500 Class 1 Slides

github.com/THOMASELOVE/500-2018

2018-01-16

If you think a good statistical analysis is expensive, try a bad one.
- Attributed to various sources

All models are wrong, some are useful. - G.E.P. Box

82.8% of all statistics are made up on the spot. - Source unknown

Today's Agenda

- Course Overview and the Web Site
- Statistical Philosophy and Problem-Solving
- What are Observational Studies?
    - How do we think about causal effects?
    - Why is randomization so important?

Course Overview

- Randomized Experiments vs. Observational Studies
  - Randomization as the "fundamental basis for inference"
  - Observational Studies and Causal Effects
- Propensity Scores: Crucial Tools for Causal Models
  - Selection Bias: key issue for observational studies
  - Dealing with Bias (both overt and hidden)
  - Subclassification (stratification) and direct regression adjustment
  - Matching and weighting using the Propensity Score
  - Sensitivity Analysis
- Instrumental Variables
- Using R, R Studio and R Markdown to accomplish all of this

Paul Rosenbaum's 2017 book *Observation & Experiment*

The Web Site

https://github.com/THOMASELOVE/500-2018

- Assignments
- Data and Code
- a short biography of me - I'm at **Thomas dot Love at case dot edu**
- Presentations
- Project
- Schedule
- Syllabus
- Other Texts

Password Protected Materials

The password is (in lower case letters) the last name of a famous part of the history of statistics. I'll be specific in class.

My Expectations

- You are interested in learning about the effects of an intervention, treatment or policy on subjects when the treatments cannot be assigned at random.
- You have little interest in technical details of methods, but serious interest in designing, conducting and analyzing observational studies skillfully.
- You have access to software (specifically R) which you can use to obtain basic hypothesis testing, regression and logistic regression results.

Assignment for Class 2

- Materials are on the Assignments page
- Submit your work via canvas.case.edu by noon on January 25.

Software

1. Download and install the most recent version of R (version 3.4.3. or later) from cran.case.edu for your operating system. If you need further help on this, let me know.
2. Download and install the most recent version of R Studio (version 1.1.383 or later) from https://www.rstudio.com/products/rstudio/download/#download for your system. If you need further help on this, let me know.

If you are rusty on (or new to) R, visit the Data and Code page and look at the R Basics materials.

Software

3. Run the following code to install some useful packages in R Studio

```
install.packages(c("arm", "BayesTree", "car", "CBPS",
    "cobalt", "devtools", "ebal", "Epi", "faraway",
    "fivethirtyeight", "foreign", "gapminder", "GGally",
    "ggrepel", "gridExtra", "Hmisc", "knitr", "lme4",
    "markdown", "MASS", "Matching", "MatchIt", "mice",
    "multcomp", "optmatch", "pander", "psych", "pwr",
    "rmarkdown", "rms", "roxygen2", "rstanarm",
    "sandwich", "survey", "survival", "tableone",
    "testthat", "tidyverse", "twang", "viridis"))
```

The Old Homework 1 Example

In previous iterations of this class, I have given a (relatively) elementary R assignment. Both the assignment and a fancy answer sketch are available with the new Homework 1 materials. I expect this assignment will be of interest if you:

- are new to R, or R Studio, or R Markdown
- are new to logistic regression
- like seeing the code of other people
- are working on assignments that will come our way after Class 2

Deadlines and Deliverables

- No exams.
- Four types of deliverables
  - Course Project (3-5 tasks, depending on how you count)
  - Observational Studies in Action (2-3 tasks)
  - Essays in reaction to Rosenbaum text (7 of these)
  - 5 garden-variety homework assignments.

A Key Goal for the Project and Course

- Help you learn how to tackle a problem, rather than just be able to perform particular statistical techniques.
  - Goal: think and solve problems when trying to infer causal effects from observational data
- But the need to think in statistical terms is omnipresent
  - Identifying researchable problems
  - Dealing with variation
  - Interplay of Design and Analysis
  - Preparing, writing and revising results, in a replicable way.

Comparative Effectiveness Studies

- We have an outcome measured on two groups of subjects (treated and control).
- We want to make a fair comparison between the treated group and the control group in terms of the outcome.
- We want to ensure that the groups are comparable in terms of **covariates** (that describe the subjects before the treatments are applied.)
- If they aren't comparable, it will be difficult for us to make a fair comparison.

Some Philosophy

But not too much.

Statistical Thinking as Problem Solving

We need to be able to:

1. Formulate a real problem in statistical terms.
2. Give advice on efficient and effective data collection.
3. Analyze data to extract the maximum amount of useful information.
4. Interpret and report the results.

- Chatfield C (1995) *Problem Solving: A Statistician's Guide*

Stages of a Statistical Investigation

Statistical thinking is required in all stages of the investigation:

1. Planning the Study
2. Collecting the Data
3. Analyzing the Results
4. Interpreting the Analyses
5. Presenting the Study

We'll spend some time in all five stages.

Early Stages of an (Idealized) Investigation

- Understand the problem, then formulate it in statistical terms.
  - Clarify the objectives very carefully. Ask as many questions as necessary. Search the literature.
- Plan the investigation and collect the data in an appropriate way.
  - Achieve a fair balance between the effort expended in collecting the data, and in analyzing them.
  - Method of collection crucial to further analysis.

Middle Stages of an (Idealized) Investigation

- Assess the structure and quality of the data.
  - Coding, typing, editing, etc.
  - Data cleaning: looking for errors, outliers, missing
  - Decide how to deal with peculiarities.
  - How much time does this take?
- Describe the data / identify interesting features
  - Descriptive summary is sometimes all you need
  - Always helpful in motivating further analyses
  - Ever done a power calculation?

Final Stages of an (Idealized) Investigation

- Select and carry out appropriate analyses
  - Often assume a particular model structure, set out in advance
  - Estimate parameters, test hypotheses
  - Check adequacy of fitted model, through residual analysis and considering refinements
- Compare findings with prior results and acquire further data as necessary
- Interpret and communicate the results

Philosophical Biases

- Emphasis on the initial examination of data
  - Essential precursor to model-building
  - Allows us to "design" our analyses suitably
  - Harder than it looks, even after the data are "clean"
- Robust near-optimal solutions beat "optimal" solutions that rely on dubious assumptions
  - Assumptions are unlikely to be satisfied exactly and may be seriously in error.
  - In observational studies, assumptions are always important. We are looking for safe, practical and reliable approaches.

Recurring Themes (from Chatfield 1995)

- Need to understand strategy, clinical concerns, **and** know techniques.
- Need to ask questions when formulating a problem and building a model to understand practical context.
- Importance of the initial data analysis for checking data quality, summarizing, and guiding further work.
    - What is "the model" going to be used for?
    - What is the question I really want to answer? Are the data amenable?
- Avoiding trouble is complementary to, and a prerequisite for, finding an optimal solution.

What this course is about...

An **observational study** concerns treatment, interventions or policies and the effects they cause, and in this respect it resembles an experiment. A study without a treatment is neither an experiment nor an observational study.

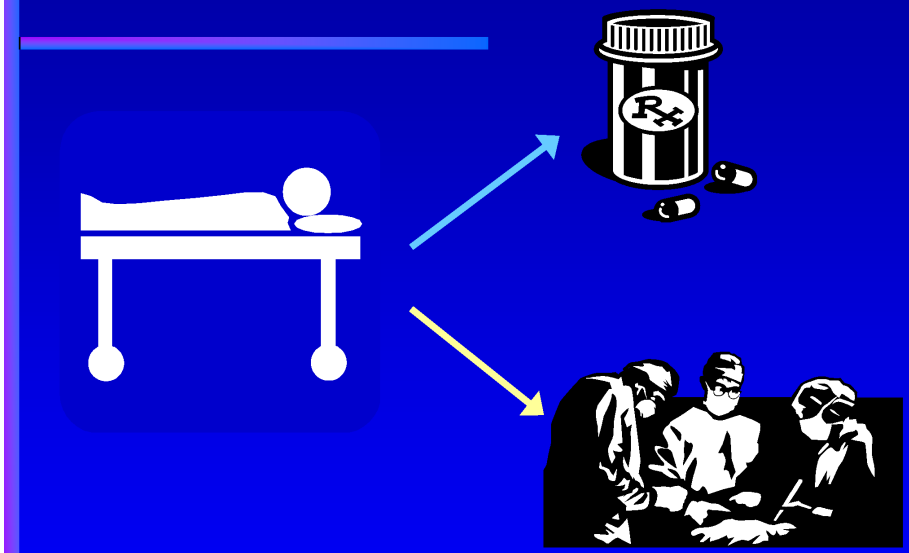
In an experiment, the assignment of treatments to subjects is controlled by the experimenter, who ensures that subjects receiving different treatments are comparable. In an observational study, this control is absent.

