Clinical Trials

STA 102: Introduction to Biostatistics

Yue Jiang

The following material was used by Yue Jiang during a live lecture.

Without the accompanying oral comments, the text is incomplete as a record of the presentation.

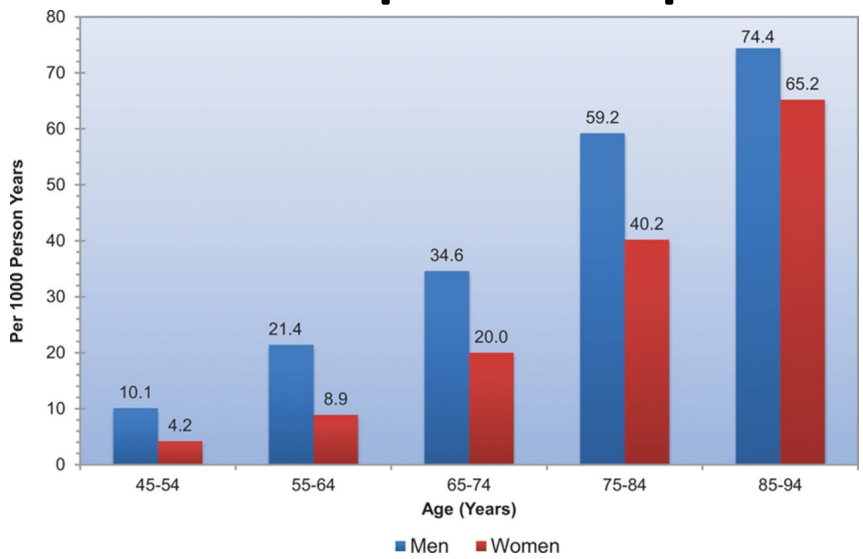
Acknowledgments

Drs. Ed Davis, Anastasia Ivanova, Lisa LaVange

Why do we need clinical trials?

CVD is listed as the underlying cause of death for approximately 1 of every 3 deaths in the US.

Incidence of CVD per 1,000 person-years



- Prevalence: Total number of cases at a given point in time
- Incidence: Number of **new** cases over time

What can we tell from this chart?

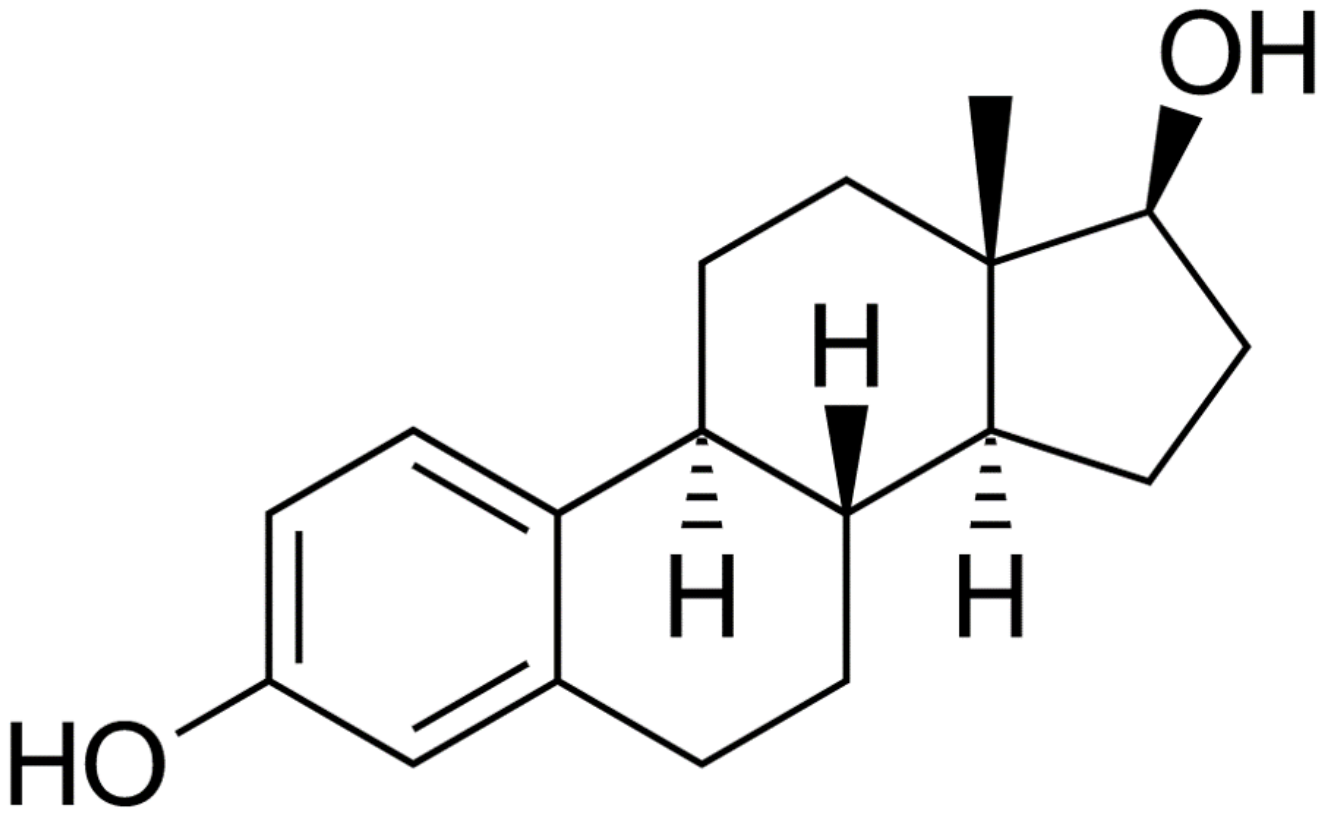
Mozaffarian D, et al. on behalf of the American Heart Association, Circulation (2015)

