500 Class 01 (Recording)

https://thomaselove.github.io/500-2025/

2025-01-16

Agenda for Recorded Lecture

- Welcome to 500!
- 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?
- The STROBE Guidelines: Items 1-12

Agenda for Zoom Call

Thursday 2025-01-16 from 10 to 11 AM. Zoom details in your email and on Canvas.

- Welcome to 500
- Logistics of the Course
- A Motivating Example: Aspirin and Mortality in Heart Patients
  - How can we avoid being misled?
  - Causal Effects as comparing potential outcomes

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 and Other Techniques
- Using R, RStudio and Quarto to accomplish all of this

Paul Rosenbaum's 2023 book *Causal Inference*

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.

This Year is Unusual

- Classes 1-8 (before Spring Break) involve:
  - a recorded 60-90 minute lecture (like this one)
  - a Zoom meeting to discuss the lecture and other issues from 10-11 AM on Thursday
- After Spring Break, starting with Class 9 (2025-03-20) we will meet in person from 8:30 AM to 11:15 AM in Wolstein Research Building room 1217 on the CWRU campus.
- For all sessions, Dr. Love is available after class for informal "office hours".
- Recordings of class sessions will be available through Zoom on Canvas when things work properly.

The Web Site

https://thomaselove.github.io/500-2025 is at the bottom of every slide

- Syllabus
- Calendar
  - links to class sessions, final word on all deadlines
- R and Data
  - Installing/Updating R, RStudio, R Packages
  - Data and Code
- Sources / References
  - Some things are **password-protected**.
- Links to Canvas, and to ways to Contact Us
- Assignments, which we'll discuss in class.

Section 1

What is this course about?

Comparative Effectiveness Studies

- We have an outcome measured on two groups of subjects
  - Let's refer to them as the treated group and the control group.
- 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.

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.

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.

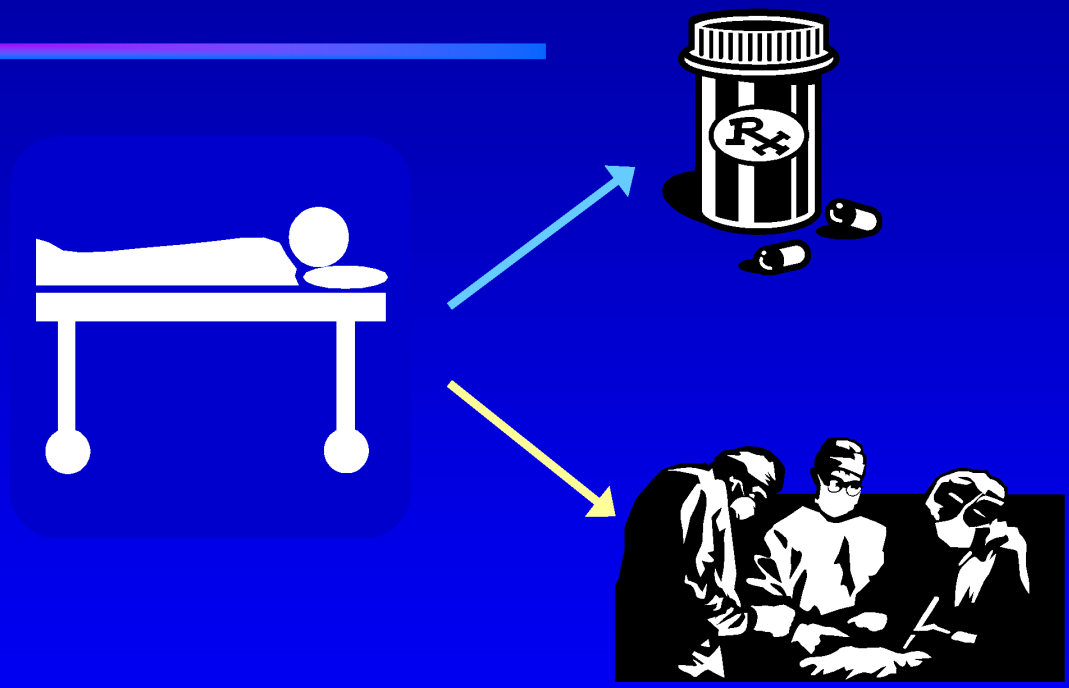
A study without a treatment is neither an experiment nor an observational study.

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.