Rosenbaum 2002, Chapter 1

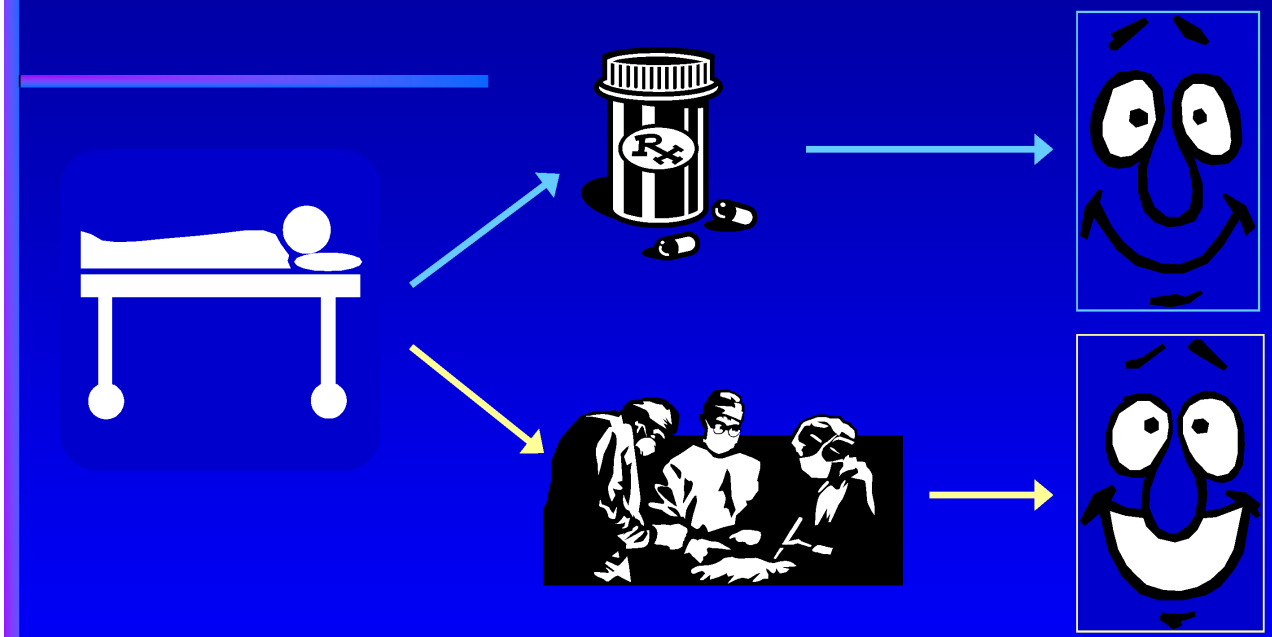
Looking for Causal Effects of an Exposure, or Treatment

