Types of Clinical Research

The widely promulgated goal of practicing evidence-based medicine is relatively recent, and for most of medical history practice has been largely anecdotal. For example, Voltaire commented that early physicians “poured drugs of which they knew little, for diseases of which they knew less, into humans of which they knew nothing.” We now know much about human physiology, pathophysiology, and pharmacology. And we know a lot about rodents. However, any physician who tries restricting practice to methods established on the basis of strong evidence in humans will quickly discover that there is distressingly little basis for current medical care. The purpose of clinical research is to bridge the gap between basic science and results in animals to patients.

Clinical research can be broadly categorized as retrospective (case reports, chart reviews, and database investigations); observational (case series); cohort, prevalence, and case-control; and randomized clinical trials. Retrospective studies are inherently weak and it is impossible to make valid conclusions about causal inference. Furthermore, data quality is usually poor since clinical records are often inaccurate and rarely contain all the detail necessary for research. And while sophisticated statistical techniques can limit the effects of bias and confounding, statistical analysis cannot eliminate the methodologic errors associated with most retrospective studies.

Observational studies range from having limited value to being important techniques. Case reports and case series are of least importance in this hierarchy. The danger is that the results of case series are almost always implicitly — or worse explicitly — compared with previous results. This is a classical “historical control” and usually generates the conclusion that the new treatment or local management is superior — a conclusion that startlingly often is shown to be false in subsequent randomized trials. But that said, analysis of large databases can help evaluate real-world effectiveness of various healthcare strategies.

Population-based observational research is the proper tool for determining the natural history of various conditions. Examples include weight gain or development of hypertension as a function of time in a population or as a function of age in individuals. But it is essential to avoid unwarranted causal inference such as the conclusion that health will be improved by preventing obesity or hypertension — which may or may not prove to be the case. Another important use of observational research is in case-control studies which may be the only valid way to study conditions so rare that randomization is impossible.

The gold standard for clinical evidence is a randomized, blinded trial because the technique is least subject to bias and confounding. The first major randomized trial evaluated the efficacy of streptomycin as a treatment for tuberculosis — in 1948. This relatively late date emphasizes how recent the entire concept of evidence-based medicine is.

The term “outcomes research” is used in a variety of contexts including clinical trials, epidemiologic studies, and health-services investigations. But in reference to clinical trials, it generally refers to large randomized studies that are powered to detect major outcomes such as duration of hospitalization, functional status, and mortality. They are, consequently, usually expensive and often multi-center.

Advantages of multi-center trials include large enrollment and somewhat greater diversity in the study population. Diversity improves generalizability, but at the cost of statistical power since variability increases. Another advantage is that problems in one or more centers, such as poor study conduct or even fraud, will be balanced by the remaining centers. For example, centers reporting unusual results can be identified statistically and subject to enhanced auditing. For this reason, the US Food and Drug Administration usually requires participation of multiple centers in pivotal trials.

The major difficulty with multi-center trials is that there are inevitably subtle — but possibly important — differences in protocol implementation because clinical routine naturally varies among sites. And of course it is considerably more expensive to operate at multiple sites than at a single center. All-in-all, multi-center trials are not necessarily superior to single-center trials; they should thus be evaluated critically just like single-center studies.

Meta-analysis is a type of systematic review that has recently become popular. Unfortunately, the popularity of meta-analysis appears to be driven more by the technique’s relative ease than its intrinsic validity. The general theory is that statistically combining a number of under-powered or marginally powered studies will produce a well-powered and robust conclusion. However, meta-analyses are subject to substantial limitations which can markedly reduce their validity.
Major limitations of meta-analyses include search and selection bias because the choice of articles to includes obviously influences the results. Publication bias is more subtle, but probably more important. The results of many studies are not published, but unpublished results are not randomly distributed. Instead, “negative” results are often withheld by the authors, concealed by corporate sponsors, or rejected by journals either because the work is under-powered or for lack of interest. It is appropriate for journals to reject under-powered studies, which is often the case for “negative” results, but these results are then rarely available for meta-analysis and contribute to a bias towards positive results. Corporate sponsors sometimes refuse to publish results of completed studies because the results are not to their liking; most investigators consider this practice to be a type of research fraud.

**Sources of Error**

There is no perfect study. All are limited by practical and ethical considerations. And it is impossible to control all potential bias and confounding factors — even in the best randomized and blinded trials. Multiple studies are thus required to convincing confirm (or disprove) a hypothesis. There are five major causes of error in clinical research: **selection bias**, **measurement bias**, **confounding**, **reverse causation**, and **chance** (random error). It is important to recognize that only the last source of error, chance, is fully addressed by statistical analysis. In contrast, statistical analysis has limited ability to compensate for systematic error.

Selection bias results when participants for study inclusion or for treatment are chosen non-randomly. In retrospective studies, this can result in subtle forms of the disease being missed or treatment being directed to patients most likely to benefit. For example, patients with better education or stronger support systems may seek or be given more aggressive treatment. Similarly, patients may comply poorly or even stop treatment, either for perceived lack of efficacy or because of side effects. The extent to which any of these non-random events occurs is usually difficult to assess.