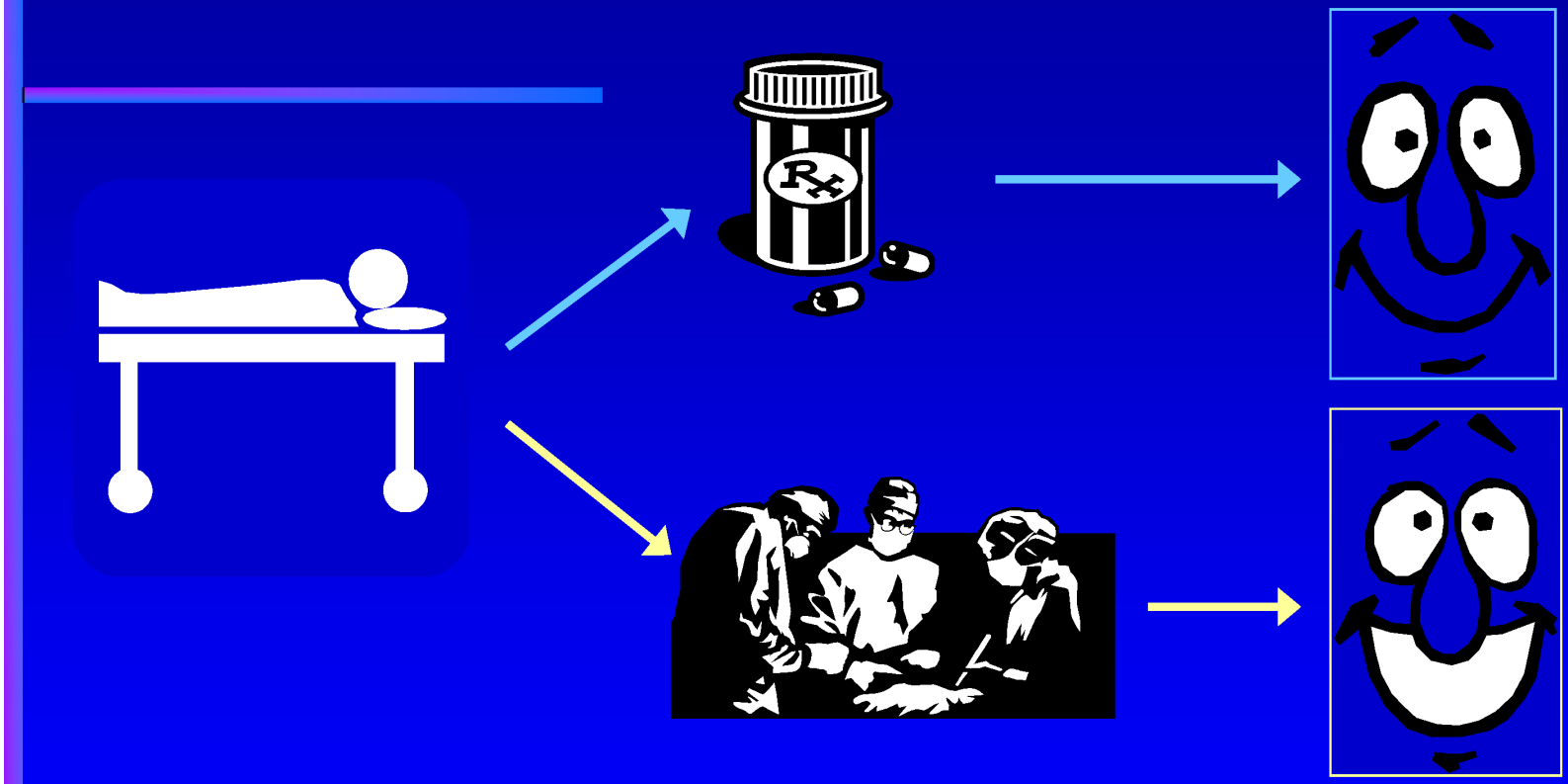
Rosenbaum 2002 Observational Studies, Chapter 1