Looking for Causal Treatment Effects

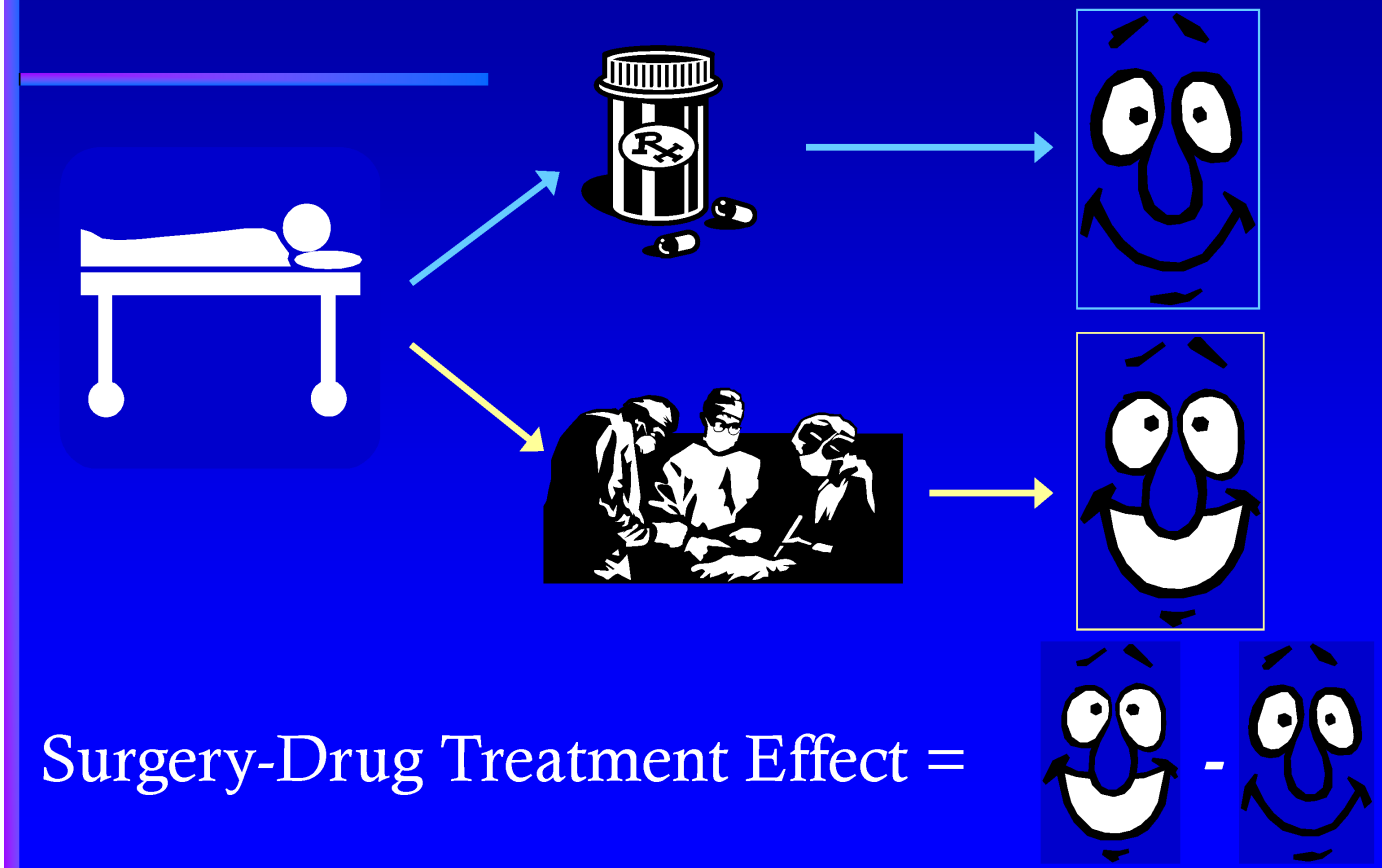
Looking for Causal Treatment Effects



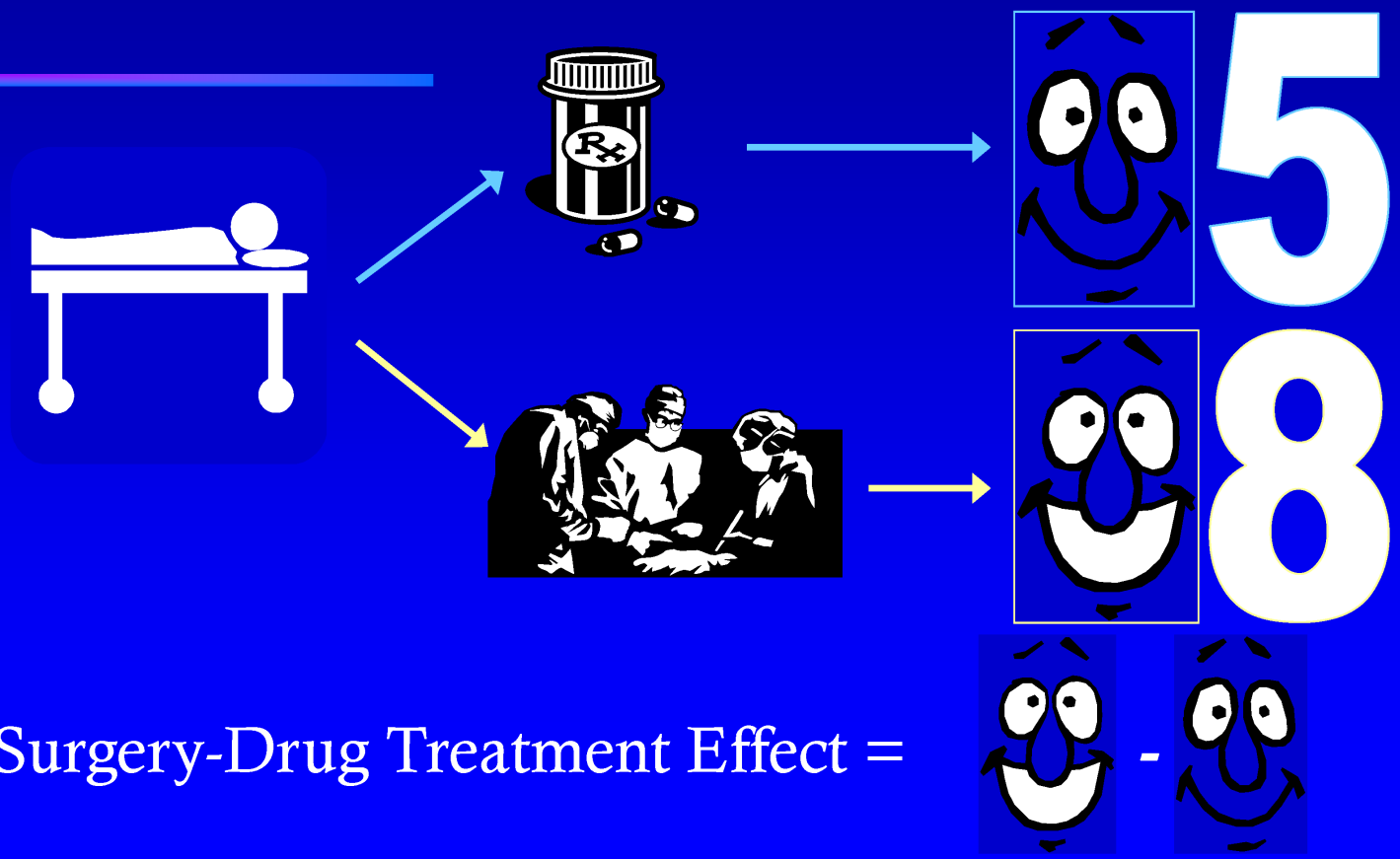
Looking for Causal Treatment Effects



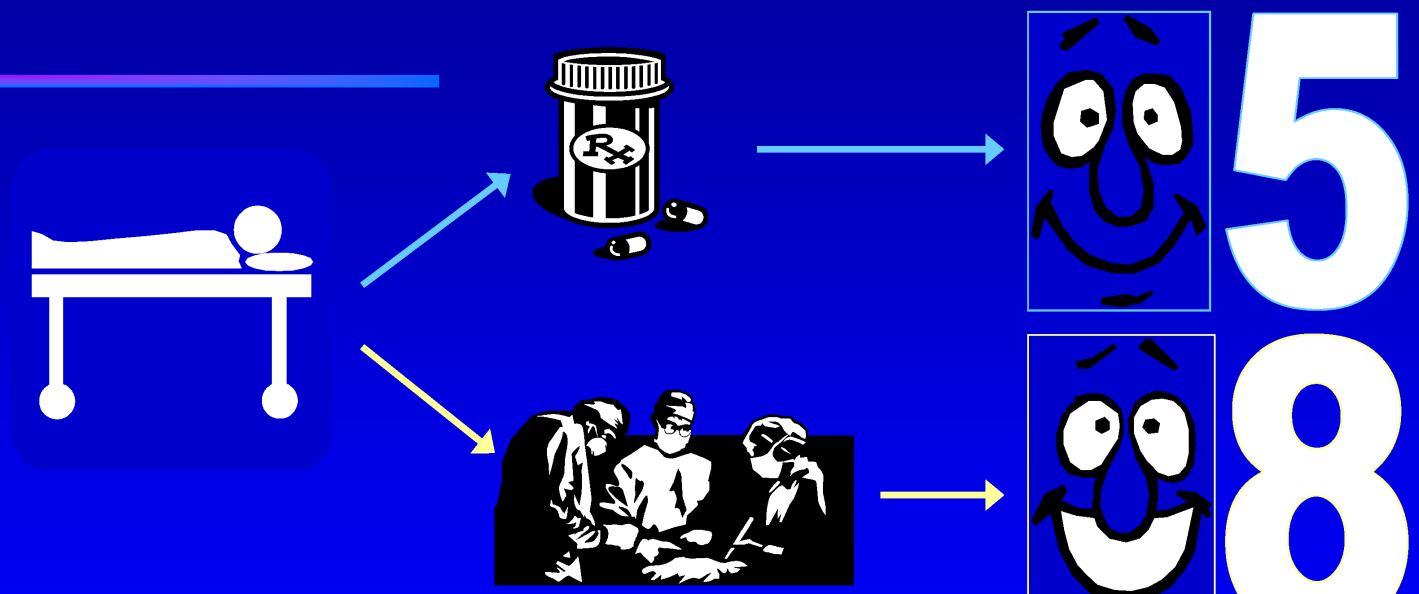
Looking for Causal Treatment Effects



Looking for Causal Treatment Effects



Looking for Causal Treatment Effects



Surgery-Drug Treatment Effect = 3

What is an experiment?

- Goal of an experiment: inference about the effects of treatments / exposures.
- Crucial: assignment of treatments to subjects controlled by the experimenter
- This should ensure that subjects receiving different treatments are comparable.
- Best choice is usually random assignment.

from Concato et al. NEJM 2000

RANDOMIZED, CONTROLLED TRIALS, OBSERVATIONAL STUDIES, AND THE HIERARCHY OF RESEARCH DESIGNS

JOHN CONCATO, M.D., M.P.H., NIRAV SHAH, M.D., M.P.H., AND RALPH I. HORWITZ, M.D.

In the hierarchy of research designs, the results of randomized, controlled trials are considered to be evidence of the highest grade, whereas observational studies are viewed as having less validity because they reportedly overestimate treatment effects.

Concato, Table 1 (USPSTF Evidence Grades, 2000)

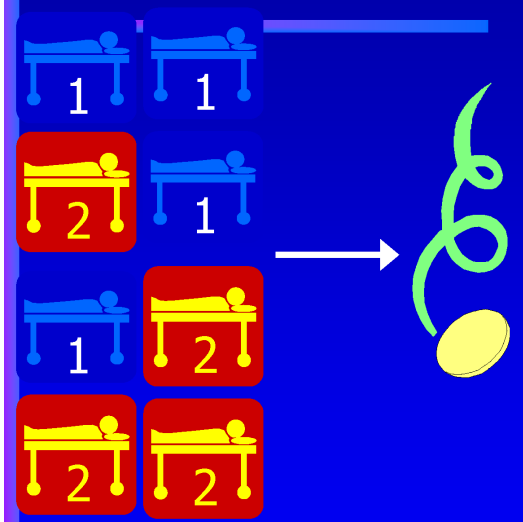
TABLE 1. GRADES OF EVIDENCE FOR THE PURPORTED QUALITY OF STUDY DESIGN.*

I	Evidence obtained from at least one properly randomized, controlled trial.
II-1	Evidence obtained from well-designed controlled trials without randomization.
II-2	Evidence obtained from well-designed cohort or case–control analytic studies, preferably from more than one center or research group.
II-3	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.
III	Opinions of respected authorities, based on clinical experience; descriptive studies and case reports; or reports of expert committees.

*The grades are those of the U.S. Preventive Services Task Force.[7]

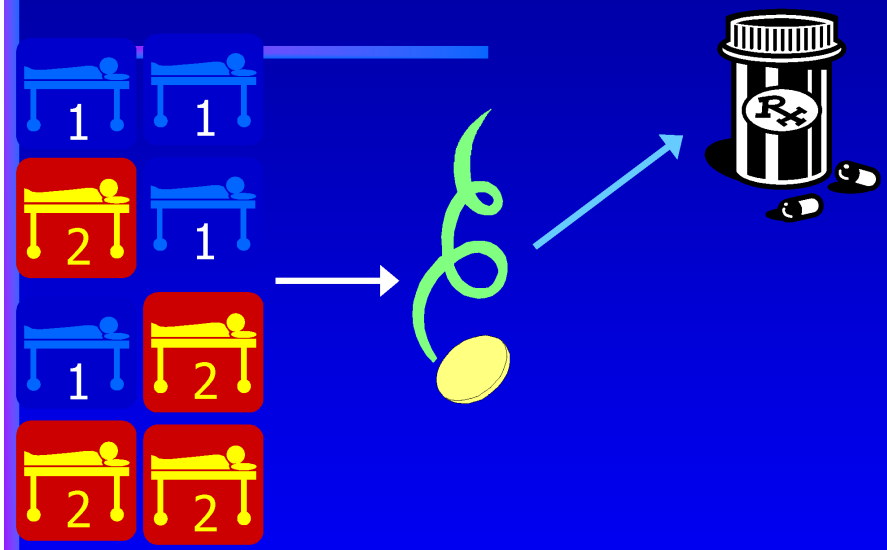
Importance of Randomization

Importance of Randomization



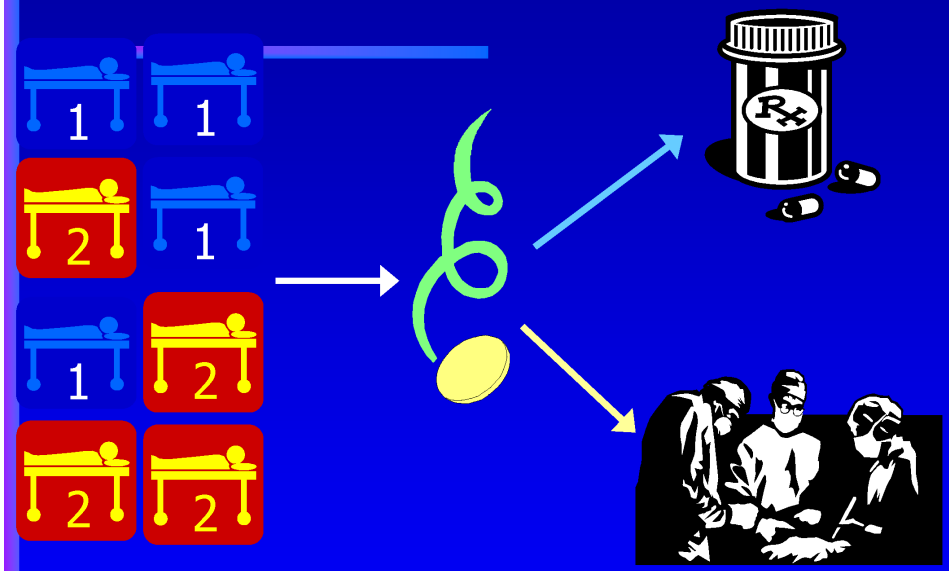
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



