WHAT IS BIOMEDICAL STATISTICS?

(OR, WHAT IS BIOSTATISTICS?)

Marie Davidian

NC State University
DEPARTMENT OF STATISTICS

davidian@stat.ncsu.edu

http://www.stat.ncsu.edu/~davidian
What is Biostatistics?

Outline

1. Biomedical and public health research
2. Why is studying humans is so complicated?
3. Example 1: Comparing treatments based on a time-to-event
4. Example 2: Longitudinal studies with repeated measures
5. Example 3: Should this study be stopped?
6. Example 4: Pharmacokinetics
7. Biostatistical research
8. Where do biostatisticians work?
9. Training to be a biostatistician
1. Biomedical and public health research

Research related to the betterment of human health:

- An enormous enterprise in the US, involving *billions* of dollars
- Funded mostly by the *federal government* and the *pharmaceutical industry*
**Broad goal:** To acquire new knowledge to help prevent, detect, diagnose, and treat *disease*

- Cancer, heart disease, diabetes, HIV/AIDS, mental illness to the common cold, flu, athlete’s foot!
- Develop new drugs and other therapies to *treat* disease
- Uncover factors and behaviors that are associated with *developing* disease and thus how to *prevent* disease
- Understand the *manifestation* and *course* of disease
- Develop new devices and procedures for *detecting* and *monitoring* disease
What is Biostatistics?

**National Institutes of Health (NIH):** The government agency responsible for coordinating most health-related research funded by the US Government

- Budget of over $23 billion in 2002
- Mainly for *grants* to researchers in universities, research institutes, hospitals to carry out studies addressing these issues
- Research on *all aspects*: prevention, detection, diagnosis, and treatment

**Pharmaceutical and biotechnology industries:**

- Focused mostly on *developing new therapies* for prevention and treatment
- *Huge* investment in money, effort, and people
Scope of research:

- *Laboratory*: identifying compounds, understanding workings of cells, etc.
- *Animal studies*: use animals as *models* for humans
- *Human studies*: *experimental* or *observational*
What is Biostatistics?

**Experimental biomedical studies: Clinical trials**

- Evaluation of treatments in a *controlled* setting

- Is a new headache medication safe? Does a new anti-fungal medication cure athlete’s foot? Does a new drug for treating HIV-infected patients work better than an older one? Does hormonal therapy reduce the risk of breast cancer? . . .

- Studies in which the participants receive treatments according to a *controlled plan*
In the news recently: The *Women’s Health Initiative* (WHI)

- A 15-year research program funded by the *National Heart, Lung, and Blood Institute* (NHLBI)

- *Estrogen+Progestin study*: A *clinical trial* to test the risks and benefits of using this therapy versus not

- 16,608 women ages 50 to 79 years were *randomly assigned* to take E+P or a *placebo* and monitored periodically for disease

- *Analysis*: Compare rates of disease in the 2 groups

- *Stopped* due to an increased risk of *invasive breast cancer* for E+P
Public health research: Epidemiology

- Study of the *causes and origins* of disease
- Second-hand smoke and lung cancer, Air pollution and respiratory illness, Diet and heart disease, ... 

- *Common approach*: Identify a group of people and *observe* their disease status and other factors

- Look for *associations* between factors and disease that may reflect *risk*

- *Often*: Follow *over time*
Often in the news: The *Framingham Heart Study*

- Started in 1948, funded by NHLBI
- *Cohort* of 5,209 men and women ages 30–62 in Framingham MA recruited and given extensive *physical exams, lifestyle interviews*
- Return every 2 years for *medical history, physical exam, and lab tests*
- In 1971, the study enrolled a *2nd-generation cohort*: 5,124 of the original participants’ adult children and spouses
- *Objective*: Understand *causes* and *risk factors* associated with *cardiovascular disease* (CVD)
- Identification of major CVD risk factors, including high blood pressure, high blood cholesterol, smoking, inactivity, . . .
- Currently recruiting a *3rd-generation cohort* of 3,500 grandchildren to facilitate study of *genetic factors*
- Evaluation of new *diagnostic technologies* like echocardiography
Little-known fact I: *Statisticians* have played and continue to play a *key role* in almost all of this!

- *Design* of studies
- *Analysis* of data
- Basis for *regulatory* requirements of the *US FDA*

Little-known fact II: Interpreting the results of such studies requires *specialized statistical methods*

- ...because of the special *challenges* involved in studying humans!

**Biomedical statistics or biostatistics:** *The field of application of statistics to address problems in biomedical and public health research*
2. Why is studying humans so complicated?

