent-to-treat* principle,

which dictates that each subject's data be included in the treatment group to which he or she is randomized. This approach is clearly in keeping with an evaluation of the effectiveness of a treatment strategy, but even when evaluation of efficacy is the goal, the clinical trial should ideally be designed in such a way that all randomized patients will contribute to the estimate of treatment efficacy. However, in limited situations, it might be judged acceptable to evaluate efficacy in a *modified intent-to-treat* subgroup of randomized patients defined on the basis of measurements made prior to randomization and ascertained in an unbiased fashion for each treatment group. In this setting, safety would still be evaluated in patients who are not in the subgroup.

In neither effectiveness nor efficacy studies would an analysis based on a *compliant* or *per-protocol* analysis population (defined as patients who adhered strictly to the prescribed dose, frequency, and duration of the assigned treatments) be considered a scientifically rigorous assessment of the treatment. Instead, when the efficacy of the treatment in a compliant population is of interest, one needs to find a way to randomize only those patients who can tolerate the treatment and who will adhere to the protocol (see below).

GOAL: INDICATION FOR A NEW TREATMENT

Ultimately, a new treatment is characterized by its “indication.” An ideal *treatment indication* will consist of a disease, a patient population, an intervention, and an outcome.

The Disease The exact medical definition of “disease” can range from primarily signs and symptoms (e.g., headache, pneumonia) to presumed causative agents (e.g., pneumococcal pneumonia, gram negative pneumonia, fungal pneumonia, or carcinomatous pneumonia). The definition of a disease is frequently not refined further than necessary to decide on an appropriate treatment strategy. In this way, the identification of a beneficial treatment often becomes the definition of the disease (e.g., all gram negative pneumonias are considered together due to the common treatment chosen in those settings). Other times, the lack of efficacy of the usual treatment is incorporated into the definition of the disease (multidrug resistant *Staph aureus*). It is common that particular diseases are diagnosed through a series of tests and procedures. The nomenclature for the disease may include the method of diagnosis (e.g., culture positive gram negative septicemia). However, it is rare that any sign or symptom be “pathognomonic”—uniquely identifying—for the disease.

The Population of Patients Because of concurrent medical conditions, a treatment might be indicated only for a subpopulation of patients who satisfy the diagnostic criteria for the disease. There might be known safe and effective therapies that are regarded as the first line treatment of the disease. In such a case, an indication for a new treatment might indicate the treatment's use only in patients for whom the standard therapy is *a priori* judged inadvisable due to concurrent medical conditions (e.g., pediatrics, pregnancy, poor renal function in a drug cleared by the kidneys) or who cannot take the standard therapy (e.g., due to lack of tolerance with respect to side effects or lack of efficacy).

The Intervention An intervention consists of a formulation of the drug(s) or device(s), a mode of administration, the dosing strategy, ancillary treatments, and the duration of treatment. Some treatments are combinations of drugs, either in a common formulation or administered separately. A mode of administration can include topical, oral, subcutaneous, intramuscular, or intravenous. In some circumstances the mode of administration may even stipulate special training for the person administering the treatment. The dose may be specified as a common level to be used by all individuals or as a dose specific to patient body size or body surface area. The dosing strategy might include a gradual increase in dose as treatment is initiated, a tapering of dose as the patient is weaned from the therapy, a regimen for increasing or decreasing the dose in response to observed patient conditions at the time of dosing (e.g., serum glucose in insulin therapy) or observed patient response to therapy (e.g., increasing the dose if the effect is not optimal or decreasing the dose in the presence of treatment toxicity).

Ancillary treatments might be administered as prophylaxis against known toxic effects (e.g., G-CSF with cancer treatments, antihistamines with drugs that tend to trigger an immune response) or for rescue from toxic effects (e.g., methotrexate followed by leucovorin rescue). One also has to indicate the frequency of administering the treatment. Finally, there is the duration of treatment, which might include “drug holidays.”

