

# Biostatistics and Research

Theresa A Scott, MS

Vanderbilt University  
Department of Biostatistics  
theresa.scott@vanderbilt.edu  
<http://biostat.mc.vanderbilt.edu/TheresaScott>

## Introduction: Clinical Research

- ▷ Includes translational research, clinical trials, patient-oriented research, epidemiologic studies, behavioral science, and health services research.
- ▷ Can be thought of from two viewpoints:
  - 1 The **anatomy** – what it's made of.
    - Research question(s);
    - Background and significance;
    - Study design and subjects; and
    - Measurements.
  - 2 The **physiology** – how it works.
    - Study findings used to draw inferences about what happened to the people in the study and about how they generalize to people outside the study.
- ▷ Biostatistics critically involved in both.

# Anatomy of Research

## ▷ **Research question(s):**

- Most often worded as objective(s)/specific aim(s).
- Should pass the 'So what?' test.
- Should be 'FINER': Feasible, Interesting, Novel, Ethical, Relevant.

## ▷ **Background and significance:** tries to answer 'Why is/are the research question(s) important?' by answering

- What is known about the topic at hand?
- What uncertainties remain?
- What kind of answers will the study provide?
- How will the findings help resolve the uncertainties?
- How will the findings lead to new scientific knowledge (and influence current practice)?

# Anatomy of Research, *cont'd*

## ▷ **Study design:** (inherently related to specific aim(s)).

- Descriptive or analytic (associations/causation)?
- Retrospective or prospective?
- Observational or interventional (ie, clinical trial)?
- If observational, cohort, cross-sectional, or case-control?
- Study of diagnostic (ie, medical) tests?
- Study using a nationally available database (eg, NAMCS, NIS)?

## ▷ **Study subjects:** (define the *population* of interest)

- 1 *Inclusion criteria* – demographics and clinical characteristics of subjects well suited to answer the research question(s); and
- 2 *Exclusion criteria* – eliminates subjects who are unethical or inappropriate to study (eg, not related to question of interest).

## Sidebar – Population and sample

- ▷ All clinical research is based, philosophically and practically, on the use of a *sample* to represent a *population* of interest.
  - *Population* – complete set of people with a specified set of characteristics (clinical, demographic (including geographic), and temporal).
    - Impractical/impossible to collect a whole population.
  - *Sample* – representative subset of the population.
  - Desire a *random* sample – a subset where every item in the population has the same probability of being in the sample.
  - In the end, left with the *study* sample – subset who actually participate in study.
  - Key to being able to generalize your findings to the population – how representative your study sample is of the population.

## Anatomy of Research, *cont'd*

- ▷ **Measurements** – *roles* of the variables collected:
  - *Outcome* – variable that is the focus of the study, whose variation or occurrence you are seeking to understand.
  - *Predictor* – variable that may influence the size or occurrence of the outcome (aka exposure variable, risk factor).
    - In a clinical trial, includes the intervention.
  - *Confounder* – variable that differs between values of the predictor variable and which also affects the outcome.
    - Needed to correctly estimate the relation between the outcome(s) and predictor(s).
  - Often more than one outcome, predictor, and confounder.
  - Additionally, variables to describe the study sample.

# Anatomy of Research, *cont'd*

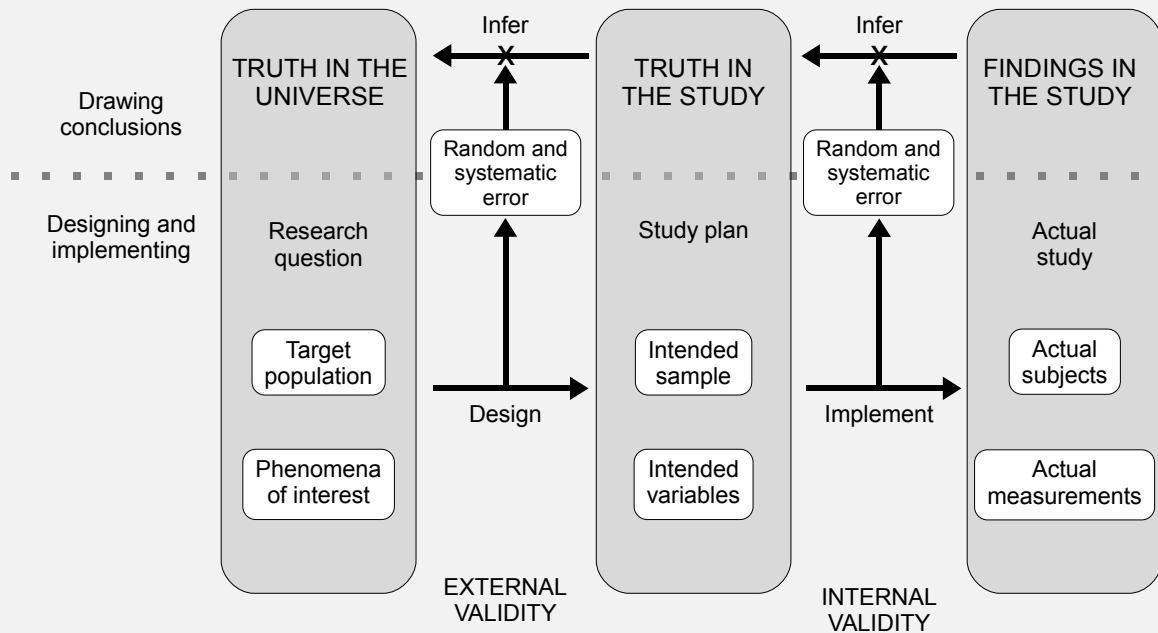
- ▷ **Measurements, *cont'd*** – types of variables collected:
  - *Continuous (quantitative data)*:
    - Have any number of possible values (eg, weight).
    - *Discrete numeric* – set of possible values is a finite (ordered) sequence of numbers (eg, pain scale of 1, 2, . . . , 10).
  - *Categorical (qualitative data)*:
    - Have only certain possible values (eg, race); often not numeric.
    - *Binary (dichotomous)* – a categorical variable with only two possible value (eg, gender).
    - *Ordinal* – a categorical variable for which there is a definite ordering of the categories (eg, severity of lower back pain as none, mild, moderate, and severe).