What can we tell from this chart?

- At any age category, men have a higher rate of CVD than women
- Older women have a higher rate of CVD than younger women

What do women have that men have in smaller amounts
AND
that younger women have much more of than older women?

Estrogen



Estrogen (estradiol)

- Pre-menopausal women: 20 - 400 pg/mL
- Post-menopausal women: 5 - 25 pg/mL
- Men: 10 - 60 pg/mL

Theory: high estrogen levels are *protective* for heart disease

Observational studies

Cross-sectional study: measure estrogen and heart disease at the same point in time

- Three studies measured atherosclerosis via angiogram, accompanied by asking women about estrogen use
- In all three studies, women taking estrogen had less artery blockage than women who were not taking estrogen

Stampfer MJ and Colditz GA, Prev. Med. (1991)

Observational studies

Case-control study: compare the estrogen use of women with heart attacks against a matching set of women who haven't experienced heart attacks

- 5 studies were performed using hospital-based controls (controls in the hospital for non-CVD reasons) and 6 studies used population-based controls
- Studies with hospital-based controls found no association (does not support theory)
- Studies with population based controls found less estrogen use among heart attack victims (supports theory)

Observational studies

Cohort study: identify a large group of women who do not have heart disease and ask them about their estrogen history. Follow these women and observe which group has more heart attacks.

- 16 cohort studies of heart disease and estrogen use in women were performed
- 15 studies found that women taking estrogen had fewer heart attacks than non-estrogen users
- Combining studies, it was estimated that estrogen reduced the risk of a heart attack by about 50% (wow!)

Stampfer MJ and Colditz GA, Prev. Med. (1991)

Newspaper headlines

Observational studies strongly support theory; physicians encouraged to put post-menopausal women on estrogen replacement therapy

Do you think estrogen is protective for heart disease in women?

Let's conduct a prospective clinical trial

"These important observations need to be confirmed in a double-blind, randomized clinical trial, since the protection is biologically plausible and the magnitude of the benefit would be quite large if selection factors can be excluded."