Looking for Causal Treatment Effects

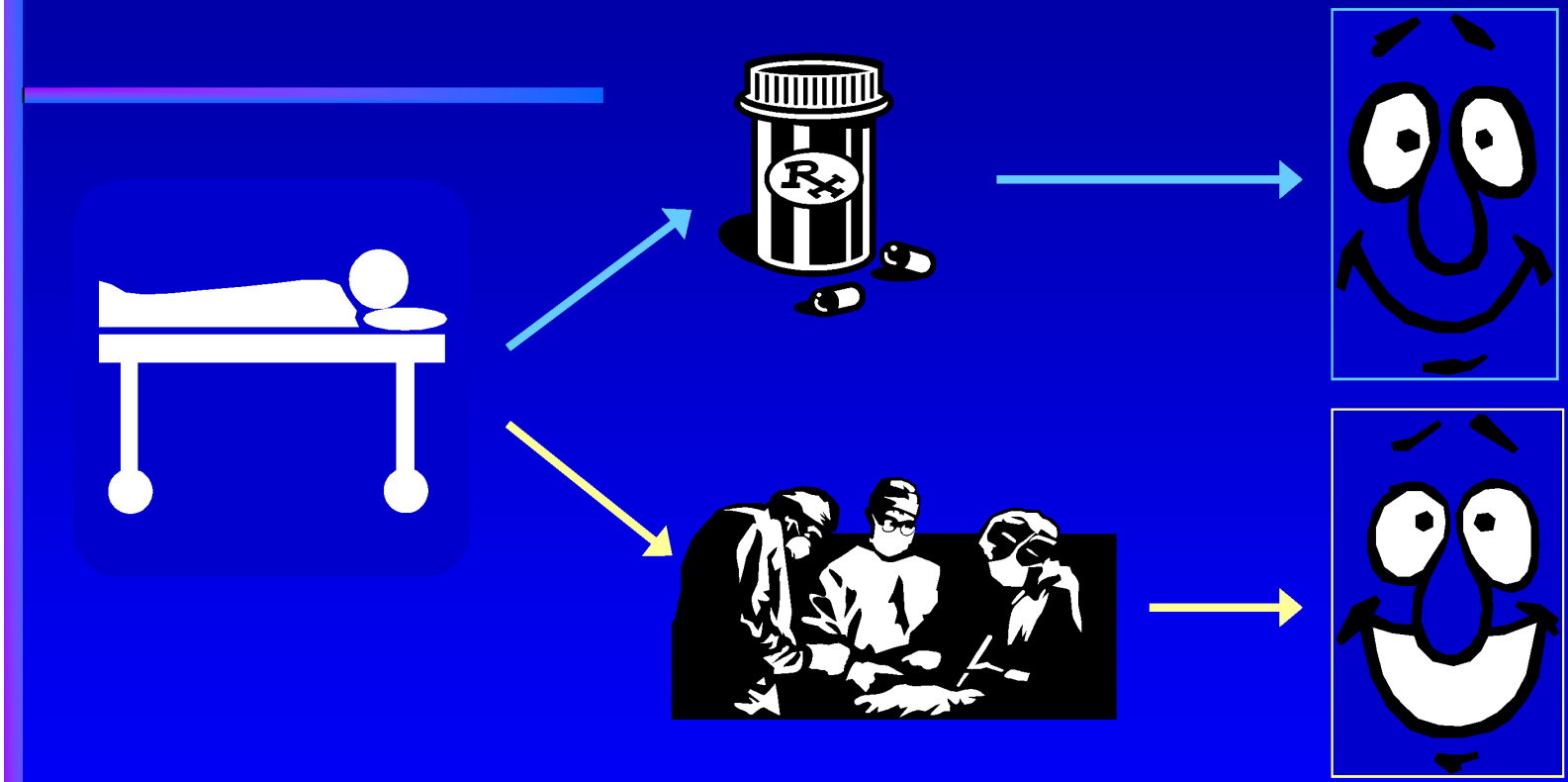
Looking for Causal Treatment Effects



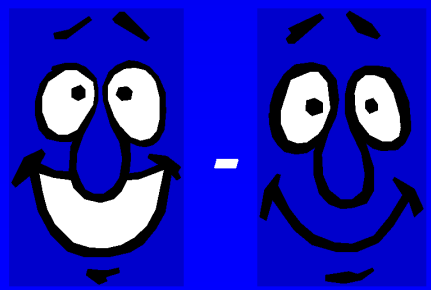
Looking for Causal Treatment Effects