The Desired Outcome The intended outcome of a treatment is typically characterized clinically, as outcomes that materially affect the clinical manifestations of the disease (e.g., lower risk of mortality, relief of symptoms, improvement in quality of life). In some settings, a strong risk factor thought to represent a surrogate outcome measure of subclinical disease or disease risk will be used (e.g., hypertension). The distinction between surrogate and clinical outcomes depends on the degree to which a patient’s sense of well-being is directly related to the outcome or the degree to which it is known that any modifications in the biomarker might possibly not be associated with an improvement in the clinical outcome (i.e., treating the symptom but not the disease). The precise definition of the outcome might explicitly include the time frame of measurement (e.g., postprandial serum glucose levels) and the method of measurement (e.g., decreasing serum glucose levels as reflected in Hemoglobin A1c), or the timeframe might only be implicitly defined.

Treatment Discovery Process

The treatment discovery process is an iterative process of studying a disease, hypothesizing and developing treatments, evaluating those treatments, and, for successful treatments, further refining the indication to account for lack of efficacy or toxicities (or both) in particular subgroups of patients. As a rule, the scientific development of a particular treatment indication is often connected with that of other treatments, and thus it may be difficult to identify the exact process that led to the adoption of some treatment. Nevertheless, the following describes a general chronology of events.

Initially, some targeted disease is characterized from observational studies (including epidemiologic studies of risk factors for the disease), clinical observation of typical disease progression and predictors of outcomes, and laboratory studies of

biochemical and histologic changes in the diseased patients' tissues. Often, this characterization of a disease starts with a constellation of symptoms and signs, and much of the ensuing observational research is directed toward finding a causative agent. Observational studies of diseased patients are then often augmented by laboratory experiments that try to further illuminate the causation of the disease and the cellular and physiologic mechanism that cause its major complications. These experiments might involve *in vitro* studies of cell lines and animal models of the disease.

Based on the understanding of the disease gained from the above types of studies, scientists might propose a potential treatment or preventive strategy. The proposed treatment is then evaluated and further refined in a series of preclinical laboratory and animal experiments. Such experiments might focus on two general approaches: *in vitro* characterization of the chemical and biochemical interactions of new drugs with cellular and extracellular constituents of the human body and *in vitro* characterization of the effects of the new therapies on cellular mechanisms using cell lines or animal experiments in suitable species. The goal of this work is to characterize:

- pharmacokinetics, measuring the effect of dose on rates of absorption and excretion of drugs from various body compartments;
- pharmacodynamics, measuring the intended or unintended effects of dose on physiologic measures;
- toxicology, measuring the effect of dose on histopathologic lesions in major organ systems;
- reproductive and embryologic effects as a function of dose; and
- *in vivo* drug-drug interactions that might lead to attenuation or potentiation of intended or unintended effects of the treatment or that might affect the pharmacokinetics of the drugs.

When sufficient preclinical studies have been performed to conclude that the treatment is basically safe, work moves to experiments in human volunteers. In order to sequentially investigate safety and then efficacy and effectiveness issues in a manner that protects human subjects from harm, the process of investigating new treatments typically goes through a phased series of clinical trials. The considerations during each phase depend on whether the investigational product is targeting disease prevention, diagnosis, or treatment, as well as the severity of disease, the type of intervention (e.g., drug, biologic agent, device, behavior), and the prior knowledge of treatment risks. The following thus describes only the general principles behind the phased investigation.

Clinical Trial Phases

Phase I clinical trials provide initial safety data to support further testing with larger samples. As the focus of these studies is primarily safety of investigation rather than efficacy or effectiveness of treatment, the study subjects are frequently a small number of healthy volunteers. A notable exception occurs when a treatment that is designed to be administered in life-threatening disease is known to have severe toxicity. For instance, in phase I cancer clinical trials, the treatment might be first tested in patients whose disease has proven resistant to all other therapies.

“First in human” clinical trials might start with a single administration of the treatment at an extremely low dose in a few subjects. In the absence of unacceptable toxicity, subsequent patients might receive increasing doses. Owing to a desire to slowly increase exposure to the treatment, patients may not be randomized across all doses. In cancer chemotherapy trials, in particular, there may be no control group. Pharmacokinetic data might be gathered on both single doses and repeated dosing to assess the rates of absorption and elimination in humans. These kinds of studies might also consider pharmacokinetics in the presence of renal or hepatic impairment, as well as pharmacokinetics in the presence of meals and other drugs. When phase I trial results show unexpected severe toxicities, further consideration of the treatment might be curtailed.