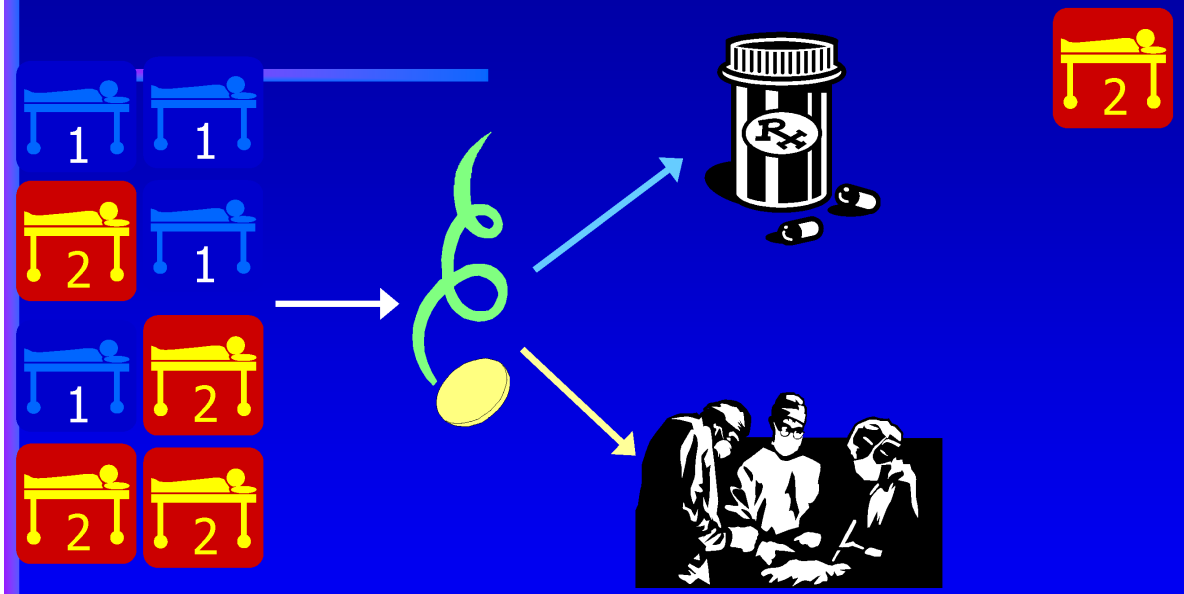
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



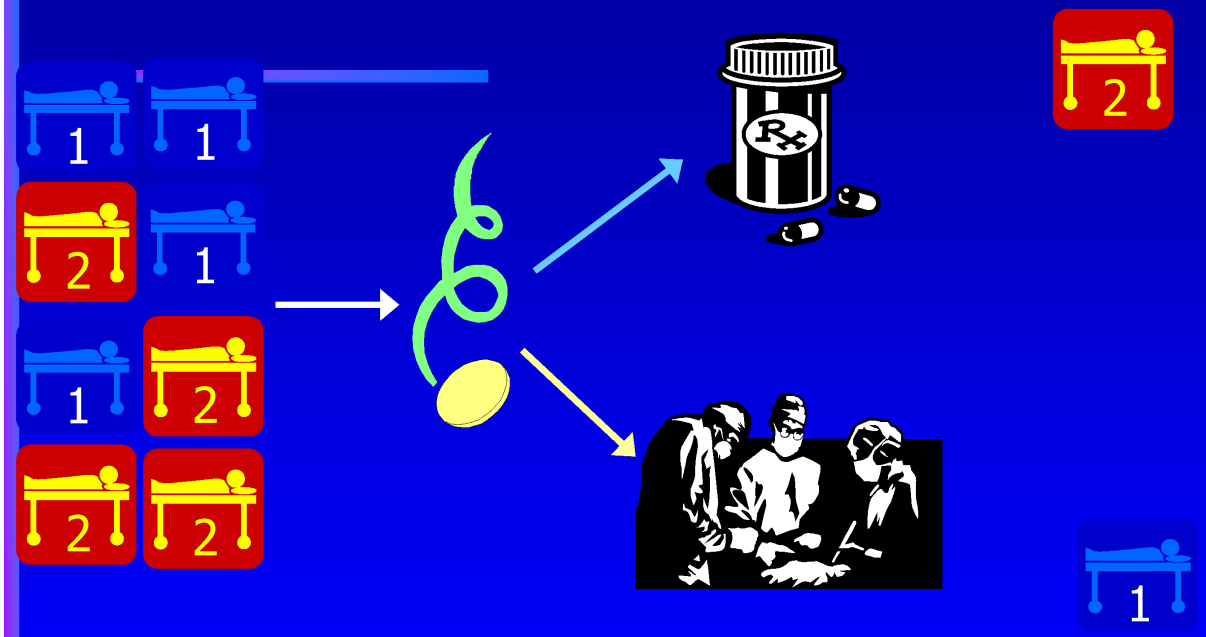
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



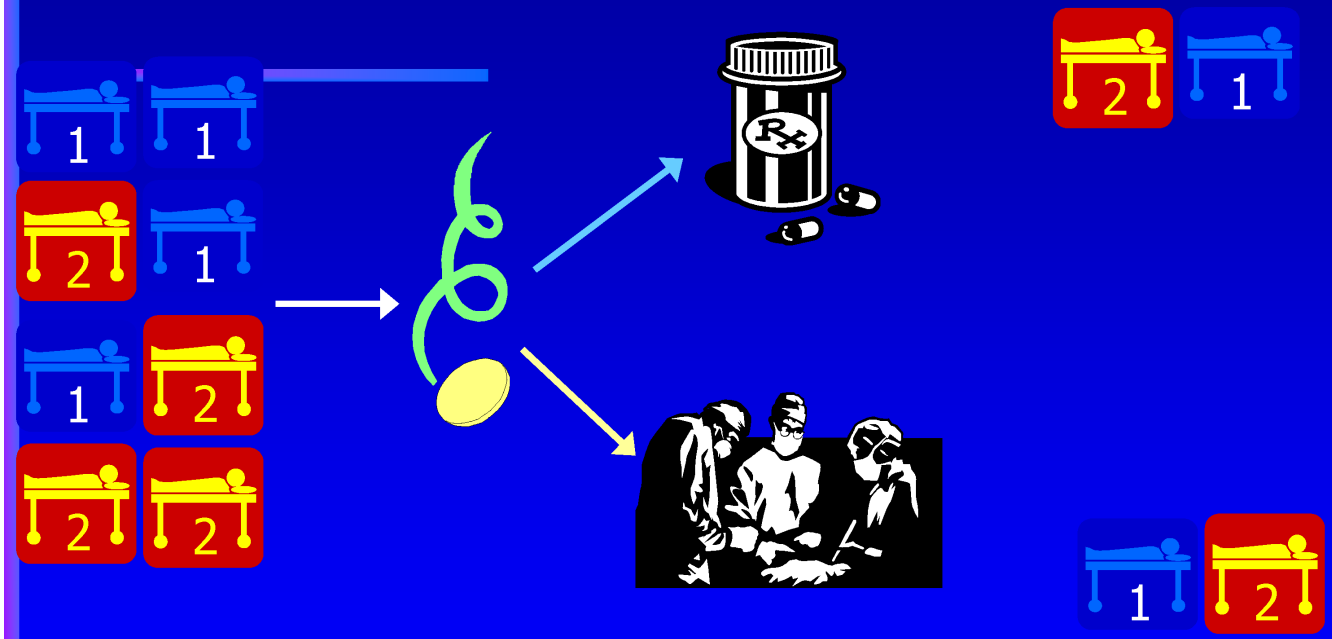
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



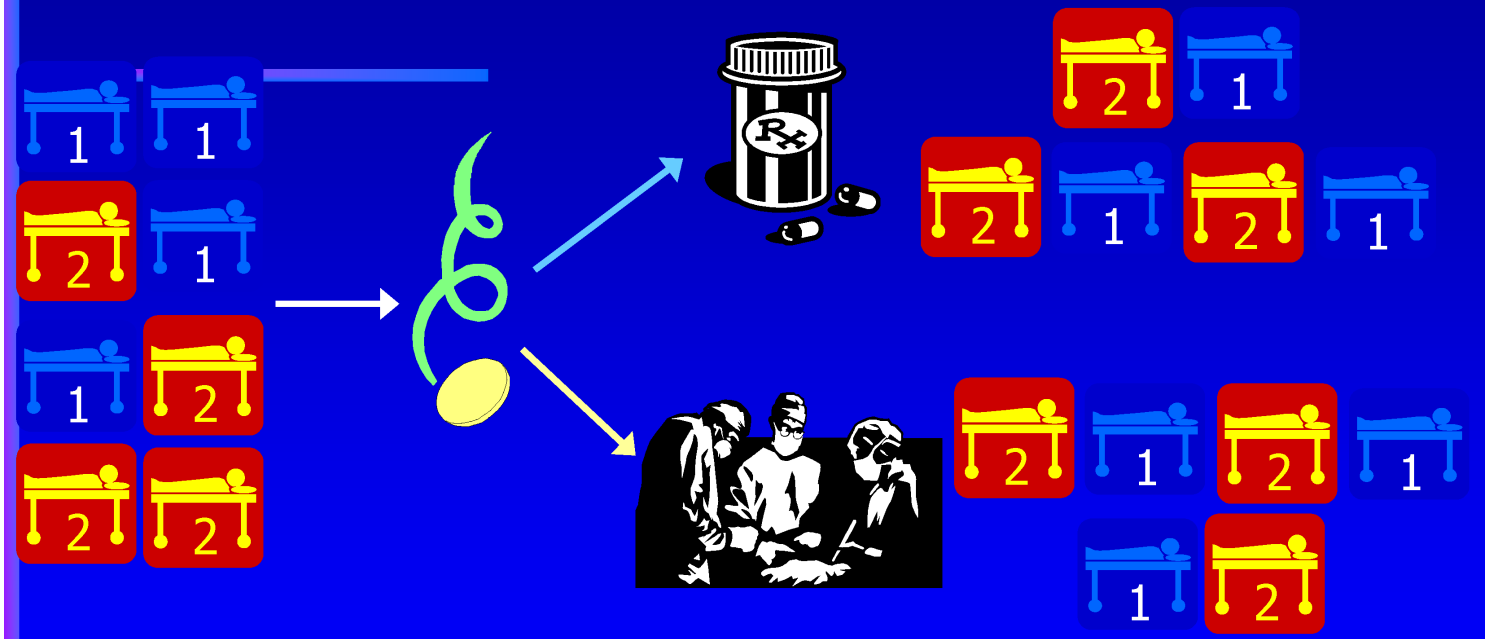
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