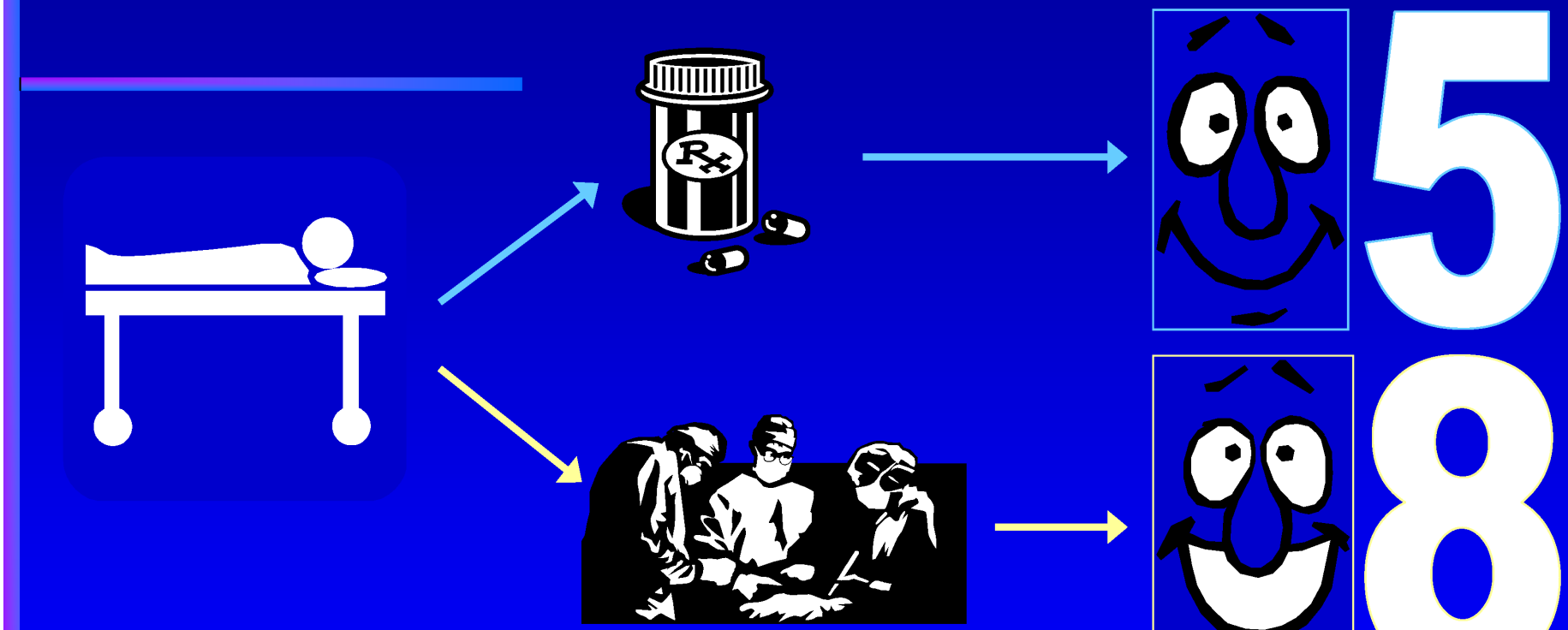
Looking for Causal Treatment Effects



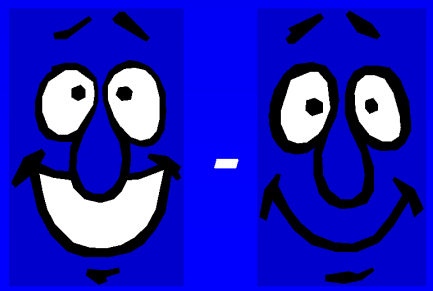
Surgery-Drug Treatment Effect =



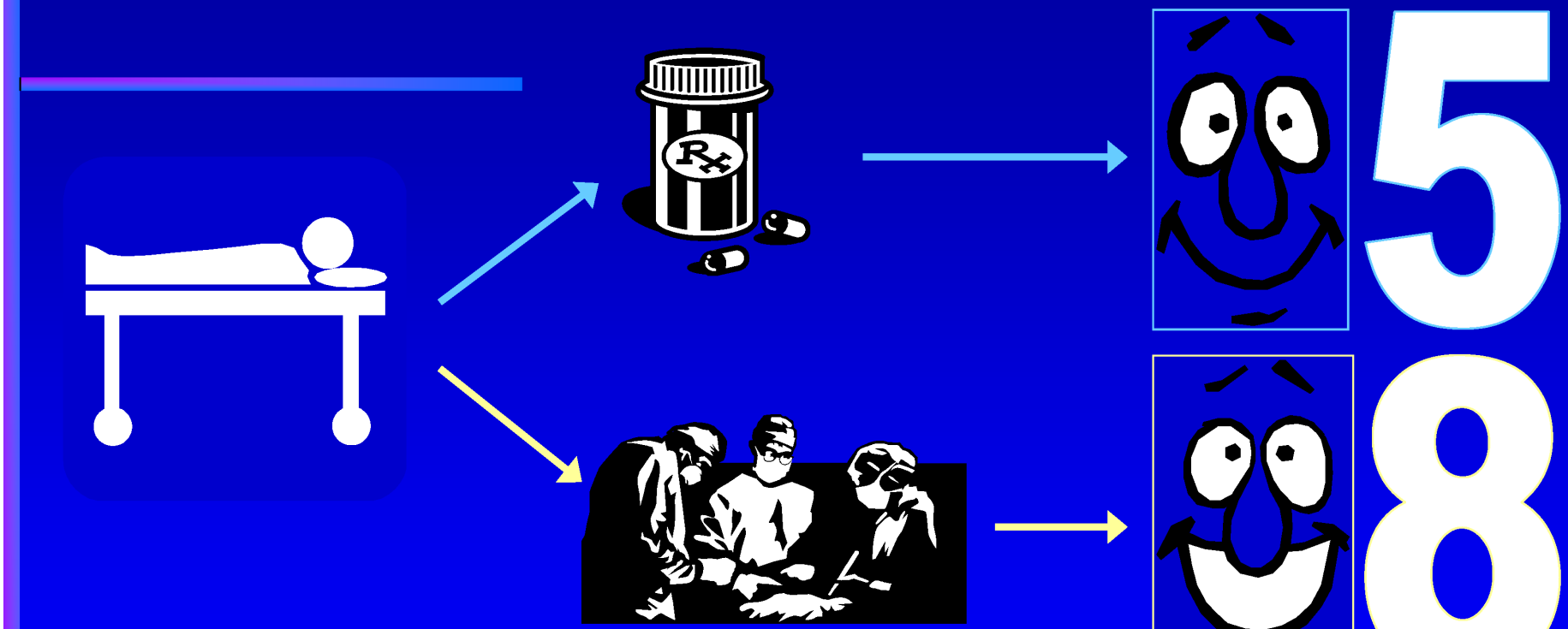
Looking for Causal Treatment Effects



