

Clinical Trials

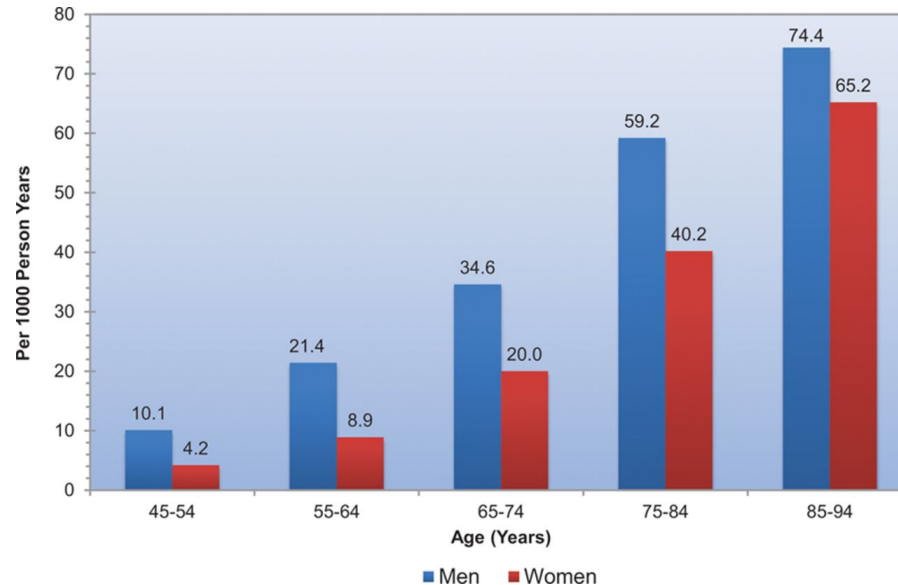
Yue Jiang

Duke University

A disclaimer

The following material was used during a live lecture. Without the accompanying oral comments and discussion, the text is incomplete as a record of the presentation. A full recording may be found via Zoom on the course Sakai site.

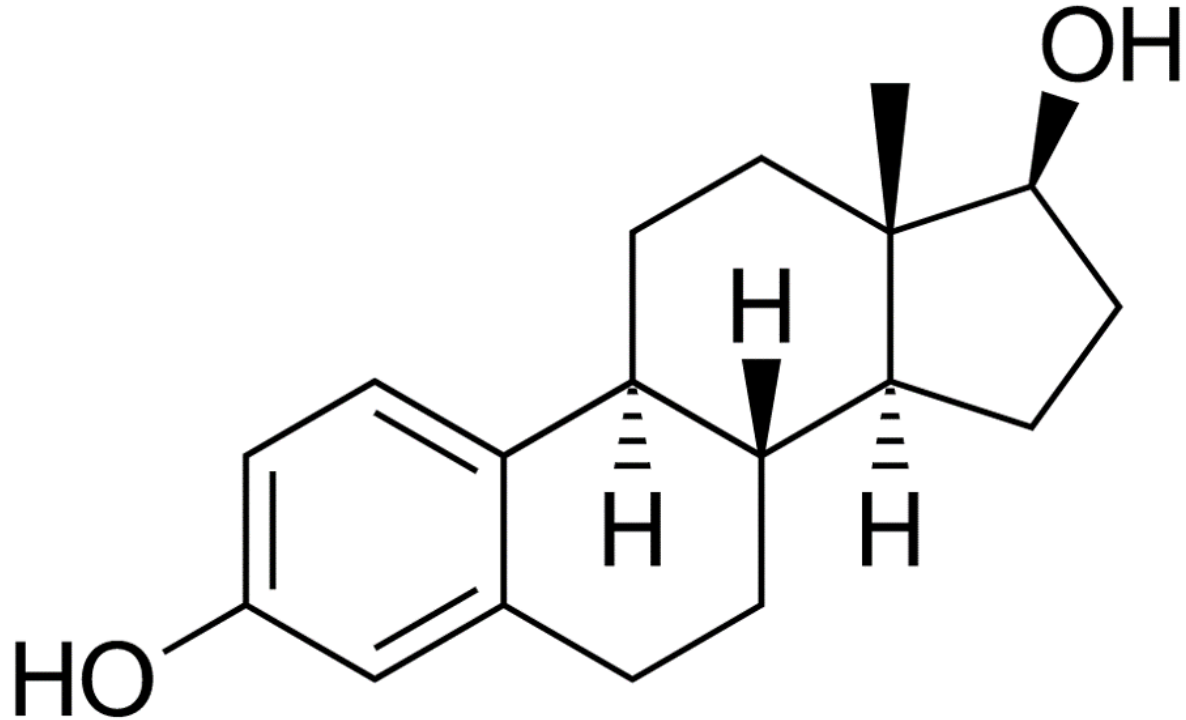
Why clinical trials?



- **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?

Why clinical trials?



Why clinical trials?