Randomization ensures that subjects receiving different treatments are comparable.

A Randomized Clinical Trial (RCT) of Coronary Surgery

- VA conducted a randomized controlled experiment for coronary artery disease
  - Coronary artery bypass surgery vs.
  - Medical therapy (Drug treatments)
- 596 patients at 13 VA hospitals
  - 286 got surgery, 310 got medical therapy
  - Random Assignment of Treatments
  - Were the subjects comparable? Is it appropriate to check?
- To whom do we wish to make inferences?
- What is our actual research question?

Baseline Comparison of VA Coronary Patients RCT

Covariate (pre-treatment)	Medical %	Surgery %
NY Heart Assoc. Class II & III	94.2	95.4
History of myocardial infarction (MI)	59.3	64.0
Definite / possible MI (electrocardiogram)	36.1	40.5
Duration of chest pain $>$ 25 mos.	50.0	51.8
History of hypertension	30.0	27.6
History of congestive heart failure	8.4	5.2
Cardiothoracic ratio $>$ 0.49	10.4	12.2
Serum cholesterol $>$ 249 mg/dl **	31.6	20.6

** $p < 0.05$ for difference between medical and surgery groups

Results of the VA Coronary Surgery Trial

- The VA study compared survival in the two groups three years after treatment.
  - Survival in the medical group was 87%
  - Survival in the surgical group was 88%
  - Both had a standard error of 2%, so the 1 percentage point difference in mortality was not significant
- Evidently, when comparable groups of patients received medical and surgical treatment at VA hospitals, outcomes were quite similar.

Why wouldn't you always do experiments?

- Any thoughts?
- Are there situations where random assignment of subjects to exposures/treatments is not possible?

Why not always do experiments?

- The treatment might be harmful and cannot be given to human subjects for experimental purposes.
- The treatment may be controlled by a political process that will not yield control.
- The treatment may be beyond the legal reach of experimental manipulation.
- Experimental subjects may have strong attachments to particular treatments.
- Sometimes, we have other problems. . .

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)

> *The Multiple Risk Factor Intervention Trial was a randomized primary prevention trial to test the effect of a multifactor intervention program on mortality from coronary heart disease (CHD) in 12,866 high-risk men aged 35 to 57 years.*

- Men were randomly assigned to a special intervention (SI) program or to usual care (UC)

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)

> *The Multiple Risk Factor Intervention Trial was a randomized primary prevention trial to test the effect of a multifactor intervention program on mortality from coronary heart disease (CHD) in 12,866 high-risk men aged 35 to 57 years.*

- Men were randomly assigned to a special intervention (SI) program or to usual care (UC)
- SI includes stepped-care treatment for hypertension, counseling for cigarette smoking, and dietary advice for lowering blood cholesterol.

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)

> *The Multiple Risk Factor Intervention Trial was a randomized primary prevention trial to test the effect of a multifactor intervention program on mortality from coronary heart disease (CHD) in 12,866 high-risk men aged 35 to 57 years.*

- Men were randomly assigned to a special intervention (SI) program or to usual care (UC)
- SI includes stepped-care treatment for hypertension, counseling for cigarette smoking, and dietary advice for lowering blood cholesterol.
- Men were followed for an average of seven years

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)

> *The Multiple Risk Factor Intervention Trial was a randomized primary prevention trial to test the effect of a multifactor intervention program on mortality from coronary heart disease (CHD) in 12,866 high-risk men aged 35 to 57 years.*

- Men were randomly assigned to a special intervention (SI) program or to usual care (UC)
- SI includes stepped-care treatment for hypertension, counseling for cigarette smoking, and dietary advice for lowering blood cholesterol.
- Men were followed for an average of seven years
- Risk factor levels declined in both groups, more in the SI group.

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)

> *The Multiple Risk Factor Intervention Trial was a randomized primary prevention trial to test the effect of a multifactor intervention program on mortality from coronary heart disease (CHD) in 12,866 high-risk men aged 35 to 57 years.*

- Men were randomly assigned to a special intervention (SI) program or to usual care (UC)
- SI includes stepped-care treatment for hypertension, counseling for cigarette smoking, and dietary advice for lowering blood cholesterol.
- Men were followed for an average of seven years
- Risk factor levels declined in both groups, more in the SI group.
- CHD mortality 17.9 deaths/1000 in SI, 19.3 in UC (not sig.)

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

1. Exclusions if . . .
   - Low risk of CHD
   - History of MI
   - Diabetes
   - Geographic Mobility is an issue
   - Cholesterol > 350
   - DBP > 115

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

1. Exclusions if . . .
    - Low risk of CHD
    - History of MI
    - Diabetes
    - Geographic Mobility is an issue
    - Cholesterol > 350
    - DBP > 115

How many men do you suppose this leaves in the study?

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

1. Exclusions if . . .
   - Low risk of CHD
   - History of MI
   - Diabetes
   - Geographic Mobility is an issue
   - Cholesterol > 350
   - DBP > 115

These exclusions affected 336,117 of the men.

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

1. Exclude 336,117 men, leaving 25,545 candidates for screening.

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

1. Exclude 336,117 men, leaving 25,545 candidates for screening.
2. Screen 25,545 men.

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

1. Exclude 336,117 men, leaving 25,545 candidates for screening.
2. Screen 25,545 men, and exclude if. . .
    - Body Weight is more than 150% of expected
    - Angina
    - Evidence of MI
    - Consuming a special diet

How many of these 25,545 men will be left?

Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

1. Exclude 336,117 men, leaving 25,545 candidates for screening.
2. Screen 25,545 men, and exclude if. . .
    - Body Weight is more than 150% of expected
    - Angina
    - Evidence of MI
    - Consuming a special diet

And this affects 12,678 of these men.

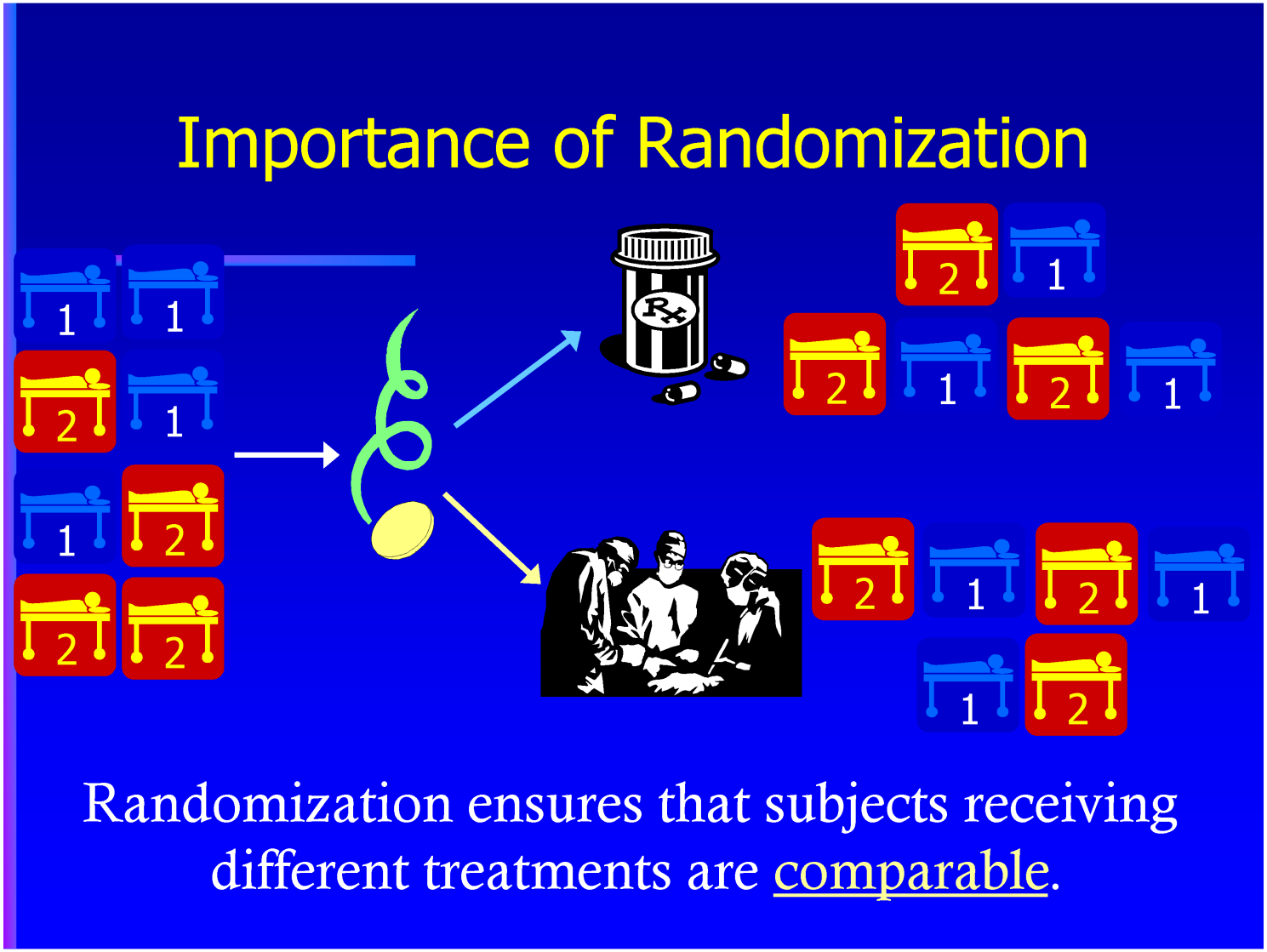
Example of RCT Subject Selection (MRFIT)

