

Background: Experimental Design and the Collection of Data

Psychology (Statistics) 484

Statistics, Ethics, and the Social and Behavioral Sciences

June 19, 2013

Beginning Quotations

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

Control what you can, and randomize the rest.

– Old statistics adage

Today's scientists have substituted mathematics for experiments, and they wander off through equation after equation, and eventually build a structure which has no relation to reality.

– Nikola Tesla

... the main business of clinical research is not enhancing or saving lives but acquiring stuff: data. It is an industry, not a social service.

– Sonia Shah (*The Body Hunters*)

There are only a handful of ways to do a study properly but a thousand ways to do it wrong.

– D. L. Sackett (1986)

Week 13: Background: Experimental Design and the Collection of Data

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

** Exam 3 given on Thursday (covers the third four weeks)

— difficulties with observational studies; the hormone replacement therapy controversy and related artifactual interpretative issues

Required Reading:

SGEP (387–419) —

Observational Studies: Interpretation

Observational Studies: Types

Observational Studies: Additional Cautions

Controlled Studies

Controlled Studies: Additional Sources of Bias

The special case of medical trials

An introductory oddity: The 1954 Salk polio vaccine trials

The Randomized Response Method

Popular Articles —

Influence of Funding Source on Outcome, Validity, and Reliability of Pharmaceutical Research, Report 10 of the Council on Scientific Affairs of the American Medical Association

Sponsorship, Authorship, and Accountability: International Committee of Medical Journal Editors (August, 2007)

Suggested Reading:

Suggested Reading on the Production of Data and Experimental Design

Film: *The Fog of War* (107 minutes)

Introduction

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

Any discussion of data collection, surveys, and experimental design will involve many words and phrases that may not have an immediately obvious interpretation. We first characterize some of these below, in no particular order:

Controlled experiment: The results from a group where a treatment was imposed are compared to the results from a second (control) group that is identical except for the treatment administration.

Observational study: (sometimes referred to as a natural or quasi-experiment, or a “status” study) A study of group(s) that precludes actual treatment manipulation; in other words, one has to “play the hand that is dealt.”

Retrospective study: Subjects are first identified by some procedure; subsequent to this identification, historical data are then collected. The often used “case-control” methodology, to be discussed shortly, is perhaps the best-known example of a retrospective study.

Prospective study: Subjects are identified and new data are then obtained.

Counterbalance: Attempts to even out various effects, such as priming or carry-over, by systematic or random presentation strategies.

Placebo: A typically harmless treatment administered as a control in comparison to a real treatment.

Nonresponse bias: Those individuals responding to questions are inherently different in unknown ways compared to those who do not respond.

Double-blind: Neither subjects nor experimenters know who is getting what treatment.

Wait list: When it is unethical to deny treatment, it still may be possible to have a randomly constructed group just wait some period of time for treatment. In the meantime, responses for the “wait” group could still be obtained and compared to those from subjects immediately given the treatment.

Placebo effect: An improvement in a person's condition regarded as the effect of the person's belief in the utility of the treatment used.

Voluntary response: Individuals who volunteer their responses or participation may be inherently different than nonvolunteers.

Framing (wording) of questions: As always, context is crucial and different responses may be given depending on how a question is asked and by whom. To emphasize the importance of wording and context in asking questions of an individual, we refer the reader to an article written by Dalia Sussman, "Opinion Polling: A Question of What to Ask" (*New York Times*, February 27, 2010).

External Validity

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

The earlier discussion of inferring causality listed eight threats to the internal validity of a quasi-experiment.

A second type of validity is also of general interest in both true and quasi-experiments, “external validity” (or possibly, “ecological validity”):

to what extent do the results provide a correct basis for generalizations to other populations, settings, or treatment and measurement variables.