The “ideal” agricultural experiment: Compare yields of wheat plants for 2 different fertilizers

- Plant wheat plants in plots in a field
- Assign plots to receive Fertilizer A or B at random
- Apply assigned fertilizers to plots
- Measure wheat yield for each plot at the end of the growing season
- Compare mean wheat yield for A vs. B by a t-test
**Features:**

- Plants actually *show up* to be planted in the plots
- The experimenter can ensure that the fertilizer *actually applied* to a plot is the *assigned* fertilizer recorded for that plot
- The plants *stick around* through the whole study and are there to be *measured* at the end

**Result:** *Data* intended to be collected

- Are *all* available
- And thus may be *analyzed* according to plan
What is Biostatistics?

The “ideal” human study: Compare *reduction in blood pressure* for two different *anti-hypertensive drugs*

- Subjects are recruited to take part in the study and asked to show up for *baseline* blood pressure reading, recording of *demographic information*, and a *month’s supply* of assigned pills with dosing instructions.

- *All* recruited subjects show up for *baseline* reading, *provide all requested information*, and *take home* supply of pills.

- *All* subjects take their assigned pills *exactly according to the instructions*.

- *All* subjects return as directed for *monthly* visits to the clinic to *receive more pills* and have *blood pressure reading* until the end of the study.

What’s wrong with this picture?
**Problem:** Humans are *notorious* for *not doing as they are told*!

- A subject may agree to participate and then *change his mind*
- A subject may show up for the initial visit but then *change her mind*
- A subject may show up for the initial visit but *refuse to provide* needed information or provide *incorrect* information
- A subject may not take the pills as *directed* (or *not all*)
- A subject may *miss* some monthly visits (e.g. vacation)
- A subject's physician may recommend she *discontinue* participation in the study
- A subject may *move away* or be hit by a *bus*
- A subject may *die* of a hypertension-related ailment
Result: Data intended to be collected on humans routinely

- May be missing or incomplete
- May be error-prone
- May not reflect what was intended because subjects may not comply with their treatment

Implications:

- It is unlikely that the data may be analyzed according to the initial plan
- Specialized analysis methods are required to take into account these “human” problems
- In fact, designs for studies that anticipate these problems are needed

Biostatistics...
3. Comparing treatments based on a time-to-event

Evaluation and comparison of treatments in clinical trials:

- *Randomization* of a fixed number subjects to treatments
- For each subject, observe a *response*
- Evaluate or compare on the basis of some summary of the response, like the *mean*
- The response is called the *primary endpoint*
For chronic diseases: Cancer, HIV, diabetes,…

- The primary endpoint on which evaluation is based is a time until an event occurs
- “Survival”
- Time from initiation of treatment to death for lung cancer patients
- Time from initiation of treatment until an AIDS-defining event for HIV-infected patients
- Time from initiation of treatment until myocardial infarction or death for patients with CVD

- Objectives: Summarize and compare the distributions of times-to-event for each treatment
Further complication:

- The study occurs over a *finite period of time* (e.g. 3–5 years)
- Logistically, subjects in a clinical trial cannot be recruited *all at once*
- Instead, subjects are recruited as they become *available*
- Some subjects enter the study *early*
- Others enter when the study is nearing an *end*
- “*Staggered entry*”
What is Biostatistics?

Study starts

Study ends

Time 0

Calendar Time

Patient Time
**Result:** Censoring

- For some subjects, regardless of *time of entry*, we do not observe the endpoint due to *competing causes*, e.g. hit by a bus
- For some subjects entering the study *late*, the study may *end* before the endpoint happens
- The endpoint is *censored* for these subjects

**Analysis:**

- *Ignoring* the censored data (throw those subjects out) leads to *inaccurate* picture of survival
- *Needed:* Methods for *fair* summary and comparison that take *censoring* into account
- *Survival Analysis*
4. Longitudinal studies with repeated measures

Often: Questions of interest have to do with how something changes over time

- Diet medications and weight loss
- Cholesterol-lowering drugs
- Anti-hypertensive medications
- HIV viral load following antiretroviral therapy
- Prostate-specific antigen for monitoring prostate cancer recurrence

How does it change? Rate of change? Different for different treatments?

A BIG problem: Subjects drop out of the study
What is Biostatistics?

Dental Study Data

age (years)

distance (mm)
What is Biostatistics?

Subject 1

Days

log10 Plasma RNA (copies/ml)

Subject 2

Days

log10 Plasma RNA (copies/ml)

Subject 3

Days

log10 Plasma RNA (copies/ml)

Subject 4

Days

log10 Plasma RNA (copies/ml)
What is Biostatistics?

**Placebo Group**

**Treatment Group**

The graphs illustrate the response over time for the Placebo Group and the Treatment Group. The x-axis represents time, and the y-axis represents the response. The Placebo Group shows a more scattered pattern, while the Treatment Group appears to have a more consistent trend with higher values.
Needed:

- A *statistical model* in which the question can be formally stated
- With analysis methods to take into account that the data are *repeated measurements* on the *same subjects*
- ... And that subjects tend to *drop out*
- *Longitudinal data analysis*
5. Should this study be stopped?