Surgery-Drug Treatment Effect =

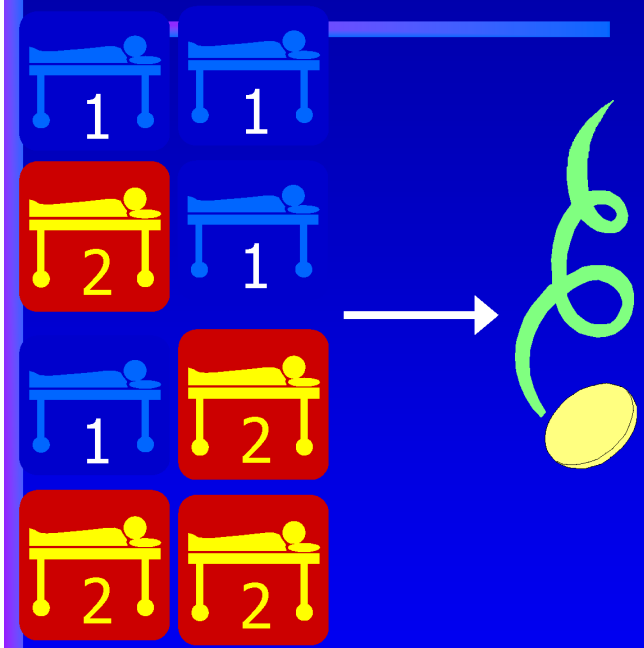


Looking for Causal Treatment Effects



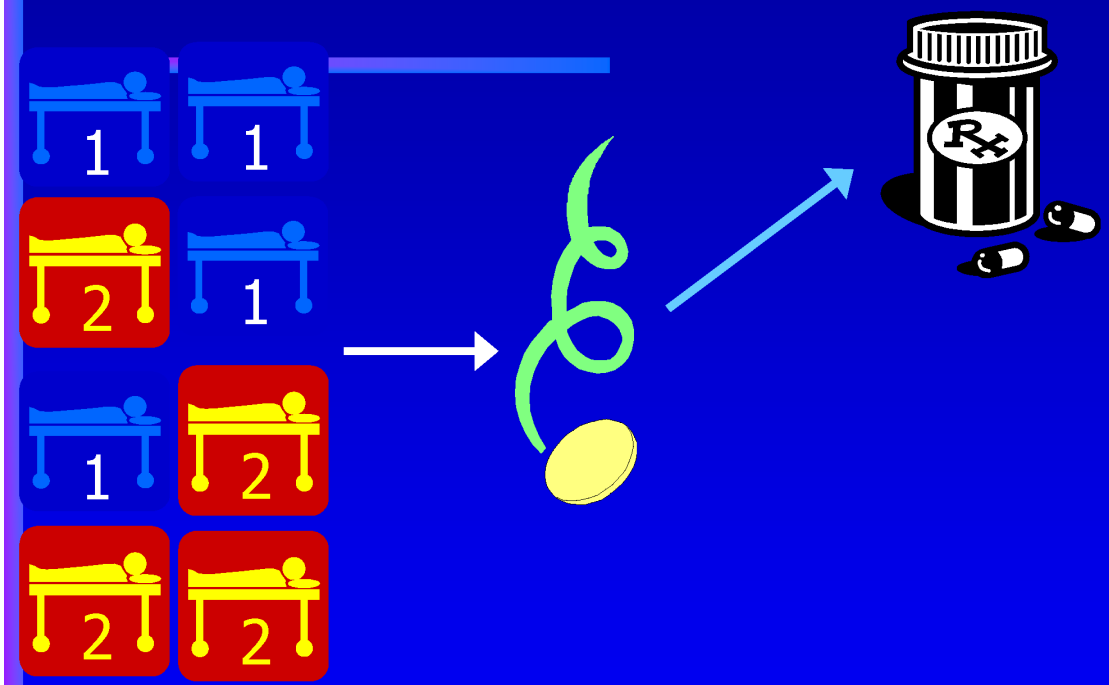
Surgery-Drug Treatment Effect = 3