We list six threats to external validity that should be noted whenever data are to be collected and interpreted (again, from Winch & Campbell, 1969):

1. *Interaction effects of testing*: the effect of a pretest in increasing or decreasing the respondent's sensitivity or responsiveness to the experimental variable, thus making the results obtained for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the experimental respondents were selected.
2. *Interaction of selection and experimental treatment*: unrepresentative responsiveness of the treated population.
3. *Reactive effects of experimental arrangements*: "artificiality"; conditions making the experimental setting atypical of conditions of regular application of the treatment;

4. *Multiple-treatment interference*: where multiple treatments are jointly applied, effects atypical of the separate application of the treatments.
5. *Irrelevant responsiveness of measures*: apparent effects produced by inclusion of irrelevant components in complex measures.
6. *Irrelevant replicability of treatments*: failure of replications of complex treatments to include those components actually responsible for the effects.

Observational Studies: Interpretation

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

Caution is needed whenever the results of an observational study are used to inform decisions regarding health practice, social policy, or other similar choices.

A good illustration is the Nurses' Health Study, started by Frank Speizer at Harvard in 1976 to assess the long-term consequences of oral contraceptive use.

This prospective cohort study of about 122,000 nurses came to a dramatic conclusion in 1985—women taking estrogen had only a third as many heart attacks as women who had never taken the drug.

The inference was made that estrogen was protective against heart attacks until women passed through menopause.

This belief provided the foundation for therapeutic practice for the next two decades, at least until the results of two clinical trials were announced.

The two trials, HERS (Hulley et al., 1998; Heart and Estrogen-progestin Replacement Study) and WHI (Rossouw et al., 2002; Women's Health Initiative) came to conclusions opposite that for a protective effect of Hormone Replacement Therapy (or, more commonly, HRT);

in fact, HRT constituted a potential health risk for all postmenopausal women, particularly for heart attacks, strokes, blood clots, breast cancer, and possibly even dementia.

Discrepancies between the results of observational studies and randomized clinical trials appear so frequently that some epidemiologists question the entire viability of the field.

A 2001 editorial in the *International Journal of Epidemiology* by George Davey Smith and Shah Ebrahim was aptly entitled “Epidemiology—Is It Time to Call It a Day?”.

This echoes a quotation from John Bailar III at the National Academy of Sciences (as reported in a Gary Taubes article in the *New York Times* [September 16, 2007]; this is part of your required reading):

The appropriate question is not whether there are uncertainties about epidemiologic data, rather, it is whether the uncertainties are so great that one cannot draw useful conclusions from the data.

Healthy-User Bias

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

The differences between what was seen in the Nurses' observational study in comparison to the clinical trials is likely due to the *healthy-user bias*.

Basically, women who were on HRT are different than those who are not, both in engaging in activities that are good for them—taking a prescribed drug, eating a healthy diet, exercise—and in their demographics—thinner, better educated, wealthier, more health conscious generally.

There are several effects similar to or part of the healthy-user bias that should be noted as possible explanatory mechanisms for associations seen in observational data.

One that has the potential to be particularly insidious is called the *compliance or adherer effect*.

Two other possible explanatory mechanisms for what we might see in an observational study are the *prescriber effect* and the *eager-patient effect*.

Apart from the situation introduced earlier of asking questions of an individual regarding matters of opinion, the appropriate framing of questions is also crucial to the collection of valid data in all health-related studies.

Observational Studies: Types

The field of epidemiology is concerned with diseases and injuries, and how they might be caused and/or prevented.

Because it is typically unethical to do a randomized clinical trial with the type of agents of interest to epidemiologists, observational data may be the only information available.

In an observational framework, four types of design are typically identified:

cohort, case-control, cross-section, ecological.