Start with 361,662 men ages 35-57

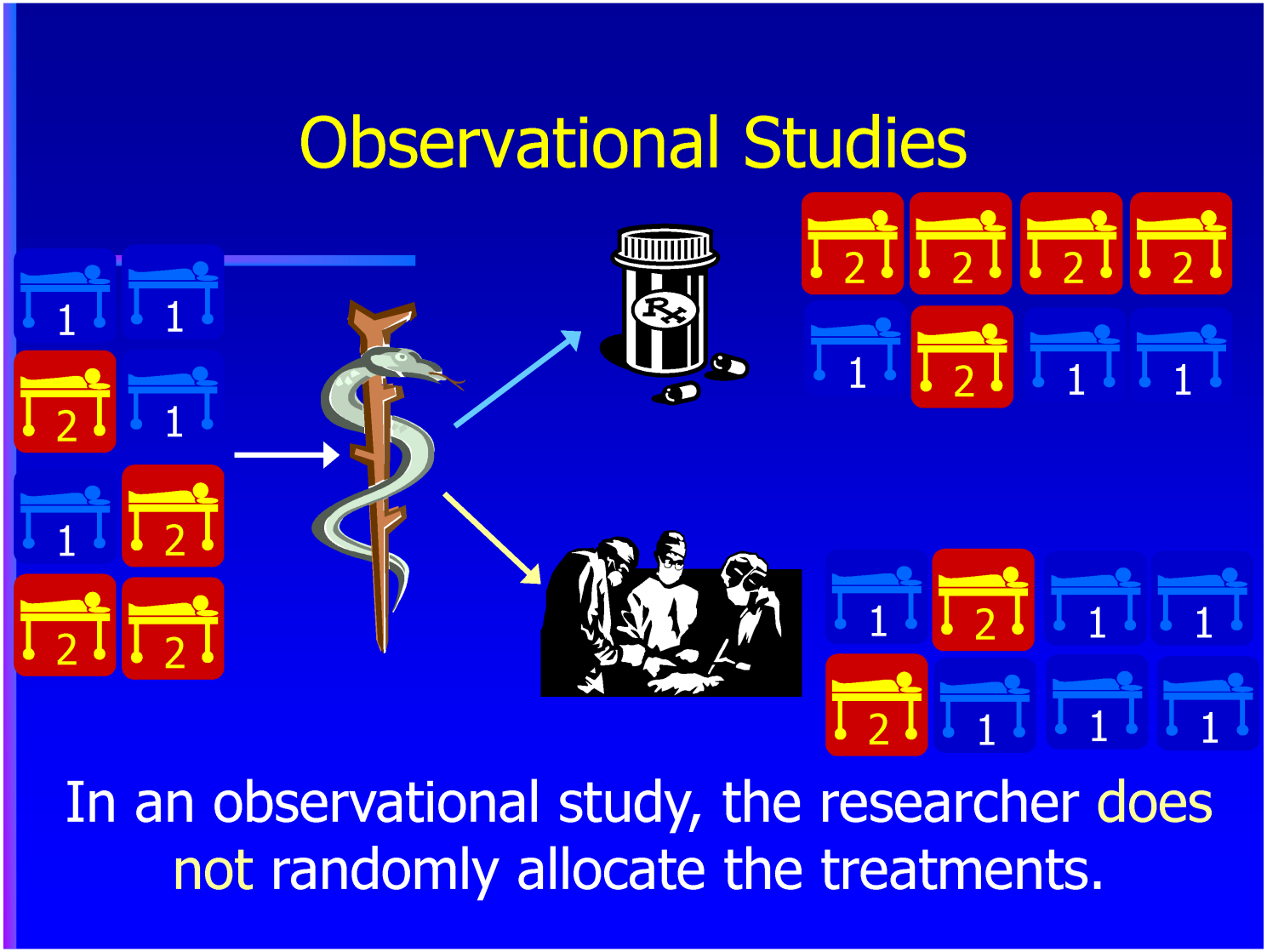
1. Exclude 336,117 men, leaving 25,545 candidates for screening.
2. Screen 25,545 men, and exclude 12,678.
3. Take the remaining sample of 12,866 and randomize into . . .
    - one group of 6,428 men
    - and the other group of 6,438 men

Bottom Line: MRFIT excluded 96.4% of potential eligibles.

Randomized Clinical Trial



Observational Studies



Observational Studies to Estimate Causal Effects

- An observational study (OS) concerns treatments and their effects, BUT the researcher does not control (cannot randomize) the assignment of treatments
- We want to compare groups receiving the two treatments who looked similar prior to the treatment assignment.
- Analytical adjustments required to account for baseline (covariate) differences.

Data Collection Strategies

- Experiments require active intervention by the investigator.
- An OS is more passive, but often attempts to look at the same sort of effect.
- Retrospective trials observe responses on carefully selected subjects, whose history is then examined to assess which variables are important in determining the condition of interest.
- Prospective trials are safer, more time-consuming.

One Observational Study

- Data for 7 different treatments for child ear infections collected from more than 400 children at a local hospital over a six-month period.
  - Large differences in re-infection rates appeared in the data for the different treatments.
  - Is this a useful indication that some treatments were better than others?

This was an observational study. What does that mean, and how does it affect our conclusions? Our design? Our analyses?

- Example is discussed in Chatfield C (1995)

Caveats for Ear Infection Study

- Doctors allocated the treatment they liked best, or thought was best for the patient.
  - Good results may indicate a good doctor, rather than a good treatment.
  - Some doctors always used the same treatment so treatment effect is hopelessly confounded with doctor effect.
  - Other doctors may have given particular treatments to particular groups (i.e. worst affected) of patients.

Are these historical data useful? Better than nothing?

How could we start to address these problems, in re-designing the study?

The Importance of Randomization

We want to compare groups who looked similar **before** they were exposed to interventions/treatments.

- Randomization tends to produce relatively comparable or "balanced" treatment groups in large experiments.
  - The covariates are not used in assigning treatments in an experiment.
  - There is no *deliberate* balancing of the covariates: it's just a nice feature of randomization.
- We have some reason to hope and expect that other (unmeasured) variables will be balanced, as well.

Without Randomization . . .

We still to compare groups who looked similar **before** they were exposed to our treatments.

- But we don't control the assignment of treatments.
  - Cannot use randomization to ensure comparability
- So how, then, do we make fair comparisons?
  - Analytical adjustments to account for baseline (covariate) differences in the groups.
  - A study is **biased** if the treatment groups differ in ways that matter for the outcome we're studying.

Benson and Hartz, NEJM 2002

A COMPARISON OF OBSERVATIONAL STUDIES AND RANDOMIZED, CONTROLLED TRIALS

KJELL BENSON, B.A., AND ARTHUR J. HARTZ, M.D., PH.D.

The fundamental criticism of observational studies is that unrecognized confounding factors may distort the results. According to the conventional wisdom, this distortion is sufficiently common and unpredictable that observational studies are not reliable and should not be funded. Our results suggest that observational studies usually do provide valid information.

Background For many years it has been claimed that observational studies find stronger treatment effects than randomized, controlled trials. We compared the results of observational studies with those of randomized, controlled trials.

Methods We searched the Abridged Index Medicus and Cochrane data bases to identify observational studies reported between 1985 and 1998 that compared two or more treatments or interventions for the same condition. We then searched the Medline and Cochrane data bases to identify all the randomized, controlled trials and observational studies comparing the same treatments for these conditions. For each treatment, the magnitudes of the effects in the various observational studies were combined by the Mantel–Haenszel or weighted analysis-of-variance procedure and then compared with the combined magnitude of the effects in the randomized, controlled trials that evaluated the same treatment.