Importance of Randomization



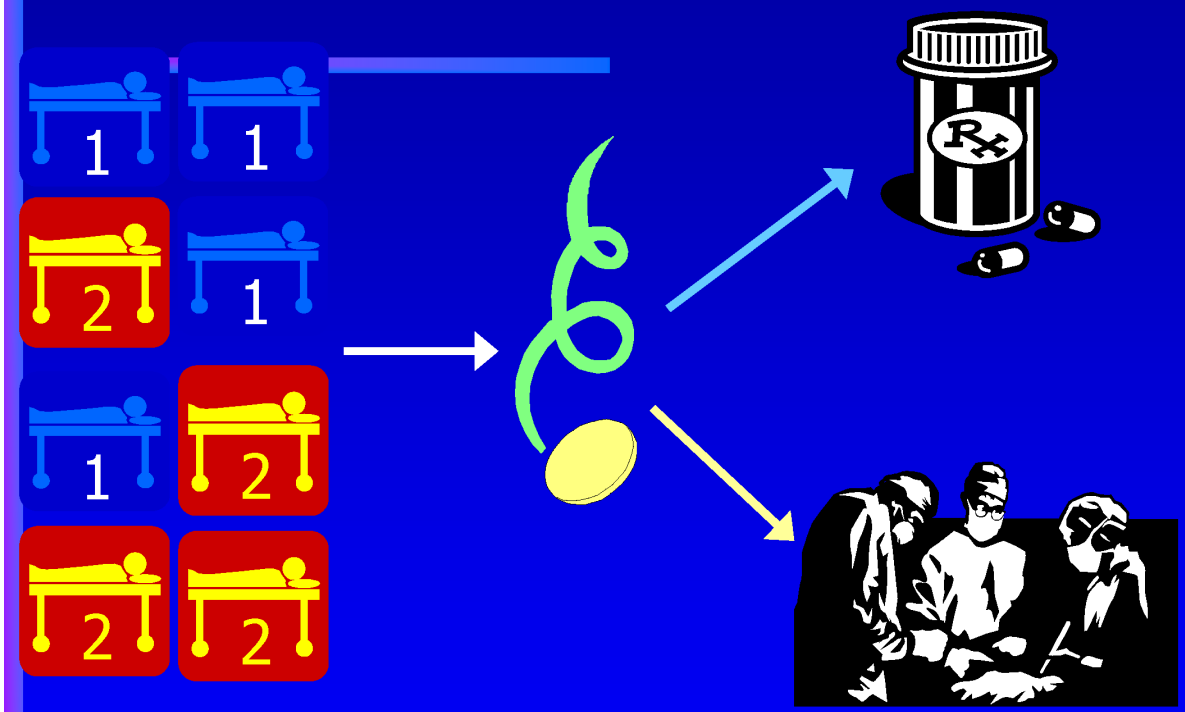
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



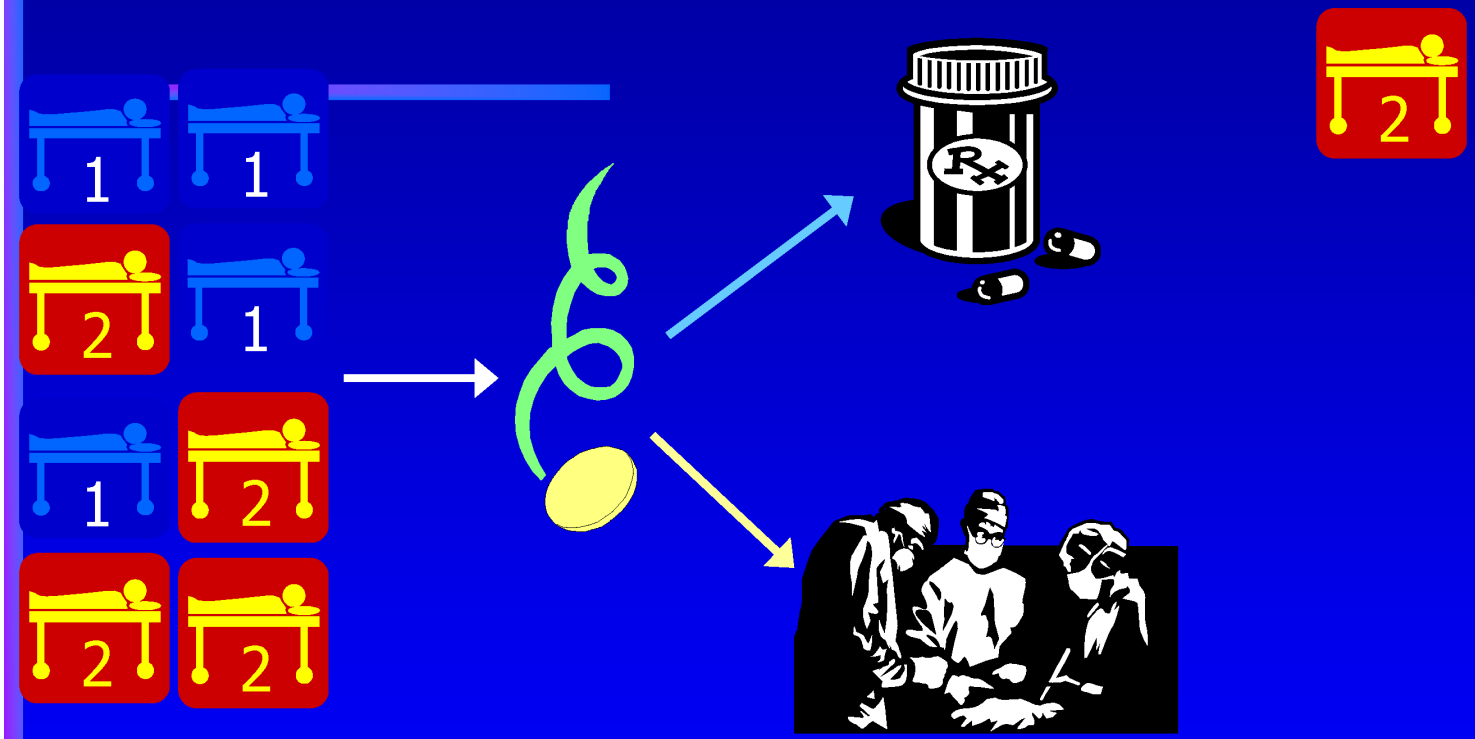
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