We give the definitions for each of these below, taken from Green, Freedman, and Gordis (2000), "Reference Guide on Epidemiology," in the *Reference Manual on Scientific Evidence*.

cohort study. The method of epidemiologic study in which groups of individuals can be identified who are, have been, or in the future may be differentially exposed to an agent or agents hypothesized to influence the probability of occurrence of a disease or other outcome. The groups are observed to find out if the exposed group is more likely to develop disease. The alternative terms for a cohort study (concurrent study, follow-up study, incidence study, longitudinal study, prospective study) describe an essential feature of the method, which is observation of the population for a sufficient number of person-years to generate reliable incidence or mortality rates in the population subsets. This generally implies study of a large population, study for a prolonged period (years), or both.

case-control study. Also, case-comparison study, case history study, case referent study, retrospective study. A study that starts with the identification of persons with a disease (or other outcome variable) and a suitable control (comparison, reference) group of persons without the disease. Such a study is often referred to as retrospective because it starts after the onset of disease and looks back to the postulated causal factors.

cross-sectional study. A study that examines the relationship between disease and variables of interest as they exist in a population at a given time. A cross-sectional study measures the presence or absence of disease and other variables in each member of the study population. The data are analyzed to determine if there is a relationship between the existence of the variables and disease. Because cross-sectional studies examine only a particular moment in time, they reflect the prevalence (existence) rather than the incidence (rate) of disease and can offer only a limited view of the causal association between the variables and disease. Because exposures to toxic agents often change over time, cross-sectional studies are rarely used to assess the toxicity of exogenous agents.

ecological study. Also, demographic study. A study of the occurrence of disease based on data from populations, rather than from individuals. An ecological study searches for associations between the incidence of disease and suspected disease-causing agents in the studied populations. Researchers often conduct ecological studies by examining easily available health statistics, making these studies relatively inexpensive in comparison with studies that measure disease and exposure to agents on an individual basis.

Cohort Studies

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

Cohort studies are usually prospective and compare the incidence of a disease in exposed and unexposed groups.

A temporal ordering is present in the relationship between the agent and the onset of the disease, so it is possible to follow a cohort to assess whether the disease occurs after exposure to the agent (a necessary consideration for any causal interpretation).

Here, the independent variable is one of exposure/nonexposure; the dependent variable is disease condition (present/absent).

Within a familiar 2×2 contingency table framework, the relative risk, defined earlier, is the ratio of proportions for those having the disease within the exposed and unexposed groups.

Case-Control Studies

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

A case-control study is well exemplified by the smoking/lung cancer investigations.

Individuals who have a disease are first identified.

A “comparable” group without the disease is then constructed, and the exposure to the agent of interest compared for the “cases” versus the “controls.”

Retrospective case-control studies can generally be completed more quickly than a cohort study that requires tracking over time.

Thus, they are suited for the study of rare diseases that would require the recruitment of a prohibitive number of subjects for a comparable cohort study.

In a case-control context, disease status is now the independent variable and exposure is the dependent.

There is a comparison of the exposure of those with the disease (the “cases”) to those without the disease (the “controls”).

Unfortunately, a calculation of relative risk is no longer meaningful.

We give a quotation from the “Reference Guide on Epidemiology”:

A relative risk cannot be calculated for a case-control study, because a case-control study begins by examining a group of persons who already have the disease. That aspect of the study design prevents a researcher from determining the rate at which individuals develop the disease. Without a rate or incidence of disease, a researcher cannot calculate a relative risk.

The Wikipedia entry on case-control studies notes the difficulties that confounding creates in generating valid interpretations.

A memorable phrase is used for this:

“it is difficult, often impossible, to separate the chooser from the choice.”

The paragraph from Wikipedia that discusses problems with case-control studies, and which includes the phrase just noted is in your required reading.

Ecological Studies

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

An ecological study is carried out at the level of groups.

The overall rate of a disease in groups is compared to other differences that might be present for these same groups.

One has to be careful not to commit the ecological fallacy and attribute what might be present at a group level to what is also true at an individual level.

It would be necessary to follow up any group-level associations with a study at the individual level to make the type of individual level conclusions one would usually wish to have.

Encouragement Design

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

In addition to the four common types of observational study, we might add a fifth that is characterized by the general form of the intervention.