Phase II clinical trials seek further safety data and preliminary evidence in support of biological effect. A slightly larger sample of subjects are administered the treatment at a dose or doses that were preliminarily judged safe in the phase I studies. Safety data are collected in a systematic fashion, including specified monitoring of any potential side effects that were identified previously. Phase II studies also serve to screen for treatments that show some sign of biological effect, such as a biological marker that is a surrogate for the clinical outcome that is of interest. Products that fail to demonstrate a certain level of biological activity might be abandoned. Such a screening process is more efficient than other approaches in finding effective treatments from a large population of ideas.

Even when the phase II clinical trials demonstrate a desired effect on the biologic outcome, it is common for investigators to use the results of the clinical trial to identify more specific factors:

- a more precise definition of the disease characteristics that would indicate the types of patients likely to benefit most from the treatment,
- a more refined definition of the population to be treated in order to eliminate subjects who might experience greater toxicity,
- a single treatment regimen (dose or dosing strategy, frequency, duration, ancillary prophylactic or rescue therapies), or
- a clinical measure to serve as the primary outcome, as well as a statistical measure to summarize the distribution of that clinical outcome across subjects.

The selection of this primary outcome (and summary measure) is based on consideration of (in order of importance): (1) the clinical measure that is most indicative of an improved clinical outcome for the patient, (2) a measure that the treatment might plausibly affect, and (3) an outcome that can be compared across treatment groups with good statistical precision.

Phase III clinical trials, which are the main focus of the panel’s report, are large confirmatory studies meant to establish an acceptable benefit/safety profile in order to gain regulatory approval for a precisely defined indication (“registrational” clinical trials). Phase III trials are well-controlled trials that provide scientifically credible and statistically strong evidence about the treatment indication hypothesized at the end of phase II investigation.

In order for a phase III trial to be regarded as confirmatory, it is crucial that the hypotheses being tested be specified before the start of the trial. Sample sizes are typically chosen to have a high probability of ruling out the possibility of ineffective therapies and to estimate the treatment effect with high statistical precision. Collection of safety data continues to play a major role, as the larger sample sizes in the phase III study afford a better opportunity to identify relatively rare serious toxicities. As a general rule, the approval process does not demand statistically proven increased rates of toxicities prior to providing warnings to patients and physicians. Depending on the disease and patient population, anecdotal occurrence of unexpected extremely serious adverse events will often dictate further study of a proposed treatment.

Evidence from phase III studies that strongly support the proposed indication will generally lead to adoption of the therapy. Sometimes, however, even when a proposed treatment has “met its outcome” in the overall study population, the indication (treatment) actually adopted might be more restrictive than was initially proposed due to lessened efficacy or heightened toxicity observed in a subgroup of patients.

Suppose, for example, that there are two subpopulations, A and B, and that the proposed therapy “met its outcome” in the combined sample. But suppose that when analyzed alone, subpopulation B did not appear to have an acceptable benefit/risk ratio (which indicates that subpopulation A exhibited a strong benefit of the treatment). Because it is not uncommon for proposed treatments to present safety issues, more focus is often placed on making sure that harmful treatments do not get adopted. In this example, subpopulation A might be approved to receive the new treatment, while regulators require additional data in support of the benefits of the treatment for subpopulation B.

There are two potential drawbacks to this “data-driven” restriction of indication. One is that if the observed difference in treatment benefit/risk is spurious, subpopulation B is deprived of a useful therapy until additional data is gathered. The other is that if the observed difference in treatment benefit/risk is spurious, the commercial sponsor will have lost income from sales to B as well as having the added expense of further studies in that subgroup.

Phase IV clinical trials are postmarketing trials that are meant to evaluate rare but serious effects that cannot be assessed in the smaller Phase III studies.