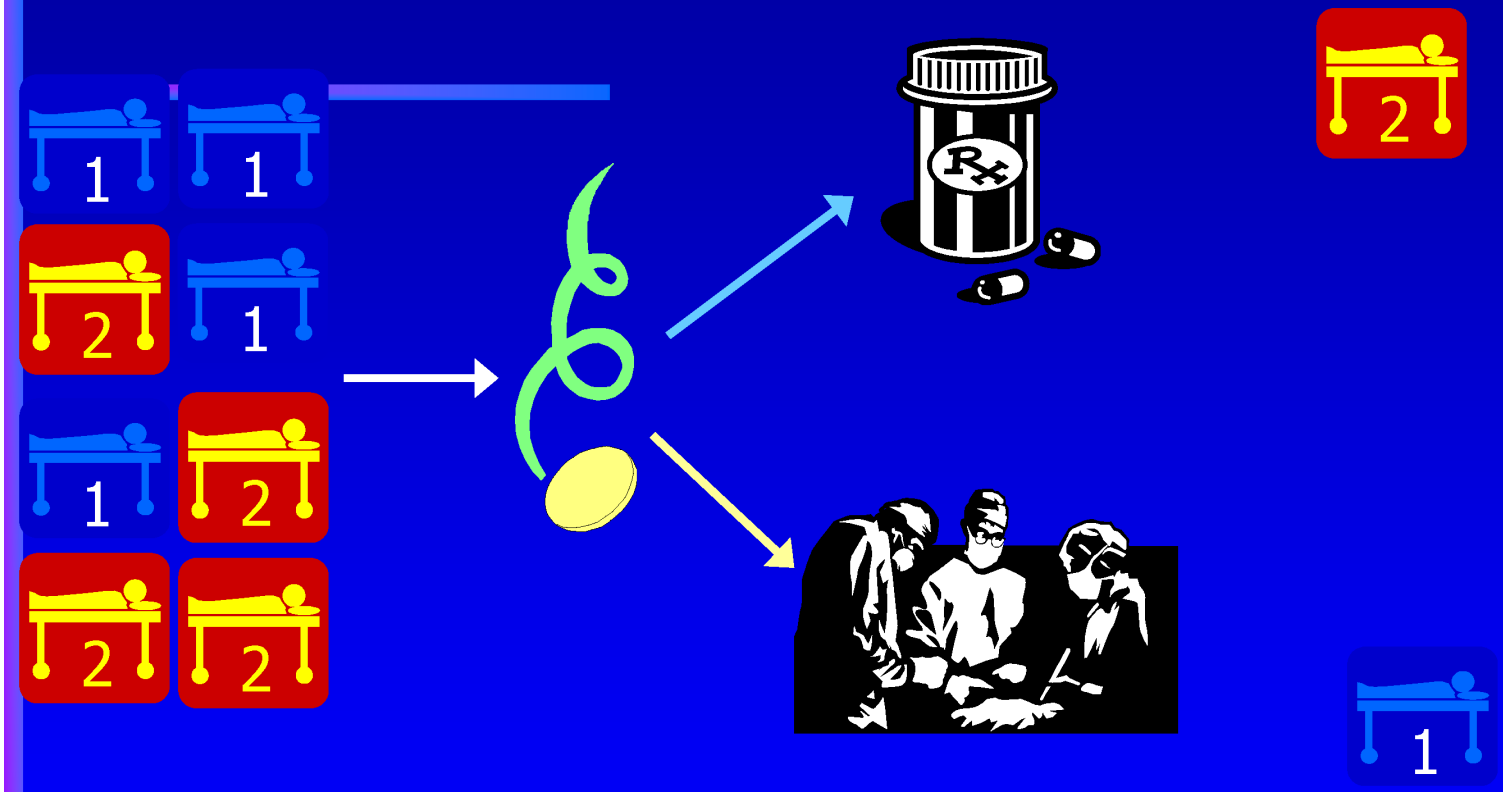
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