An *encouragement design* is one in which the active treatment is just the encouragement to do something, for example, take a drug to reduce blood pressure, change a diet to be more healthy, exercise regularly.

Instead of trying to evaluate an actual treatment regime when the compliance with it is unknown or difficult to control, the target of inference changes to one of encouragement having an effect or not.

Bias

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

Irrespective of the type of study adopted, it is important to consider how bias may affect the conclusions.

Bias refers to anything, other than sampling error, that results in a specious association, and which thereby compromises validity.

Generally, there are two broad categories of bias.

The first, *selection bias*, is about who gets into a study. Are there systematic differences in characteristics between those who get in and those who don't? Because of some reason that relates to what is being studied, people may be unwilling to be part of a study or they drop out after entering. Obviously, this affects the inferences that can be legitimately drawn.

The second category of bias is *information bias*, a flaw in the measurement of either or both the exposure and the disease.

As part of this, there is *recall bias* where people with the disease may remember past exposures better. This bias may be particularly problematic in a case-control study where the control group may not remember their exposures correctly.

In general, *misclassification bias* is when people are misclassified according to exposure or disease status. For disease, the diagnostic criteria should be good.

Some Definitions

Terms commonly used in epidemiology: These are grouped below according to the chapters of the *Reference Manual on Scientific Evidence* in which they appeared.

“Reference Guide on Epidemiology”:

agent: Also, risk factor. A factor, such as a drug, microorganism, chemical substance, or form of radiation, whose presence or absence can result in the occurrence of a disease. A disease may be caused by a single agent or a number of independent alternative agents, or the combined presence of a complex of two or more factors may be necessary for the development of the disease.

bias: Any effect at any stage of investigation or inference tending to produce results that depart systematically from the true values. In epidemiology, the term bias does not necessarily carry an imputation of prejudice or other subjective factor, such as the experimenter's desire for a particular outcome. This differs from conventional usage, in which bias refers to a partisan point of view.

pathognomonic: An agent is pathognomonic when it must be present for a disease to occur. Thus, asbestos is a pathognomonic agent for asbestosis. See signature disease.

secular-trend study: Also, time-line study. A study that examines changes over a period of time, generally years or decades. Examples include the decline of tuberculosis mortality and the rise, followed by a decline, in coronary heart disease mortality in the United States in the past fifty years.

signature disease. A disease that is associated uniquely with exposure to an agent (for example, asbestosis and exposure to asbestos). See pathognomonic.

teratogen: An agent that produces abnormalities in the embryo or fetus by disturbing maternal health or by acting directly on the fetus *in utero*.

differential misclassification: A form of bias that is due to the misclassification of individuals or a variable of interest when the misclassification varies among study groups. This type of bias occurs when, for example, individuals in a study are incorrectly determined to be unexposed to the agent being studied when in fact they are exposed.

etiology: The cause of disease or other outcome of interest. [The phrase “etiology unknown” simply means that it is of an unknown cause.]

misclassification bias. The erroneous classification of an individual in a study as exposed to the agent when the individual was not, or incorrectly classifying a study individual with regard to disease. Misclassification bias may exist in all study groups (nondifferential misclassification) or may vary among groups (differential misclassification).

“Reference Guide on Toxicology” (Bernard D. Goldstein & Mary Sue Henifin):

epigenetic: Pertaining to nongenetic mechanisms by which certain agents cause diseases, such as cancer.

mutagen: A substance that causes physical changes in chromosomes or biochemical changes in genes.

“Reference Guide on Medical Testimony” (Mary Sue Henifin, Howard M. Kipen, & Susan R. Poulter):

differential diagnosis: The term used by physicians to refer to the process of determining which of two or more diseases with similar symptoms and signs the patient is suffering from, by means of comparing the various competing diagnostic hypotheses with the clinical findings. [The aim is to identify the disease to determine the treatment.]

differential etiology: A term used on occasion by expert witnesses or courts to describe the investigation and reasoning that leads to a determination of external causation, sometimes more specifically described by the witness or court as a process of identifying external causes by a process of elimination. [Here, the goal is to identify the cause(s) of the disease but not to determine treatment.]

pathogenesis. The mode of origin or development of any disease or morbid process.