– *Barrett-Connor and Bush, JAMA (1991)*

HERS

The Heart and Estrogen/prosgestin Replacement Study

- 2763 post-menopausal women with previous heart disease recruited
- Half given estrogen + progestin; half given placebo
- Women were followed for 4.1 years

No statistically significant difference in heart attacks, bypass surgery, congestive heart failure, peripheral artery disease, or stroke (*Hulley et al., JAMA (1998)*)

ERA

The Estrogen Replacement and Atherosclerosis Trial

- 309 women with angiographically verified disease
- Randomly divided into 3 groups: estrogen, estrogen + progestin, and placebo
- Followed for 3.2 years

Average artery diameters: 1.87mm, 1.84mm, and 1.87 mm, respectively (*Herrington et al., NEJM (2000)*)

WHI

The Women's Health Initiative

- Two trials, examining estrogen (with or without progestin) vs. placebo, depending on hysterectomy status
- 16,608 women with an intact uterus and no history of CVD followed for 5.2 years
- 10,739 women with hysterectomy and no CVD followed for 6.8 years

No beneficial effect of hormone replacement found on the risk of heart attacks (in fact, higher breast cancer risk among women with intact uteruses) (*Rossouw et al., JAMA (2002); Manson et al., NEJM (2007)*)

Uh oh...

Why did observational studies find such a strong effect, but clinical trials find no effect?

Selection bias in cohort studies

Physicians write prescriptions for hormone replacement therapy. No one writes prescriptions for "no hormone replacement therapy."

Every woman taking HRT had a personal physician. These are women who are more likely to be affluent, have health insurance, and be more health conscious.

"...women who take HRT differ from those who don't in many ways, virtually all of which associate with lower heart disease risk"
– Gary Taubes, writing for the NYT.

Remember, correlation $\neq$ causation.

What is a clinical trial?

1. Clinical trials have defined subjects (usually humans)
2. Clinical trials involve direct intervention
3. Clinical trials are prospective (follow subjects over time)
4. Clinical trials are performed under conditions controlled by researchers, often involving a placebo arm and randomization of assigned treatment groups

The British East India Company

History

- 1848: Bartlett argued that no proof of efficacy could be obtained without a control group
- 1865: Sutton describes first use of placebo (mint water for rheumatic fever)
- 1904: Cushny and Peebles performed a crossover trial for a sleep drug

Anyone remember this paper?

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

The first example from the paper...

SECTION IX. *Illustrations of Method.*

Illustration I. As an instance of the kind of use which may be made of the tables, I take the following figures from a table by A. R. Cushny and A. R. Peebles in the *Journal of Physiology* for 1904, showing the different effects of the optical isomers of hyoscyamine hydrobromide in producing sleep. The sleep of 10 patients was measured without hypnotic and after treatment (1) with D. hyoscyamine hydrobromide, (2) with L. hyoscyamine hydrobromide. The average number of hours' sleep gained by the use of the drug is tabulated below.

The conclusion arrived at was that in the usual dose 2 was, but 1 was not, of value as a soporific.

Additional hours' sleep gained by the use of hyoscyamine hydrobromide.

Patient	1 (Dextro-)	2 (Laevo-)	Difference (2–1)
1.	+ ·7	+ 1·9	+ 1·2
2.	− 1·6	+ ·8	+ 2·4
3.	− ·2	+ 1·1	+ 1·3
4.	− 1·2	+ ·1	+ 1·3
5.	− 1	− ·1	0
6.	+ 3·4	+ 4·4	+ 1·0
7.	+ 3·7	+ 5·5	+ 1·8
8.	+ ·8	+ 1·6	+ ·8
9.	0	+ 4·6	+ 4·6
10.	+ 2·0	+ 3·4	+ 1·4
	Mean + ·75	Mean + 2·33	Mean + 1·58
	S. D. 1·70	S. D. 1·90	S. D. 1·17

Some more history

- 1915: Greenwood and Yule describe random allocation of patients to treatment groups
- 1926: Fisher formally introduces randomization
- 1927: Ferguson et al. use blinding in their study of cold vaccines

Randomization

Allocation of treatments to patients using a random mechanism so that neither the patient or the physician knows in advance which treatment will be assigned.

Why is randomization important?