The above description is most applicable to the evaluation of new therapies. In disease prevention, some authors have suggested that phase III trials focus on efficacy by demonstrating the treatment benefit of prevention through some surrogate biomarker of the disease (e.g., colon polyps as a precursor lesion for colon cancer) and phase IV trials focus on effectiveness, using the clinically relevant outcome in a population-based sample of the types of individuals likely to receive an adopted treatment. In doing so, these studies may also aim to evaluate changes in individual behavior that might mitigate the efficacy of the treatment (e.g., increased risk-taking behavior when vaccinated for HIV or treated for peanut allergies).

GOAL OF EFFECTIVENESS VERSUS EFFICACY

Evidence-based medicine often involves a stepwise process that closely parallels the parts of a treatment indication described above. These steps consider: (1) patient population (the definition of the disease and any restrictions on patient characteristics apart from disease manifestations), (2) intervention, (3) comparison (alternatives to the intervention that might be considered), and (4) outcome (the clinical condition that is desired), and they are referred to as PICO. Distinctions between effectiveness and efficacy are given below in the context of PICO.

Effectiveness: Phase III Trials

An effective treatment is one that provides improvement in the general health of the population viewed as a whole: “general health” in some sense considers the average state of health of the population. An “effectiveness trial” enrolls a representative sample from the population of patients who would eventually receive the treatment. The effectiveness study should strive for an inclusive setting, which might include both independently living as well as institutionalized patients, and it should, insofar as safety permits, not restrict patients based on concomitant disease unless such a restriction will be in the ultimate indication.

Ideally, the eligibility criteria would consist of inclusion criteria that define the population of patients that would ultimately be in the indication for the treatment, the criteria would also delineate patients who might be inappropriate for a randomized controlled trial evaluating an unproven therapy. The intervention would then be administered as it will be given to a patient, which might include (a) decreased dosage due to lack of tolerability, (b) lack of compliance on the part of the patient, (c) ancillary treatments used to prevent or treat unintended effects of the treatment, and (d) other changes of patient or treating clinician behavior.

The trial would then compare the new intervention to the treatment the population of patients would otherwise (in the absence of the new intervention) have received and evaluate an outcome that is the best summary of “general health” for the patient population, which might be affected by (a) other changes in behavior that are associated with receiving the treatment, and (b) the timing of the intervention and the timing and methods of measurement for the outcome.

Efficacy: Phase II Trials

Treatment efficacy can be defined in a subset of the patients who would eventually be treated, and it can be based on an outcome that is merely thought to be an indicator of eventual clinical benefit. An “efficacy trial” might enroll patients from a defined subset of diseased patients who are most likely to show evidence of treatment effect. This could be because (a) they have been previously (prior to randomization) found to be able to tolerate the treatment (e.g., during a screening “run-in” phase), (b) they have been previously (prior to randomization) found to be compliant with randomized controlled trials procedures (e.g., during a screening “run-in” phase), or (c) there is reason to believe prior to randomization that they are relatively more likely to have a beneficial treatment effect than the population of all patients with the disease. A

key point is that all randomized patients will be analyzed for their outcomes. Any “enrichment” of the sample to maximize response needs to be done prior to randomization.

The intervention must be clearly defined and may differ from the eventual (“effectiveness”) intervention for several reasons: (a) the care providers administering the treatment are more highly trained, (b) the treatment protocol is more rigidly enforced, (c) inducements for high compliance are used, or (d) ancillary treatments and additional treatments in the presence of lack of efficacy are restricted or proscribed. In addition, the care that the comparison groups receive might not be the standard of care the patient would have received in the absence of the randomized controlled trial. Moreover, the primary outcome may not be the clinical outcome of greatest public health relevance because it may be measured using techniques or schedules that do not coincide with usual clinical practice (e.g., heightened radiographic surveillance for subclinical progression of cancer rather than examinations based on clinical events), or it may be an intermediate marker that is believed, but not known, to be a necessary and sufficient indicator of the true clinical outcome (e.g., tumor response in a cancer clinical trial, arrhythmias in studies of survival following myocardial infarction).