From the Cochrane Collaboration glossary:

phase I, II, III, IV trials: A series of levels of trials required of drugs before (and after) they are routinely used in clinical practice: Phase I trials assess toxic effects on humans (not many people participate in them, and usually without controls); Phase II trials assess therapeutic benefit (usually involving a few hundred people, usually with controls, but not always); Phase III trials compare the new treatment against standard (or placebo) treatment (usually a full randomized controlled trial). At this point, a drug can be approved for community use. Phase IV monitors a new treatment in the community, often to evaluate long-term safety and effectiveness.

idiopathic: of unknown cause. Any disease that is of uncertain or unknown origin may be termed “idiopathic.”

Observational Studies: Additional Cautions

A particularly influential but problematic meta-analysis appeared in 1991 by Meir Stampfer and Graham Colditz, entitled “Estrogen Replacement Therapy and Coronary Heart Disease: A Quantitative Assessment of the Epidemiologic Evidence” (*Preventive Medicine*, 20, 47–63).

A reprinting of this article occurred in 2004 (*International Journal of Epidemiology*) along with a collection of essays that tried to deal with the discrepancy between the results of this meta-analysis and subsequent clinical trials.

One particularly good commentary was from Diana Petitti, “Hormone Replacement Therapy and Coronary Heart Disease: Four Lessons” (*International Journal of Epidemiology*, 33, 461–463).

Paraphrasing, the four lessons referred to in this article are:

(1) Do not turn a blind eye to contradiction and ignore contradictory evidence; instead, try to understand the reasons behind the contradictions.

At the time of the Stampfer–Colditz (1991) meta-analysis, there were data available from two major sources that contradicted the meta-analysis—The Coronary Drug Project and many studies on the use of oral contraceptives—which would be difficult to reconcile with the Stampfer–Colditz conclusions.

(2) Do not be seduced by mechanism.

Even where a plausible mechanism exists, do not assume that we know everything about that mechanism and how it might interact with other factors.

Apparently, Stampfer and Colditz (1991) were so seduced:
“This benefit is consistent with the effect of estrogens on lipoprotein subfractions (decreasing low-density lipoprotein levels and elevating high-density lipoprotein levels).”

This is good advice more generally.

Just because one can conjure up a reason for explaining why a particular result may have been observed, doesn't automatically mean that it is therefore true.

For example, Caspi and colleagues (2003) identified a particular serotonin gene as possibly being related to depression, and we know that the common antidepressants all act on serotonin as re-uptake inhibitors.

So, here is a reasonable mechanism to account for the relationship they apparently saw.

The failure to replicate the Caspi results mentioned in Chapter 16, should be read against the seduction of an unproven serotonin mechanism.

(3) Suspend belief. In commenting on researchers defending observational studies, Pettiti (2004) notes:

“belief caused them to be unstrenuous in considering confounding as an explanation for the studies”. Don’t be seduced by your desire to prove your case. Stampfer and Colditz ignored the well-known (in 1991) effects of social class, education, and socioeconomic status on coronary heart disease.

We are reminded of a quip generally attributed to George Box: “Don’t emulate Pygmalion and fall in love with your model.”

Box provided a more detailed version of this sentiment in his R. A. Fisher Memorial Lecture::

The good scientist must have the flexibility and courage to seek out, recognize, and exploit such errors—especially his own. In particular, using Bacon's analogy, he must not be like Pygmalion and fall in love with his model.

(4) Maintain scepticism. Question whether the factor under investigation can really be that important; consider what other differences might characterize the case and control groups.

Be wary of extrapolating results beyond the limits of reasonable certainty (for example, with grandiose forecasts of number of “lives saved”).