Can you think of any potential *downsides* of randomization?

Randomization

- Randomization eliminates conscious bias on behalf of the intervener
- Groups become alike in expectation; they may still be unbalanced in practice on certain covariates – but this is ok!
- **With randomization, we may make causal claims**
- Two competing goals in randomization schemes: to keep imbalance low throughout and in the final allocation, and to keep predictability low
- Variance of estimators may be lowered, for instance by stratified randomization by certain factors

Why might we want to lower variance of our estimators?

Blinding

To further eliminate bias, randomized trials are often "blinded" (masked):

- Open: no one masked
- Single-blind: participant masked
- Double blind: participant and intervener masked

Randomization and double-blinding also eliminates *unconscious* bias in terms of care, management, and evaluation.

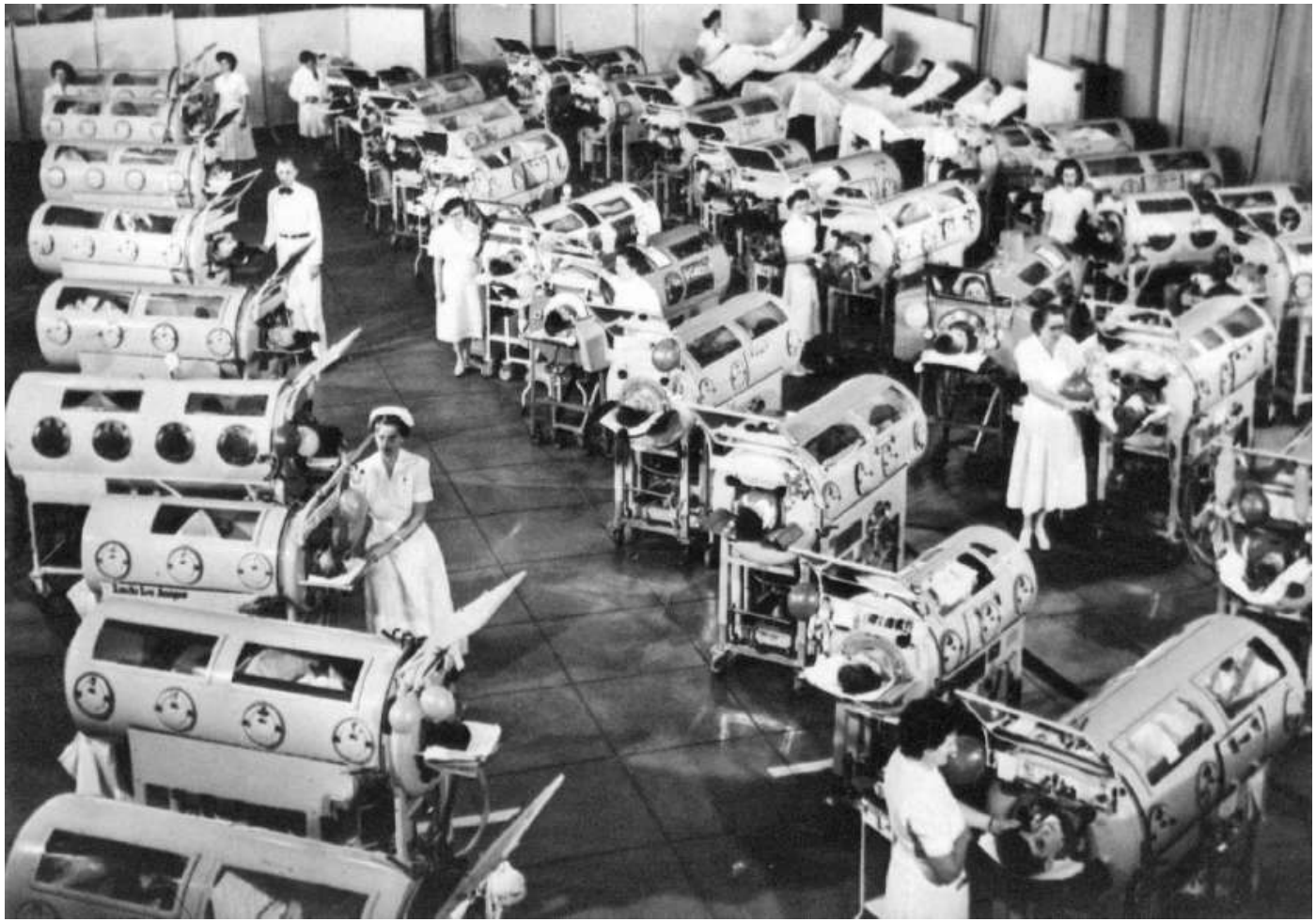
Blinding is not always feasible (e.g., some surgeries)