Selection bias is largely prevented by proper randomization in which patients, caregivers, and investigators are all prevented from altering the designated treatment allocation. But even within the context of proper randomization, enrollment criteria can be designed to favor patients most likely to benefit from an experimental treatment. And patients who volunteer for trials may differ in important ways from those who do not. Both factors reduces the extent to which the results can be generalized to a wider population.

Measurement bias results when data quality or availability varies non-randomly. In retrospective studies, the quality of records is usually poor since they were designed for clinical purposes rather than research. The difficulty is that record quality may vary non-randomly; for example, patients given new treatments may be watched more closely and enthusiastic clinicians may over-estimate treatment benefit or under-estimate associated complications. The importance of placebo effect should never be underestimated. This is especially the case for subjective responses such as pain and quality-of-life, but placebo effects have also repeatedly been shown to influence supposedly objective outcomes. Measurement bias is largely prevented by double-blinding (i.e., both patient and clinician are unaware of treatment allocation).

Confounding is defined by an apparent association between two factors that is influenced or caused by a third factor. This is a subtle, but important, source of error since alternative influences may not even be suspected. And in retrospective studies, potential confounding factors may be well known but unavailable for analysis by virtue of not being included in clinical records. For example, blood transfusions are strongly associated with adverse outcomes including mortality. However, blood transfusions are far more likely to be required by patients who are sick (i.e., have anemia of chronic disease) or have the longest and largest operations. The extent to which these factors contribute can be difficult to determine retrospectively. As with selection bias, randomization provides considerable protection against confounding.

Reverse causation is a special type of confounding in which the treatment or factor of interest causes or unmasks the disease in question. For example, retrospective analysis might show that morphine use is significantly more common in patients with gall bladder disease than in those without disease. However, morphine worsens gall bladder symptoms by provoking sphincter of Oddi constriction, which thus promotes early diagnosis of gall bladder disease. A conclusion that morphine causes gall bladder disease would thus be incorrect.

**Study Design**

Studies should have a formal hypothesis, usually the last sentence of the protocol or manuscript’s introduction. The hypothesis is a specific statement of the study’s goals and identifies a priori primary and secondary outcomes.
Research reports lacking a hypothesis should be viewed with considerable caution as it is the hypothesis that largely determines the optimal study design.

When trials are randomized, allocation should be based on computer-generated codes (stratified, if appropriate). Importantly, treatment allocation should be concealed from both subjects and investigators until the last possible minute to minimize the risk of selection bias. Data analysis for major outcome trials is usually based on intention-to-treat; that is, they include results from all randomized subjects whether or not they received or completed the designated treatment. However, it can be appropriate to base analysis on the treatment subjects actually received in small studies such as those evaluating physiological mechanisms or pharmacokinetics.

Defined inclusion and exclusion criteria will determine who participates in a given study. Subject selection is the “art” of clinical trials since it is a delicate balance between choosing patients at risk for the condition and likely to benefit from treatment, response variability, and availability of subjects. On one hand, broadening enrollment criteria facilitates recruitment, but simultaneously adds variability which increases sample-size requirements. On the other, tighter enrollment criteria reduces the number of subjects required, but may make adequate recruitment impossible. The results of studies most immediately apply to subjects meeting the inclusion and exclusion criteria. A consequence of broad or restrictive enrollment strategies are results that are more or less generalizable to larger populations.

Protocols should be designed with formal a priori sample size estimates, including designated interim analysis points (if any). The assumptions used should be included in research reports, including expected baseline outcome rates and anticipated treatment effects, along with the basis for terminating the study for futility. The reason for restricting analysis to defined points, and taking an appropriate statistical penalty, is that investigators over-estimate significance by taking “multiple-looks” and terminating a trial when \( P \) is first <0.05.

Blinding (also called masking) is the only reliable way to prevent measurement bias. Blinding, along with randomization, are thus the two most important aspects of study design. Studies are considered single-blind when one party, usually the patient, is unaware of the designated treatment. An unblinded investigator is sometimes necessary because it is impossible to mask all interventions; however, outcomes should, whenever possible, be evaluated by a separate set of investigators who are unaware of the patients’ group assignments. Best is to mask both patients and investigators, making the study fully double-blinded. Data entry and statistical analysis require considerably more judgment than generally appreciated, and bias can creep into either process. It is thus best to maintain blinding throughout analysis.

**Data Analysis and Interpretation**

Perhaps the most important thing to understand about statistics is that statistical analysis is largely designed to detect random error, but has little ability to detect bias, confounding, and other sources of systematic error. Statistics thus compensate poorly for inadequate study design or execution. A corollary is that a significant \( P \) value by no means assures that the result of a particular investigation is likely to be “true.”

Common statistical/design errors include failing to define a priori primary and secondary outcomes which promotes “fishing” for a significant result. The problem with this approach is that it has the potential to vastly over-estimate the actual degree of statistical validity (alpha power). Similarly, sample-size and stopping rules are often not defined a priori which again promotes “multiple-looks” which diminishes true statistical power. The planned statistical analysis should also be defined before the study — and of course the right statistical tests should be used.