As a general admonishment that might be best to keep in mind, remember the *law of unintended consequences*;

any intervention in a complex system invariably creates unanticipated and often undesirable outcomes.

Controlled Studies

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

The two most common controlled studies are named (for purposes of PubMed publication), a “randomized controlled trial” and a “controlled clinical trial.”

We give definitions from the National Library of Medicine:

Controlled Clinical Trial: A Controlled Clinical Trial (CCT) is a work consisting of a clinical trial involving one or more test treatments, at least one control treatment, specified outcome measures for evaluating the studied intervention, and a bias-free method for assigning patients to the test treatment. The treatment may be drugs, devices, or procedures studied for diagnostic, therapeutic, or prophylactic effectiveness.

Randomized Controlled Trial: A Randomized Controlled Trial (RCT) is a work consisting of a clinical trial that involves at least one test treatment and one control treatment, concurrent enrollment and followup of the test- and control-treated groups, and in which the treatments to be administered are selected by a random process, such as the use of a random-numbers table.

An RCT is usually considered the gold standard for providing valid evidence of a (causal) effect in whatever area it is used (with a CCT coming in at a more distant second).

These controlled studies, however, are subject to some of the same interpretive anomalies that plague observational studies.

First, in considering how subjects are recruited, there is a need to follow whatever rules of informed consent are imposed by the institution overseeing the study.

This implies that subjects must be told about the possible downsides of what may be administered, that they might be allocated to, say, a placebo or alternative condition, and they can terminate their participation at any time they might wish.

Generally, informed consent is a voluntary and documented confirmation of a subject's willingness to participate in a trial, and before any protocol-related procedures or treatments are performed.

The process of obtaining informed consent gives a potential subject the chance to just say “no” to entering the trial at the outset, or permission to stop participation at any time.

Based on the difficulties seen in observational studies, it is clear that individuals who opt to begin participation and/or who continue are basically different from those who don't start or those who drop out.

Intention-to-treat Analyses

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

Once the data from a controlled study are available, it is tempting to engage in a process of data dredging (the older version of modern data mining), to see if various effects can be teased apart for subgroups.

It deserves reminding that sample size issues and the culling of chance occurrences must always be accounted for.

A particularly contentious part of this process arises when an “intention-to-treat” analysis is performed (alternatively labeled as “analyze as randomized” or “as randomized, so analyzed”).

An intention-to-treat analysis in an RCT uses *all* patients randomly assigned to the treatments, irrespective of whether they completed or even received the designated or intended treatment(s).

Two terms defined in the Cochrane Collaboration Glossary (2005) are useful when discussing issues raised by an “intention-to-treat” analysis:

per protocol analysis: An analysis of the subset of participants from a randomized controlled trial who complied with the protocol sufficiently to ensure that their data would be likely to exhibit the effect of treatment. This subset may be defined after considering exposure to treatment, availability of measurements and absence of major protocol violations. The per protocol analysis strategy may be subject to bias as the reasons for noncompliance may be related to treatment.

performance bias: Systematic differences between intervention groups in care provided apart from the intervention being evaluated. For example, if participants know they are in the control group, they may be more likely to use other forms of care. If care providers are aware of the group a particular participant is in, they might act differently. Blinding of study participants (both the recipients and providers of care) is used to protect against performance bias.

Controlled Studies: Additional Sources of Bias

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

The *Cochrane Handbook for Systematic Reviews of Interventions* includes a chapter entitled “Assessing Risk of Bias in Included Studies.”

Although directed toward an eventual meta-analysis, it includes a valuable discussion of bias sources that should be assessed in considering just a single controlled study.

We begin with a classification scheme into five broad areas of bias:

Selection bias: Systematic differences between the baseline characteristics of the groups to be compared;

Performance bias: Systematic differences between groups in the care provided, or in exposure to factors other than the interventions of interest;

Attrition bias: Systematic differences between groups in the number of withdrawals from a study;

Detection bias: Systematic differences between groups in how outcomes are determined;

Reporting bias: Systematic differences between reported and unreported findings.