Recall: The WHI estrogen+progestin trial

- Was stopped due to increased breast cancer risk
- How was this decision made?
- Data Safety Monitoring Board: includes physicians, ethicists, patient advocates, and statisticians (DSMB)
- DSMB meets periodically, reviews data available so far, makes a decision whether to continue or stop
- Today, it is routine (and sometimes required) for a clinical trial to have a DSMB

How is the decision made?
What is Biostatistics?

Why stop?

- Serious *adverse events*
- Overwhelming *superiority* of one treatment
- *No apparent difference* between treatments
- Logistical or design problems *too difficult* to fix
- Clinical trials are *expensive*

**How to stop?** Need a *formal set of rules* for stopping *early*
**Ordinarily:** Do a study, collect *all* data, analyze at the end

- Single *hypothesis test* at level of significance $\alpha$

**Here:** Look at data *sequentially*

- DSMB meets, looks at *partial* data so far
- *Interim* hypothesis test to look at the difference based on partial data
- If trial is *never* stopped, want level of significance $\alpha$ at the end
- *How large* does the test statistic need to be at an “*interim look*” (all data not available) so that at the end, the chance we would make a mistake is $\leq \alpha$?
- *Group sequential statistical methods* for design and analysis
6. Pharmacokinetics

How does a drug work? From a pharmacologist’s point of view

- Achieve therapeutic objective
- Minimize problems
- Come up with dosing regimen to do this
- How much? How often? Under what conditions?
- Need to understand “what the body does to the drug”
What is Biostatistics?

- Dose
- Concentration
- Response

$PK$, $PD$
What is Biostatistics?

What goes on inside: *Absorption, Distribution, Metabolism, Elimination*

Routes of drug administration: *Intravenously, Intramuscularly, Subcutaneously, Orally, ...*
Data for 4 subjects given same oral dose of anti-asthmatic theophylline:
One-compartment model with first-order absorption, elimination:
oral dose $D$

\[
\begin{align*}
\frac{dX(t)}{dt} &= k_a X_a(t) - k_e X(t), \quad X(0) = 0 \\
\frac{dX_a(t)}{dt} &= -k_a X_a(t), \quad X_a(0) = FX(0)
\end{align*}
\]

$X(t)$ = amount in “blood” compartment, $X_a(t)$ = amount at absorption site

\[
C(t) = \frac{X(t)}{V} = \frac{k_a DF}{V(k_a - k_e)} \{\exp(-k_e t) - \exp(-k_a t)\},
\]

$V$ = “volume” of compartment
Pharmacokinetics: Learn about *pharmacokinetic parameters* like $k_a$, $k_e$, $V$ that *govern* how concentrations arise

- For *individual* subjects
- And how *variable* they are across subjects

**Needed:**

- A *statistical model* that allows each subject to have his/her *own* parameters
- And describes how these *vary*
- ...From which the parameters and their variability can be *estimated*
- *Nonlinear mixed effects models*
Biomedical and public health research:

- **Statistics** plays an **indispensable** role
- **New challenges** require **new methods**
- **Research** to develop new methods is required
- Such research is **supported** by **grants** to provide biostatistical researchers **time**, **computing support**, and **research assistants** (aka **graduate students**)
- Theoretical properties, computer simulations, tests on data from real studies
8. Where do biostatisticians work?

Everywhere biomedical and public health research takes place:

- **Pharmaceutical industry**
- Industry *supporting* the pharmaceutical industry, *contract research organizations* (CROs, e.g. Quintiles)
- **Government**: NIH, FDA, NCHS, ...
- **Institutions**: E.g., Cancer centers (Sloan-Kettering, Dana-Farber, M.D. Anderson), Medical centers (Mayo Clinic, Cleveland Clinic, Duke Clinical Research Institute)
- **Academia**: Biostatistics departments in schools of medicine or public health, Statistics departments like this one!

**Opportunities:** *There is a serious and growing shortage of biostatisticians with advanced degrees!*
9. Training to be a biostatistician

**Background:** *Undergraduate* training in

- Statistics, mathematics, engineering, biological sciences

**Where?** Departments of *biostatistics* or *statistics*

- Advantage of biostatistics department: *Proximity* to medical and public health research
- Advantage of statistics department: Often more theoretical *foundations* in curricula

**Degrees:** Masters and Doctoral
**Coursework:** *Bread and butter*

- Survival analysis
- Longitudinal data analysis
- Categorical data analysis
- Clinical trials and epidemiology

**Coursework:** *Specialized*

- Missing data methods
- Statistical computing
- Genetics/genomics
- Advanced statistical modeling
What is Biostatistics?

**Practical experience:** *Very desirable!*

- *Graduate Industrial Traineeship:* Work in pharmaceutical industry, research institutions
- *Research Assistantship*
10. Concluding remarks

**Biostatistics:** A *huge* field with *many opportunities*

- Chance to apply quantitative skills and training to *betterment of human health*
- Chance to work a part of a *team* with physicians and other experts
- *Huge shortage* means lots of opportunities