Notes on Efficacy Versus Effectiveness

True effectiveness can never be tested in an unbiased fashion because the randomized controlled trial setting itself is artificial and because observational studies are always subject to unmeasured bias. Nonetheless, it is important that in phase III trials, the effectiveness of a therapy be assessed as accurately and precisely as possible.

An efficacious treatment may not be effective for at least four major reasons. First, the kind of patients who were not represented in the efficacy trial have worse clinical outcomes that overwhelm any benefit seen in the efficacy trial sample. This may happen because (a) they have a heightened susceptibility to serious adverse events leading to poor clinical outcomes; (b) they cannot tolerate the treatment, and the therapeutic window for administering alternative therapies has passed; (c) the broader population of patients includes individuals whose disease is so mild or so severe that the intervention provides no benefit, but those patients do experience toxicities; or (d) off-label use of the therapy confers risk but no benefit at a level that outweighs the benefit in the more restrictive population in the indication.

Second, the intervention tested in the efficacy trial differs from the intervention that would be realized in the more inclusive population of patients with the disease or condition, because (a) the skill of the investigators administering the intervention is necessary for the treatment benefit, but that skill is not present in the general setting; (b) the efficacy trial restricted use of ancillary treatments that interact negatively with the experimental treatment; (c) the efficacy trial restricted use of ancillary treatments that are in wide use in the population and provide the same benefit as the treatment (perhaps with fewer toxicities); or (d) the compliance of patients with the experimental therapy is markedly worse than was achieved in the efficacy trial, and the toxic effects of the therapy are manifested with lesser exposure than the beneficial effects.

Third, the comparison group in the efficacy trial does not encompass the true standard of care that patients would receive in the absence of the experimental treatment, and the experimental treatment does not provide added benefit over that standard of care.

Fourth, the primary outcome used in the efficacy trial is not predictive of the true clinical outcome, because (a) the predictive value of an intermediate marker is affected by the treatment (i.e., “treating a symptom, not the disease”); (b) the schedule of outcome assessment in the efficacy trial led to additional beneficial ancillary treatments that are not realized in standard medical care; or (c) the population of patients changed their behavior (e.g., risk taking) when taking a treatment that is or is thought to be protective.

DEFINITIONS

Treatments

As noted above, it is useful to differentiate between the concepts of a simple *treatment*, a *treatment regimen*, and a *treatment strategy*.

- Treatment (sometimes referred to as nominal experimental treatment) includes formulation, administration, dose (fixed, per weight, per body surface area, adaptive), frequency (including drug holidays), and duration.
- Treatment regimen includes nominal experimental treatment as above, prescribed prophylactic treatments to prevent adverse events, dose modifications in the presence of adverse events or demonstrated efficacy, and prescribed ancillary treatments for known adverse events.
- Treatment strategy includes treatment regimen as above, patient compliance, ancillary treatments according to the usual standards of care, and rescue treatments for lack of effect following the usual standards of care with prior characterization of potential rescue treatments. Rescue treatment may represent: (a) a second-line (less effective) treatment used in failure of primary therapy, (b) a crossover to an established standard of care that is used as the control treatment, (c) a crossover to the experimental treatment, or (d) a progression to a treatment known to be more effective, but avoided for other reasons (e.g., opiates in pain relief). The treatment strategy is what is truly tested when randomized controlled trial data is analyzed.

Study Design

The following are some common study designs for randomized clinical trials:

- Randomized cohort design: eligible patients are randomized to therapy and followed for outcomes.
- Prerandomization run-in: patients are started on a placebo to measure compliance with treatment and study procedures or are started on an experimental treatment to ensure tolerance.
- Randomized withdrawal: All subjects start on experimental treatment, and proof of efficacy is based on worsened clinical status following randomized withdrawal.

Time Frame of Measurement

The following are some common time frames for randomized clinical trials:

- Single fixed study time: outcomes are assessed at some fixed time defined post randomization.
- Interval study time: outcome is averaged over a specified interval of time postrandomization, or outcome is contrasted over a specified interval of time post randomization.
- Single fixed event time: outcome is assessed at some time defined by a particular event (e.g., childbirth).
- Interval based on event: outcome is assessed over the interval up to a particular event (e.g., time until liver transplant).
- Administratively censored time to event: outcome is time to some defined event, length of follow-up may vary by individual, and censoring occurs only due to time from randomization to data analysis
- Time to event subject to competing risks: outcome is time to some defined event providing it occurs prior to another (nuisance) event that would preclude ability to measure, length of follow-up may vary by individual, and (scientific relevance depends on whether competing risk is noninformative and whether other processes will alter risk of the (nuisance) event.