Several additional terms introduced in this *Cochrane Handbook* chapter pertain to the classification of bias just given:

sequence generation and *allocation concealment* concern selection bias;

blinding is relevant to performance, attrition, and detection bias;

incomplete outcome data is pertinent to attrition bias;

and *selective reporting* is obviously connected to reporting bias.

Brief characterizations follow:

Sequence generation refers to the rule for subject allocation to treatment by some chance mechanism;

Allocation concealment concerns the steps taken to insure implementation of the sequence generation by preventing prior knowledge of the ensuing allocations;

Blinding reduces the risk of study participants or personnel knowing the intervention received;

Incomplete outcome data points to possibly biased outcomes because of attrition or study participant exclusion;

Selective reporting is the (generally unethical) censoring of data on study outcomes.

The Special Case of Medical Trials

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

Controlled studies carried out to obtain FDA (Food and Drug Administration) approval for some medical intervention, such as a drug, implant, or other device, have their own set of biases, with some unique to medical trials carried out to seek regulatory approvals.

Obviously, all biases should be of concern in the interpretation of clinical trial data and how these might skew the outcome of a regulatory argument.

Even though a study may begin as a well-designed randomized controlled trial, because of differential dropout and other forms of experimental mortality, that is not where one usually ends up.

It is important to know from where and how the subjects were recruited.

For foreign trials, in particular, the kind of inducements offered should be known, and then, exactly who was so enticed to participate.

Follow-up information should be available on the characteristics of the subject population (for example, general health and nutrition, age, sex, social class, education), and whether these may interact with the medical procedure being assessed to slant the results in particular ways.

For nonforeign trials, the bonuses doctors receive for recruiting to the clinical trial could be seen as problematic inducements that might lead to registering patients who are not really eligible for the trial.

The general question is always the following:

were the correct patients enrolled so the trial provides information on safety and effectiveness directly relevant for the target group expected to receive the medical intervention?

A second central question would be a careful characterization of the treatments administered, both as to type and dosage.

For instance, if a “me-too” drug, defined as one similar to others already available, is being evaluated, did the comparison involve an inert placebo or an inappropriate dose level of a competitor?

Also, was the length of the trial too short to show the adverse events that might be apparent only in Phase IV postmarketing monitoring (think of Vioxx, Avandia, and the host of other withdrawn products)?

Are all the data from the trial reported and not just selected portions that demonstrate what the petitioners wish to put forward?

Were treatments compared to alternatives no longer under patent, and was the new treatment substantially better than generic alternatives?

A third question involves what is being measured.

Many times only surrogate outcomes are available (for example, lowered cholesterol), because the relevant clinical outcome is not available (for example, death from heart failure).

Has the definition been changed, possibly arbitrarily, for the condition being treated—for obesity, hypertension, high cholesterol levels—and what effect does this have on the surrogate endpoint being assessed?

Also, what is the strength of connection to the ultimate clinical endpoint?

Is the treatment and trial for an ethically dubious disorder?

Strong effects seem to be obtained more often when drug companies conduct the trials involving patentable products.

Are all trials being reported and registered before they begin?

Are the trials being conducted outside the reach of the Declaration of Helsinki?

Do the sponsors have complete control over what gets reported in the open literature?

What is the fate of private Contract Research Organizations (CROs) running clinical trials that don't obtain the results a sponsor would like?

Marcia Angell, a physician and the first woman to serve as editor-in-chief of the *New England Journal of Medicine*, published a book in 2004 about how Big Pharma operates;

it has the self-explanatory title, *The Truth About the Drug Companies: How They Deceive Us and What to Do about It*.

In commenting on the conduct of medical trials, Angell argues that drug companies should not be allowed to control the testing of their own medical products.

Moreover, clinical trial data should be the joint property of, say, NIH and the researchers who carried out the trials, and not in any way under the control of the sponsoring drug company.