Estrogen levels:

- 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

Potential study designs

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

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

- 5 studies used 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
- However, studies with population based controls did find less estrogen use among heart attack victims

Potential study designs

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 using meta-analysis, it was estimated that estrogen was associated with reduced risk of a heart attack of about 50\% (wow!)

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

Potential study designs

“...the protection is biologically plausible and the magnitude of the benefit would be quite large...”

Barrett-Connor and Bush, JAMA (1991)

Clearly, these studies strongly support the theory, and physicians were encouraged to put post-menopausal women on estrogen replacement therapy to avoid CVD.

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

Further studies

HERS: The Heart and Estrogen/Progestin Replacement Study

- 2,763 post-menopausal women with previous heart disease
- Half given estrogen + progestin; half given placebo
- Women 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)

Further studies

ERA: The Estrogen Replacement and Atherosclerosis Trial

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

Average artery diameters of 1.87mm., 1.84mm., and 1.87mm., respectively

Herrington et al., NEJM (2000)

Further studies

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)

What happened?

Why did the initial studies find such a strong effect, but further studies find no effect?

What happened?

"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)

Selection bias" physicians write prescriptions for hormone replacement therapy. No one writes prescriptions for "no hormone replacement therapy."

...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, NYT

Clinical trials

Clinical trials are often **experimental studies** as opposed to **observational studies**.

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

History of clinical trials



History of clinical trials

- 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

VOLUME VI

MARCH, 1908

No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

By STUDENT.

History of clinical trials

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

History of clinical trials

- 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

Why is randomization important? Can you think of any *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 along certain covariates, but this is ok!). Importantly, with randomization, we may make **causal claims**.

There are two competing goals in randomization: to keep imbalance low throughout and in the final allocation, and to keep predictability of treatment low.

Variance of estimators may be lowered by **stratifying** on certain factors.

Why might we want to lower the variability of our estimators?

Blinding

To further eliminate bias, randomized trials are often **blinded** or masked

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

Randomization and blinding help eliminate *unconscious bias* in terms of care, management, and evaluation

How might we assess blinding?

More history

- 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)

More history



More history

- 1954: 1.8 million children participated in the largest ever clinical trial to assess the effectiveness of Salk's vaccine (0.8 million the randomized component). Salk provided the first effective vaccine for polio and a mass immunization campaign was launched (soon replaced by the Savin oral vaccine in 1962).
- 1979: The last transmission of wild polio in the US

Why was such a large number needed?

More history



Pharmaceutical industry regulation



Am. J. Ph.]

7

[December, 1901

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

Clinical trials are generally sponsored by government agencies (e.g., US NIH, VA, etc.), non-profit agencies (e.g., the Wellcome Trust), or pharmaceutical or medical device companies

- 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
- The burden of proof was on the FDA to show that a drug was useless or harmful

Pharmaceutical industry regulation

Thalidomide was 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
- A 1961 paper identified serious birth defects linked to thalidomide

The FDA never approved thalidomide; over 8,000 thalidomide babies were born in Europe vs. just 40 in the US

- The Kefauver-Harris Amendments in 1962 required active approval from FDA before active marketing
- 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 investigations"

Pharmaceutical industry regulation

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
- By 1990s, FDA began to accept multiple imputation, sample size re-estimation from interim analyses, better control for multiple testing, and simulation- based approaches

Why is not adequately controlling for multiple tests dangerous?

The Bonferroni-Holm method

1. Sort p-values in ascending order $p_{(1)} < p_{(2)} < \dots < p_{(m)}$
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.

The Bonferroni-Holm method also strongly controls family-wise error rate without being as conservative as the Bonferroni method

[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 drug development process

- **Phase I trials:** small-scale study on healthy volunteers to determine safety and dosage
- **Phase II trials:** moderately sized trials on patients to evaluate efficacy and side effects
- **Phase III 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 (or analogous body in other countries)

Phase I trials

Often the first study of a new drug in humans; aim to determine pharmacokinetic and pharmacodynamic properties and find a range of well-tolerated doses. Most Phase I studies for non-life-threatening diseases are placebo-controlled.

For some drugs (e.g., chemotherapy), 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.

How might we use an adaptive trial design to use data accumulating in the trial to modify the trial according to pre-specified rules? How might they increase power and efficiency?

Phase II trials

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 using an **active control** group.

Aim is to determine primary and secondary endpoints, and 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

Potential designs

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

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

Phase III and beyond

Phase III trials are often large, multi-center trials involving multiple doses and active controls; they may last for years and are often very expensive and time-consuming.

Patient population is more heterogeneous compared to Phase II trials, and comprise the majority of the submitted material to regulatory bodies.

Post-marketing surveillance often occurs in further trials.

Why does a more heterogeneous population potentially reduce power? How can that be combatted?

Two types of analyses

Intention-to-treat (ITT): every patient randomized in the study will be analyzed, including those who are non-compliant.

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

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