Scientific Outcomes

Two scientific outcomes for randomized clinical trials are common. One is clinical outcomes, which include survival, specific quality-of-life factors (e.g., serious events leading to hospitalization, diminished functioning such as nonfatal myocardial infarctions, resolution of a chief complaint such as headache), and general quality of life. The other is surrogate outcomes, which include modification of risk factors for clinical outcomes (e.g., blood pressure, HbA1c), intermediate subclinical outcomes (e.g., tumor progression), and biomarkers (e.g., PSA)

There are also studies with multiple outcomes. They include:

(1) Coprimary outcomes: the treatment must demonstrate effect on each of several outcomes separately, though there are situations in which the individuals do not need to meet each of the outcomes, which would include cases when safety and efficacy are evaluated separately.

(2) An outcome index: an index for each individual is defined as the sum or average of measurements made on several different outcomes.

(3) Composite outcomes, which include: (a) good outcome is defined by an individual's meeting all outcomes, (b) bad outcome is defined by an individual's failing to meet any of the outcomes, and (c) time of bad outcome is defined as the earliest occurrence of any undesirable event

Subject Participation

Eight levels of participation for subject participants can be specified.

- *Screening only*: an individual is considered for inclusion, may have protocol-specified measurements, and there are no protocol-specified interventions or treatments.
- *Run in*: individuals receive protocol specified interventions, namely, placebo (for general compliance behavior) or experimental (for tolerance to treatment, and are nonrandomized (for individual specific measures) or randomized (for investigator training). Evaluation of outcomes will not be included in evaluation of efficacy or effectiveness, but if an individual receives an experimental treatment, safety outcomes likely will be evaluated.
- *Enrolled*: patients are included who were ever assigned (by the criterion specified in the protocol) to receive the study intervention, and, in a randomized study, they are subjects who receive a randomization code.
- *Active participation with treatment*: subjects adhere to some part of nominal treatment or treatment regimen, or subjects adhere to monitoring schedule.
- *Active follow-up after study treatment discontinuation*: a subject has stopped nominal treatment or treatment regimen but is adhering to full monitoring schedule.
- *Reduced follow-up after study treatment discontinuation*: a subject has stopped nominal treatment or treatment regimen, as well as most invasive or inconvenient monitoring schedule, and but is still followed for passively observable major clinical outcomes (e.g., survival).
- *Loss to follow-up*: clinical investigators cannot contact the participant, though participation may resume as “active participation with treatment,” “active follow-up after study treatment discontinuation,” or “reduced follow-up after study treatment discontinuation” in the event the participant is later found.
- *Withdrawn consent*: the subject has withdrawn consent for further participation of any kind.

Analysis Populations

The scientific and statistical validity of a randomized controlled trial depends on the comparability of the treatment groups. That comparability is achieved (on average) at the start of a study by randomizing patients to treatments. All events that occur postrandomization are then, plausibly, the result of the treatment assignment. Several terms are used to describe the analysis populations often discussed in clinical trials:

- *Intent to treat*: covers all patients who were ever enrolled. Patients are included in their assigned treatment group. In a randomized study, this population guarantees comparability of treatment groups, and only this analysis population allows generalizability to specified eligibility criteria.
- *Modified intent to treat*: covers a subsample of enrolled patients for which the comparability of randomized groups and generalization of specified eligibility criteria is valid. Subjects are included in their assigned treatment group

regardless of treatment actually received. Exclusion of enrolled patients is based on criteria defined prior to enrollment, though the reporting of the measurements used as the basis might be delayed for logistical reasons. No postrandomization events can be allowed to influence eligibility. The purpose is generally to focus an efficacy (not effectiveness) analysis for a subset of patients for whom the treatment is hypothesized to work best, but logistics precludes identification of that group in real time. As an example, an efficacy trial of a treatment for gram negative sepsis may use only those patients whose blood cultures obtained prior to enrollment are found to be positive for gram negative organisms on the laboratory reading performed 48 hours post specimen collection