We give a short quotation from Angell's book that summarizes this position:

To ensure that clinical trials serve a genuine medical need and to see that they are properly designed, conducted, and reported, I propose that an Institute for Prescription Drug Trials be established within the National Institutes of Health to administer clinical trials of prescription drugs.

In an afterword to her book, Angell gives a few suggestions about what individuals can do to protect their interests when it comes to the pharmaceutical industry.

First, when your doctor prescribes a new drug, ask for evidence that this is better than alternative treatments.

Also, has the evidence been published in a peer-reviewed journal, or is it just from drug company representatives?

To our members of Congress, ask about financial ties to the pharmaceutical industry.

And finally, ignore all direct-to-consumer (DTC) drug advertising. These ads are meant to sell drugs and not to educate consumers in any altruistic manner.

An Introductory Oddity: The 1954 Salk Polio Vaccine Trials

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

The 1954 Salk polio vaccine trials was the biggest public health experiment ever conducted.

One field trial, labeled an observed control experiment, was carried out by the National Foundation for Infantile Paralysis.

It involved the vaccination, with parental consent, of second graders at selected schools in selected parts of the country.

A control group would be the first and third graders at these same schools, and indirectly those second graders for whom parental consent was not obtained.

The rates for polio contraction (per 100,000) are given below for the three groups:

Grade 2 (Vaccine): 25/100,000;
Grade 2 (No consent): 44/100,000;
Grades 1 and 3 (Controls): 54/100,000.

The interesting observation we will return to below is that the Grade 2 (No consent) group is between the other two in the probability of polio contraction.

Counterintuitively, the refusal to give consent seems to be partially protective.

The second field trial was a (double-blind) randomized controlled experiment.

A sample of children were chosen, all of whose parents consented to vaccination.

The sample was randomly divided into two, with half receiving the Salk vaccine and the other half a placebo of inert salt water.

There is a third group formed from those children with no parental consent and who therefore were not vaccinated.

We give the rates of polio contraction (per 100,000) for the three groups:

Vaccinated: 28/100,000;

Control: 71/100,000;

No consent: 46/100,000.

Again, not giving consent appears to confer some type of immunity; the probability for contracting polio for the “no consent” group is between the other two.

The seeming oddity in the ordering of probabilities, where “no consent” seems to confer some advantage, is commonly explained by two “facts”:

(a) children from higher-income families are more vulnerable to polio; children raised in less hygienic surroundings tend to contract mild polio and immunity early in childhood while still under protection from their mother's antibodies;

(b) parental consent to vaccination appears to increase as a function of education and income, where the better-off parents are much more likely to give consent.

The “no consent” groups appear to have more natural immunity to polio than children from the better-off families.

This may be one of the only situations we know of where children growing up in more resource-constrained contexts are conferred some type of advantage.

The Randomized Response Method

Background:
Experimental
Design and
the Collection
of Data

Psychology
(Statistics)
484

As noted earlier, how questions are framed and the context in which they are asked are crucial for understanding the meaning of the given responses.

This is true both in matters of opinion polling and for collecting data on, say, the health practices of subjects.

In these situations, the questions asked are usually not sensitive, and when framed correctly, honest answers are expected.

For more sensitive questions about illegal behavior, (reprehensible) personal habits, suspect health-related behaviors, questionable attitudes, and so on, asking a question outright may not garner a truthful answer.

The randomized response method is one mechanism for obtaining “accurate” data for a sensitive matter at a group level (but not at the individual level).

It was first proposed in 1965 by Stanley Warner in the *Journal of the American Statistical Association*, “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias” (60, 63–69).

A modified strategy was proposed by Bernard Greenberg and colleagues in 1969, again in *JASA*: “The Unrelated Question Randomized Response Model: Theoretical Framework” (64, 520–539).

We illustrate Warner’s method and then Greenberg’s with an example in the required reading.