Results There were 136 reports about 19 diverse treatments, such as calcium-channel–blocker therapy for coronary artery disease, appendectomy, and interventions for subfertility. In most cases, the estimates of the treatment effects from observational studies and randomized, controlled trials were similar. In only 2 of the 19 analyses of treatment effects did the combined magnitude of the effect in observational studies lie outside the 95 percent confidence interval for the combined magnitude in the randomized, controlled trials.

Conclusions We found little evidence that estimates of treatment effects in observational studies reported after 1984 are either consistently larger than or qualitatively different from those obtained in randomized, controlled trials. (N Engl J Med 2000;342: 1878-86.)

What do you expect the Figures to show?

Benson and Hartz, Figure 1

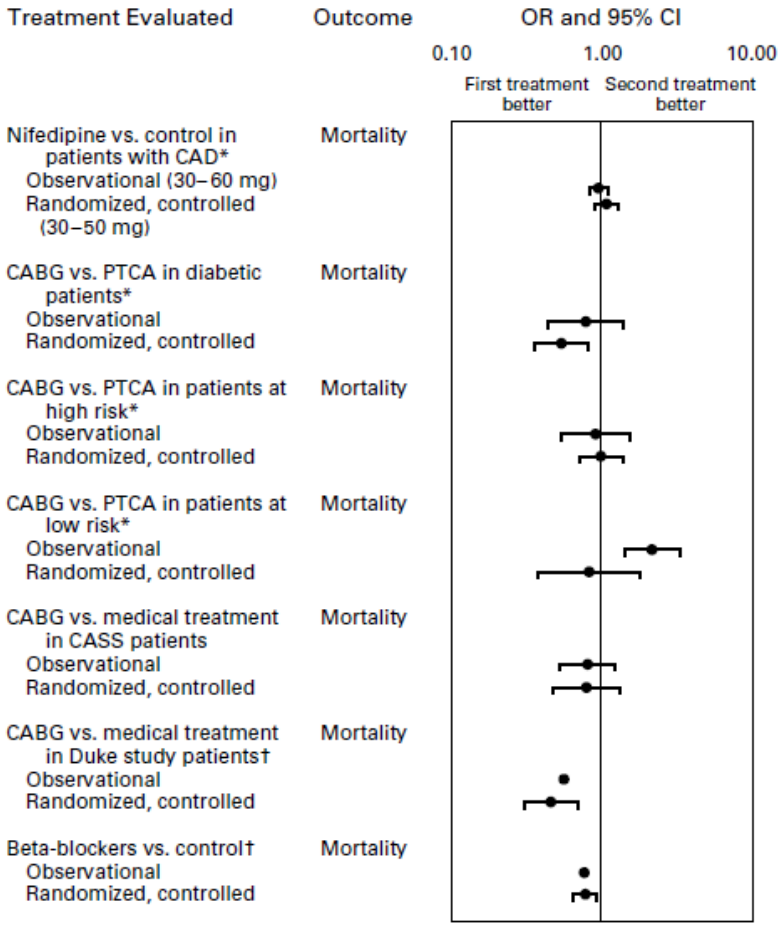


Figure 1. Results of Observational Studies and Randomized, Controlled Trials of Cardiologic Treatments.

The figure is based on data from eight articles.[13-20] Some articles contain data from more than one study. OR denotes odds ratio, CI confidence interval, CAD coronary artery disease, CABG coronary-artery bypass graft surgery, PTCA percutaneous transluminal coronary angioplasty, CASS Coronary Artery Surgery Study, and Duke the Duke University Cardiovascular Disease Databank. Asterisks indicate studies that reported relative risks rather than odds ratios. Daggers indicate studies that reported neither a confidence interval nor a P value for the odds ratio.

Impact of randomization?

Benson and Hartz, Figure 4

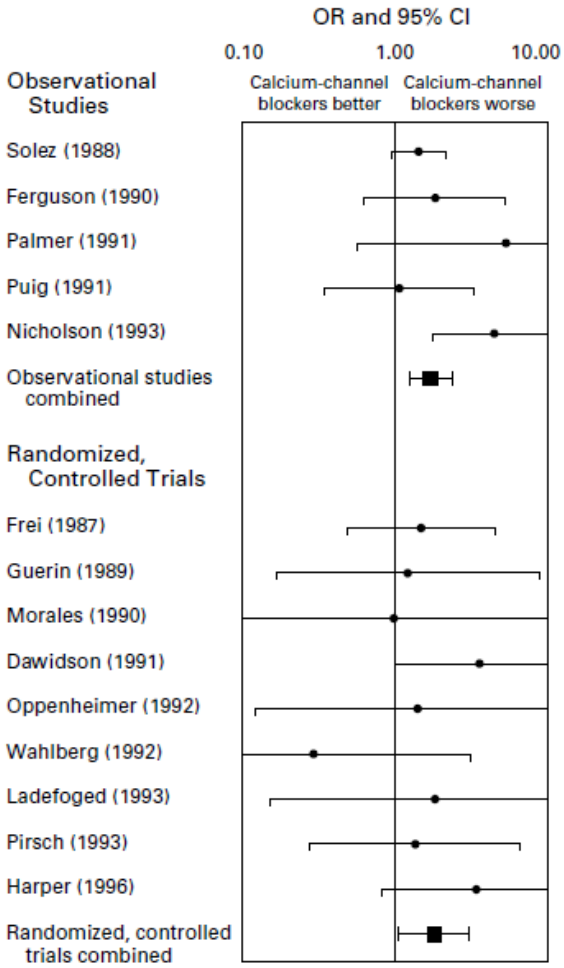


Figure 4. Odds Ratio for Graft Survival after Kidney Transplantation in Patients Receiving Calcium-Channel Blockers as Compared with Controls.

The figure is based on data from six articles.[58-63] The nine randomized, controlled trials were analyzed by Ladefoged and Andersen.[63] OR denotes odds ratio, and CI confidence interval.

What can we (and *can't we*) conclude in this setting?

Benson and Hartz, Discussion (p. 1884)

Our finding that observational studies and randomized, controlled trials usually produce similar results **differs from the conclusions of previous authors**.

A study *(Chalmers et al NEJM, 1977)* reviewed the evidence of the effectiveness of anticoagulants in the treatment of acute myocardial infarction, using eight observational studies and six randomized, controlled trials.

- The differences in mortality rates between control and treatment groups were larger in the observational studies than in the randomized, controlled trials.
  - The observational studies reviewed were published before 1975, and
  - the authors did not use current meta-analytic techniques for pooling data.
- The results of the comparison might have differed if current *(i.e. 2002)* methods had been used to combine the results of several trials.

Benson and Hartz, Discussion (continuing)

Three other studies[1] commonly cited to show the inadequacy of observational data, as well as one that found no bias in observational data[2], also compared observational studies and randomized, controlled trials that **evaluated different treatments**.

As compared with these previous studies, our study has the advantage that the comparisons were stratified according to treatment.

In addition, the studies that we reviewed were more recent and therefore may have used better methods than those in the earlier reviews.

[1] Sacks et al Am H Med 1982, Colditz et al Stat Med 1989, Miller et al Stat Med 1989

[2] Ottenbacher Controlled Clin Trials 1992

Benson and Hartz, Discussion (conclusion)

The fundamental criticism of observational studies is that unrecognized confounding factors may distort the results. According to the conventional wisdom, this distortion is sufficiently common and unpredictable that observational studies are not reliable and should not be funded. Our results suggest that observational studies usually do provide valid information.

They could be used to exploit the many recently developed, clinically rich data bases. Only with a greater willingness to analyze these data bases is it possible to achieve a realistic understanding of how observational studies can best be used.

Smith and Pell, BMJ 2003

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

Smith and Pell, BMJ 2003

HULTON/GETTY

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

It is a truth universally acknowledged that a medical intervention justified by observational data must be in want of verification through a randomised controlled trial.

Observational studies have been tainted by accusations of data dredging, confounding, and bias.

- For example, observational studies showed lower rates of ischaemic heart disease among women using hormone replacement therapy, and these data were interpreted as advocating hormone replacement for healthy women, women with established ischaemic heart disease, and women with risk factors for ischaemic heart disease.
- However, randomised controlled trials showed that hormone replacement therapy actually increased the risk of ischaemic heart disease, indicating that the apparent protective effects seen in observational studies were due to bias.

Cases such as this one show that medical interventions based solely on observational data should be carefully scrutinised, and the parachute is no exception.

The "Healthy Cohort" Effect, from Smith and Pell

One of the major weaknesses of observational data is the possibility of bias, including selection bias and reporting bias, which can be obviated largely by using randomised controlled trials. The relevance to parachute use is that individuals jumping from aircraft without the help of a parachute are likely to have a high prevalence of pre-existing psychiatric morbidity.

Individuals who use parachutes are likely to have less psychiatric morbidity and may also differ in key demographic factors, such as income and cigarette use.

It follows, therefore, that the apparent protective effect of parachutes may be merely an example of the "**healthy cohort**" effect. Observational studies typically use multivariate analytical approaches, using maximum likelihood based modelling methods to try to adjust estimates of relative risk for these biases.