- *Experimental treatment population (per protocol)*: Covers the subset of enrolled patients who received any amount of the study drug (or other treatment). Patients are included in the assigned treatment group. This group does not include patients who were randomized but never received any treatment. Comparability of treatment groups is compromised in an unblinded study because the reasons for not administering the assigned treatment might be based on investigators' or subjects' biases.
- *Safety population*: covers the patients included for the experimental treatment population, but any patient receiving the experimental intervention is analyzed with the experimental group.

Types of Clinical Trial Data

The types of data collected in a clinical trial can be characterized by their ultimate use:

- *Prerandomization*: data include determination of eligibility, indication of stage or severity of disease, indication of concomitant risk factors, indication of important subgroups for specified or exploratory analyses.
- *Postrandomization primary treatment compliance*: data include information on compliance (dose reduction, delay, and termination, protocol-specified adaptations versus patient/provider choice) and on realized treatment, which includes duration of treatment, cumulative dosing, and dose intensity.
- *Postrandomization concomitant or ancillary treatments*: data include safety and efficacy outcomes, including intermediate measures and surrogate measures, as well as measures of secondary outcomes.

Mechanisms of Missing Data

There are a variety of ways that data that are intended to be collected in a clinical trial can be missing. A patient can fail to be included in the denominator for which measurement is scientifically relevant for at least two major reasons. One is that the patient was never included for scientific reasons (e.g., pregnancy test in men) or for efficiency reasons. The second is that the patient is no longer included due to end of protocol-specified time frame due to scientific reasons (e.g., death), efficacy or efficiency

reasons (e.g., symptom relief in a trial separating efficacy and safety analyses), or ethical reasons (e.g., crossover to a known, more effective rescue therapy).

There is also item nonresponse, which can be due to: (1) clinical inadvisability for specific invasive procedures (e.g., liver biopsy in patient with bleeding disorder), (2) patient refusal for specific invasive procedures or measurements (e.g., refused biopsies), (3) patient refusal to answer specific questions (e.g., sex behavior, income), or (4) patient's missing clinic visits on time-sensitive measurements.

There is administrative missingness, when the protocol allows study termination prior to complete data collection on each subject, leading to missing repeated measures or censored time to event. There is also missingness from competing risks (e.g., censoring by death from other causes in a cancer clinical trial), missingness due to treatment noncompliance (which is relevant when trying to evaluate a treatment or treatment regimen, rather than a treatment strategy), missingness due to loss to follow-up, missingness due to withdrawal of consent, and missingness due to data editing (values out of range).

Sensitivity Analysis

In the body of the report we focus our discussion of sensitivity analyses on sensitivity to the assumption about the underlying mechanism producing the missing values. There are other aspects of a statistical model for which sensitivity should be assessed. Here is an outline of the steps leading to a comprehensive sensitivity analysis for such models:

(1) *Presumed mechanisms of missing data:* steps would include identification of data likely to be missing, speculation on mechanisms leading to that missing data, and specification of analyses of missing data patterns

(2) *Planned analyses to deal with missing data:* presumed model assumes either missing completely at random, missing at random, or missing not at random (as defined in Chapter 3); the population with available data that will be used (e.g., complete cases, all available data, etc.); the variables that will be used; how variables will be modeled; distributional assumptions; the statistical model; and the statistical paradigm (Bayesian, frequentist, likelihood).

(3) *Sensitivity analyses:* one will need (a) a framework for exploring effect of distributional assumptions, (b) a framework for exploring effect of variable modeling (e.g., linear, dichotomized, interactions), (c) a framework for exploring effect of considering other variables, (d) a framework for exploring effect of changing population used for modeling, (e) a framework for exploring effect of assumptions of missing at random or missing not at random, and, finally, (f) possible augmented data collection that can shed light on assumptions.