How might we assess blinding?

History (part 3)

- Arthur Bradford Hill drove the modernization of clinical trials conducted by the UK Medical Research Council
- 1948: first use of a properly randomized control group (streptomycin for pulmonary TB)
- 1950: first use of a placebo-controlled, double-blind study (antihistamines for the common cold)

Poliomyelitis



Salk's trial

In 1954, **1.8 million** children participated in the largest ever clinical trial to assess the effectiveness of Salk's vaccine

0.8 million children participated in a randomized component (1 million in an unrandomized component)

The large number was needed because of the low incidence rate, and also the need to demonstrate evidence of any potential effect as soon as possible

Salk provided the first effective vaccine for polio and a mass immunization campaign was launched (soon replaced by the Sabin oral vaccine in 1962)

Polio vaccination

The last transmission of wild polio in the US occurred in 1979.

Who sponsors clinical trials?

- Government agencies (e.g., the US NIH, the VA, or DoD)
- Non-profit agencies (e.g., the Wellcome Trust)
- Pharmaceutical or medical device companies

Pharmaceutical industry regulation

BAYER Pharmaceutical Products

HEROIN—HYDROCHLORIDE

is pre-eminently adapted for the manufacture of cough elixirs, cough balsams, cough drops, cough lozenges, and cough medicines of any kind. Price in 1 oz. packages, $4.85 per ounce; less in larger quantities. The efficient dose being very small (1-48 to 1-24 gr.), it is

The Cheapest Specific for the Relief of Coughs

(In bronchitis, phthisis, whooping cough, etc., etc.)

WRITE FOR LITERATURE TO

FARBENFABRIKEN OF ELBERFELD COMPANY

SELLING AGENTS

P. O. Box 2160 40 Stone Street, NEW YORK

Pharmaceutical industry regulation

- Prior to 1938, pharmaceuticals were largely unregulated
- The Food and Drug Act created the FDA in 1938
- The FDA had 180 days to review evidence of drug safety
- However, the burden of proof was on the FDA to show that a drug was useless or harmful

Thalidomide

- Brought to market in Germany in 1957 as an OTC sedative (sometimes used to induce sleep in early pregnancy)
- US license obtained in 1959; application to market filed with FDA in 1960
- FDA review in 1961 identified peripheral neuropathy as a potential side effect
- In November 1961, a paper identified serious birth defects linked to thalidomide

Rise of the FDA

- The FDA never approved thalidomide; over 8,000 thalidomide babies born in Europe vs. 40 in the US
- The Kefauver-Harris Amendments in 1962 required active approval from the FDA before drugs were allowed to be marketed
- In 1979, the FDA expanded to meet the Drug Efficacy Study Implementation (DESI) Act, requiring manufacturers to provide "substantial evidence" of efficacy, based on "adequate and well-controlled investigation**s**"

Rise of the FDA

- DESI reviews were of low quality by today's standards (low power; missing data problems; multiple testing ignored)
- Reviews weren't very sophisticated (lots and lots of t-tests)
- In the early 1980s, statistics began to be more important (e.g., by the 90s, the FDA began to accept multiple imputation, sample size re-estimation from interim analyses, better control for multiple testing, and simulation-based approaches)
- Regulatory requirements became more rigorous

Review: multiple testing

Why is not adequately controlling for multiple tests dangerous?