Distasteful as these statistical adjustments are for the cognoscenti of evidence based medicine, no such analyses exist for assessing the presumed effects of the parachute.

from Smith and Pell

What is already known about this topic

Parachutes are widely used to prevent death and major injury after gravitational challenge

---

Parachute use is associated with adverse effects due to failure of the intervention and iatrogenic injury

---

Studies of free fall do not show 100% mortality

What this study adds

No randomised controlled trials of parachute use have been undertaken

---

The basis for parachute use is purely observational, and its apparent efficacy could potentially be explained by a "healthy cohort" effect

---

Individuals who insist that all interventions need to be validated by a randomised controlled trial need to come down to earth with a bump

A call to (broken) arms

Only two options exist.

The first is that we accept that, under exceptional circumstances, common sense might be applied when considering the potential risks and benefits of interventions.

The second is that we continue our quest for the holy grail of exclusively evidence based interventions and preclude parachute use outside the context of a properly conducted trial.

The dependency we have created in our population may make recruitment of the unenlightened masses to such a trial difficult. If so, we feel assured that those who advocate evidence based medicine and criticise use of interventions that lack an evidence base will not hesitate to demonstrate their commitment by volunteering for a double blind, randomised, placebo controlled, crossover trial.

Smith and Pell, 2003

Contributors

GCSS had the original idea. JPP tried to talk him out of it.

JPP did the first literature search but GCSS lost it.

GCSS drafted the manuscript but JPP deleted all the best jokes.

GCSS is the guarantor, and JPP says it serves him right.

USPSTF Grade Definitions

Grade	Definition	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer or provide this service.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.	Offer or provide this service.
C	The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small.	Offer or provide this service for selected patients depending on individual circumstances.
D	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.	Discourage the use of this service.
I Statement	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.	Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

https://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions

Levels of Certainty (USPSTF)

Levels of Certainty Regarding Net Benefit

Level of Certainty*	Description
High	The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies.
Moderate	The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as: • The number, size, or quality of individual studies. • Inconsistency of findings across individual studies. • Limited generalizability of findings to routine primary care practice. • Lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.
Low	The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of: • The limited number or size of studies. • Important flaws in study design or methods. • Inconsistency of findings across individual studies. • Gaps in the chain of evidence. • Findings not generalizable to routine primary care practice. • Lack of information on important health outcomes. More information may allow estimation of effects on health outcomes.

*The USPSTF defines certainty as "likelihood that the USPSTF assessment of the net benefit of a preventive service is correct." The net benefit is defined as benefit minus harm of the preventive service as implemented in a general, primary care population. The USPSTF assigns a certainty level based on the nature of the overall evidence available to assess the net benefit of a preventive service.

Simple Observational Studies, Again

- We have an outcome measured on two groups of subjects (treated and control).
- We want to make a fair comparison between the treated group and the control group in terms of the outcome.
- We can obtain covariates that describe the subjects before they received treatments, but we **can't ensure** that the groups will be comparable in terms of the covariates.

The Key Role of Assumptions

We'd like to describe cause-effect relationships from non-experimental data. This is challenging.

> *. . . the elucidation of causal relationships from observational studies must be shaped by knowledge (or assumptions) about how the data were generated; such assumptions are crucial to causal inference.*

- Judea Pearl (2000) Causal Inference in the Health Sciences: A Conceptual Introduction *Health Services & Outcomes Research Methodology* 2: 189-220.

How Randomization Works

1. Identify experimental units.
   - Inferences refer only to these units, typically.
2. Define a collection of possible assignments of treatments to units.
   - Exclude unreasonable assignments from the collection.
3. Define a stochastic mechanism for selecting one assignment from the collection.
   - Complete randomization vs. Blocked randomization
   - Biased coin / "balancing" randomization
4. Select one assignment from the collection using the mechanism.
5. Use the stochastic mechanism as the sole basis for inference.

Randomized vs. Non-Randomized Studies

- In a non-randomized study, we'd no longer KNOW the distribution of treatment assignments.
- We need to make some assumption about the distribution in order to make inferences.
- Moreover, there may be little basis on which to ground or defend this assumption. It may be wrong, or open to challenge.

The Role of Assumptions

Scenario	Goal of Analysis	Role of Assumptions
Randomized Experiments	Testing H_0: No treatment effect	None

The Role of Assumptions

Scenario	Goal of Analysis	Role of Assumptions
Randomized Experiments	Testing H_0: No treatment effect	None
Randomized Experiments	Estimating treatment effects, CIs	Minor

The Role of Assumptions

Scenario	Goal of Analysis	Role of Assumptions
Randomized Experiments	Testing H_0: No treatment effect	None
Randomized Experiments	Estimating treatment effects, CIs	Minor
Observational Studies	Anything	**MASSIVE**

Why are Experiments *Better* Than Observational Studies?

Scientific questions are not settled on a particular date by a single event. Rather, we speak of the "weight of evidence."

- Experiments leave fewer grounds for doubt.
- Experiments often settle questions faster.
- Uncertainty about treatment effects is greater in the absence of randomization.
- With observational studies, we are especially concerned about sensitivity to hidden bias.

Advantages of Smart Observational Studies

- Address chief criticism of randomized trials: limited generalizability / external validity
- Enable examination of exposure in "real life"
- May enable examination of effect size and "entrenched practices"
- Broader array of exposures and outcomes can be explored with an observational study than in controlled experiments.
- Due to frequently large size, can provide information about exposures with small effect sizes (toxicity of treatments)
- Data are widely (increasingly) available
- Often reduced cost and time to get answer

BUT No randomization forces the investigator to think hard about **how exposures were assigned** or determined.

Rosenbaum Example 1. Vitamin C and the treatment of advanced cancer

The observational study in question (Pauling et al.) used 100 patients (cases given large vitamin C doses) believed to be terminally ill from advanced cancer matched to 10 historical controls each.

- As a group, vitamin C patients survived 4x as long.

Rosenbaum Example 1. Vitamin C and the treatment of advanced cancer

The observational study in question (Pauling et al.) used 100 patients (cases given large vitamin C doses) believed to be terminally ill from advanced cancer matched to 10 historical controls each.

- As a group, vitamin C patients survived 4x as long.
- But a Mayo Clinic RCT showed no difference, with the placebo group surviving slightly longer [NS].

Rosenbaum Example 1. Vitamin C and the treatment of advanced cancer

The observational study in question (Pauling et al.) used 100 patients (cases given large vitamin C doses) believed to be terminally ill from advanced cancer matched to 10 historical controls each.

- As a group, vitamin C patients survived 4x as long.
- But a Mayo Clinic RCT showed no difference, with the placebo group surviving slightly longer [NS].
- Why did this happen? If OS and RCT conflict, which should we believe?

Characteristics of Excellent Observational Studies

- Careful choice of research hypothesis: narrow, controlled examination of a broad theory
- Use of a control group (subjects who did not receive the treatment) carefully selected
- Careful choice of treatment: Sharply distinct treatments that could happen to anyone
- Competing **theories**, not just H_0 and H_A: desirability of multiple working hypotheses.

Rosenbaum Example 2. Smoking and Heart Disease

An observational study looks at "age-adjusted" mortality rates from heart disease of British doctors with various smoking behaviors.

Specifying an **Elaborate Theory**: What should we see if smoking causes coronary disease?

Rosenbaum Example 2. Smoking and Heart Disease

An observational study looks at "age-adjusted" mortality rates from heart disease of British doctors with various smoking behaviors.

Specifying an Elaborate Theory: What should we see if smoking causes coronary disease?

- Is there a dose-response effect?
  - Non- vs. Light vs. Moderate vs. Heavy?
  - Where does quitting fit in?
- What does the elaborate theory suggest?
- Can we use design and data to test multiple elements of the theory at once?

Rosenbaum Example 3. DES and Vaginal Cancer

Could diethylstilbestrol (DES), given to pregnant women, cause vaginal cancer in their daughters?

- Match 8 cases to four referents each born at about the same time in same hospital, and on same service.
- Compare DES use of mothers of cases to referents?
  - 7 of 8 cases had moms on DES, 0 of 32 referents did.

Key issue: **Sensitivity to Bias**: how severe would the unseen problems in this study have to be to produce such a relationship if DES were in fact harmless?

Rosenbaum Example 4. Achievement in Public / Catholic High Schools

Responding to Criticisms

- Aim for tangible, specific and plausible alternative interpretations of the available data
- Cochran: "The investigator should always list and discuss all alternative explanations of his results (including different hypotheses and biases in the results) that occur to him."

Do Catholic schools eliminate difficult students (by expelling them) while public schools don't?

Characteristics of Excellent Observational Studies: A Reminder

- Careful choice of research hypothesis: narrow, controlled examination of a broad theory
- Use of a control group (subjects who did not receive the treatment) carefully selected
- Careful choice of treatment: Sharply distinct treatments that could happen to anyone
- Competing **theories**, not just H_0 and H_A: desirability of multiple working hypotheses.

Next Class - Thursday 2017-01-25 at 8:30 AM

- Rosenbaum (2017) Chapters 1-4 (Part I. Randomized Experiments)
  1. A Randomized Trial
  2. Structure
  3. Causal Inference in Randomized Experiments
  4. Irrationality and Polio
- Homework 1 discussion
  1. Mock Proposals from the DIG study (we'll share)
  2. Analyzing Data - building a logistic regression model
- How Can We Avoid Being Misled by Observational Studies?
  - What is **selection bias** and why should I care about it?
  - What can be done to deal with selection bias in observational studies?