# Physiology of Research

- ▷ *Goal* – to draw inferences from findings in the study about the nature of the universe around it.
- ▷ When planning the study, goal is to maximize the *validity* of the inferences at the end of the study.
  - *Internal validity* – degree to which the investigator draws the correct conclusion about what actually happened in the study.
  - *External validity (also called generalizability)* – degree to which these conclusions can be appropriately applied to people and events outside the study.
- ▷ *Challenge* – to design and implement a study plan with adequate control over two major threats to these inferences:
  - *Random error* (wrong results due to chance).
  - *Systematic error* (wrong results due to bias).

# Physiology of Research, *cont'd*

▷ Pictorial summary of how research works:



## Biostatistics

▷ Application of *statistics* to a wide range of topics in *biology* (eg, medicine).

- Sometimes referred to as biometry or biometrics.

▷ Science of/set of tools for

- the collection, summarization, and analysis of data from biological research studies; and
- the presentation, explanation/interpretation of, and *inference* from, the results.

▷ Where in the timeline of research can you benefit from working with a biostatistician?

- “From inception to publication and beyond”.
- **IMPORTANT:** Get a biostatistician involved *as soon as possible*.

## Biostatistics, *cont'd*

- ▷ Why is it important to understand (bio)statistics?
  - *Understanding the literature*
    - Should be able to critique the science & the statistics of a study.
  - *Designing research studies*
    - Much easier to plan for everything before the study is conducted than to try to fix things once the study is finished.
  - *Producing manuscripts*
    - Reviewers very critical of whether the statistical analyses used were appropriate, what conclusions were drawn, and what inferences were made.
  - *Acquiring grants*
    - Reviewers very critical of study design, sample size calculation/justification, and proposed statistical analyses.

## Biostatistics, *cont'd*

- ▷ What are some of the ways a biostatistician can help you?
  - Research question(s):
    - Clarify the specific aim(s)/objective(s) and word them in such a way as to convey the statistical analysis used/to be used.
  - Background and significance:
    - Provide another set of eyes in order to ensure the 'So what?' factor is justified and understandable to reviewers (ie, those not familiar with the area of research/topic at hand).
  - Study design:
    - Determine the most appropriate study design, maximizing the efficiency and increasing the interpretability and generalizability in the process.
    - Ensure there are no flaws in the design that will affect the statistics and the ability to find anything or make inferences.

## Biostatistics, *cont'd*

- ▷ What are some of the ways a biostatistician can help you?, *cont'd*
  - Study design, *cont'd*:
    - Use modern experimental designs that allow for early termination of experiments when sufficient evidence for effects is demonstrated, or extends experiments when results are equivocal.
  - Study subjects:
    - Calculate and justify the sample size/power, incorporating any hypotheses, budgetary constraints, recruiting limitations, lost to follow-up, statistical needs, etc.
  - Measurements:
    - Assist with the selection/refining of the variables to collect, ensuring they relate to the study objective(s) and conform to the proposed statistical analysis.

## Biostatistics, *cont'd*

- ▷ What are some of the ways a biostatistician can help you?, *cont'd*
  - Measurements, *cont'd*:
    - Assist with the design and development of methods, operations and procedures to collect, edit, verify, and manage data, including the design of questionnaires and other data collection instruments.
  - In general:
    - Perform the planned statistical analysis, including data verification and data correction as necessary, using appropriate statistical analysis software and ensuring reproducible research.
    - Interpret and communicate the results in lay terms, explaining all statistical methods used.

▷ What are some of the ways a biostatistician can help you?, *cont'd*

■ In general, *cont'd*:

- Assist with the development of IRB and grant proposals, increasing their likelihood of being funded – includes providing formal write-up of sample size/power calculation/justification and statistical analysis plan, ensuring each specific aim/objective is addressed.
- Assist with the writing and revision of abstracts, posters, slide presentations, and manuscripts for publication, including providing statistical methods and results sections, figures and tables – increasing the likelihood of acceptance and improving the quality of the result.

## Recommended Reading

- *Designing Clinical Research*
  - 3rd edition, by Stephen B Hulley, et al
- *Essential Medical Statistics*
  - 2nd edition, by Betty R Kirkwood and Jonathan AC Sterne
- “*Statistics Notes in the British Medical Journal*”
  - by Doug Altman and Martin Bland
  - <http://www-users.york.ac.uk/~mb55/pubs/pbstnote.htm>
  - Particularly, *Absence of evidence is not evidence of absence*