Bonferroni-Holm

1. Sort p-values in ascending order $p_{(1)} < p_{(2)} < \cdots < p_{(m)}$
2. Compare $p_{(1)}$ to α/m (just like normal Bonferroni)
3. If significant, we've reduced our multiple testing problem by one test
4. Compare the next smallest p-value $p_{(2)}$ to $\alpha/(m-1)$
5. ...and so on. Stop at the first non-significant p-value

Strongly controls family-wise error rate

A Simple Sequentially Rejective Multiple Test Procedure - jstor

by S Holm - 1979 - Cited by 18506 - Related articles

Scand J Statist 6: 65-70, **1979**. A Simple Sequentially Rejective Multiple Test Procedure. STURE **HOLM**. Chalmers University of Technology, Goteborg.

The FDA's role

- Drugs, biologics, medical devices, and procedures must be evaluated and demonstrated to be safe and effective in clinical trials
- Clinical trials can evaluate treatment, prevention, quality of life, or early detection

The drug development process

- Phase I clinical trials: small-scale study on healthy volunteers to determine safety and dosage
- Phase II clinical trials: moderately sized trials on patients to evaluate efficacy and side effects
- Phase III clinical trials: large, "pivotal" trials, often multi-center, to verify efficacy and monitor adverse events over a longer period

After pivotal trials are completed, if successful, a firm may file an application to the FDA (depending on what type of drug, e.g., an NDA, BLA, etc.)

Phase 1 trials

- First study of a new drug in humans
- Aims to determine pharmacokinetic and pharmacodynamic properties
- Aims to find a range of well-tolerated doses (including the maximum tolerated dose)
- Most phase 1 studies for non-life threatening diseases are placebo controlled

Balancing efficacy with safety

For some cancer drugs, the therapeutic effect is believed to increase with dose, but this dose may be limited by drug toxicity

We do not want to treat many patients at low, ineffective doses, or high, excessively toxic doses

Ethical considerations require treating patients at lower doses before administering higher doses.

Adaptive designs

Adaptive designs allow us to be more flexible in a clinical trial's design to utilizing observed data accumulating in the trial to modify the trial according to pre-specified rules.

How might an adaptive designs increase statistical power and efficiency? How might an adaptive design be more ethical than a fixed design?

Adaptive trial design is an active field of statistical research.

Phase 2 tests

Phase 2 tests are often the "real test of the drug to do what it is supposed to do" – Donahue and Ruberg (1997).

- Used to determine which doses to carry to pivotal trials, sometimes comparing an active control group
- Aim to determine primary and secondary endpoints
- Often used to estimate treatment effects for future power analyses, recruitment rate, logistics, and potential side effects or toxicity
- For non-life threatening diseases, usually the first studies of patients with the disease – those most likely to benefit and least likely to experience toxicities

Potential designs

- Parallel (fixed group): easy to analyze and interpret
- Crossover studies: reduce the total number of patients needed. However, treatment groups differ with respect to recent exposure to other treatments (carry-over effects)

Sometimes outcomes based on surrogate endpoints, outcomes that can be measured quickly and believed to be related to the clinical outcome of interest (for instance, tumor shrinkage vs. long-term survival)

Phase 3 and beyond

Phase 3 trials are often large, multi-center trials involving multiple doses and active controls

- May last for years – often very expensive and time-consuming
- Often consists of a more heterogeneous population compared to Phase 2 trials
- Comprise the majority of the submitted material to regulatory agencies

Post-marketing surveillance often occurs in further trials

Two types of analyses

Intention-to-treat: every patient randomized in the study will be analyzed.

Per-protocol: only analyze the patients that were fully compliant

- ITT analysis includes patients who drop out prematurely or are non-compliant
- Tends to under-estimate the treatment effect
- More accurately reflects the "real-world"

Which approach is more conservative (in terms of minimizing type I error rate)?

Ethical considerations

It is always important to take into account ethical considerations when conducting clinical trials.

Beneficence: *primum non nocere*

Equipoise: when should a clinical trial be run?

Wrapping up...

Clinical trials are an active field of statistical research!

Today was just a brief introduction to this interesting field.