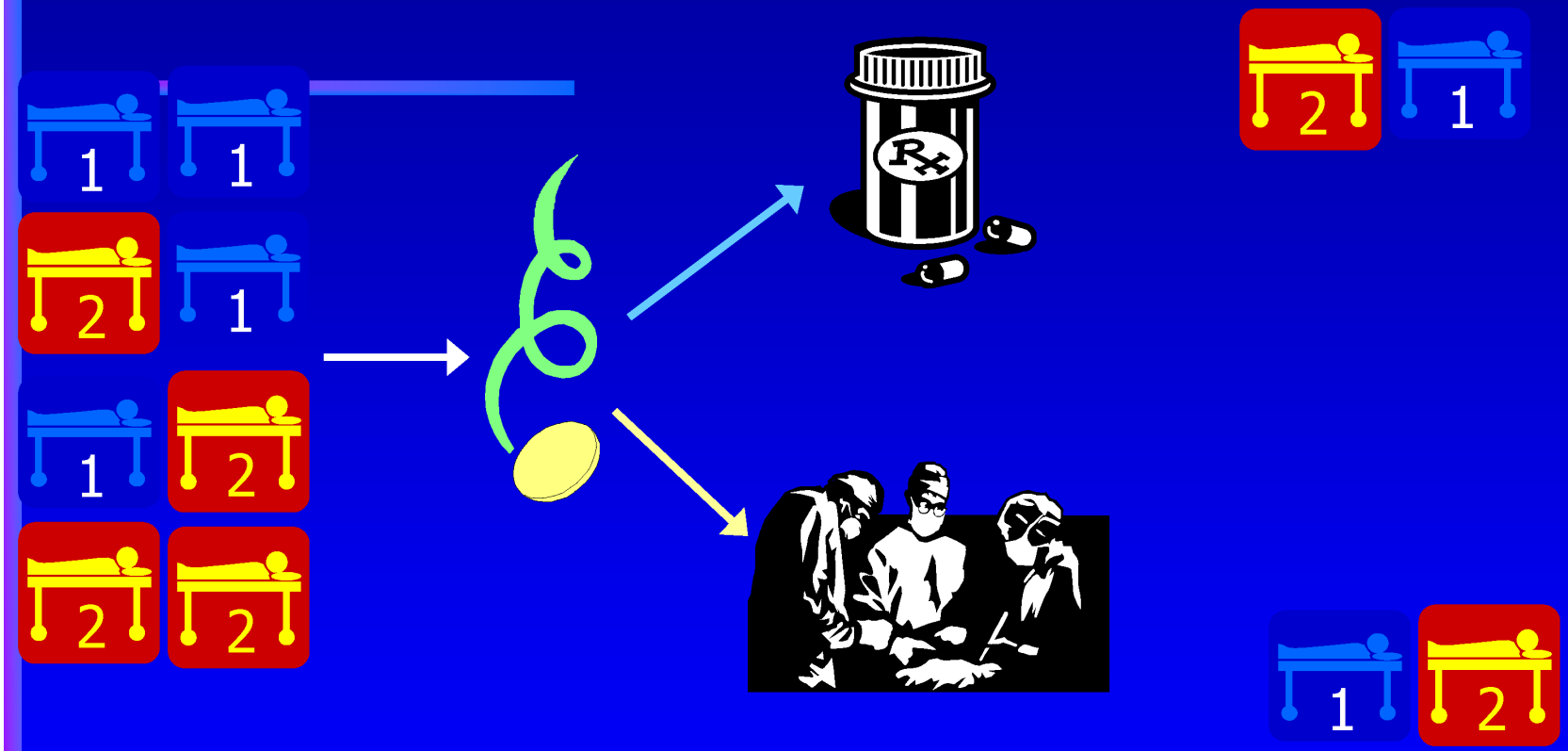
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



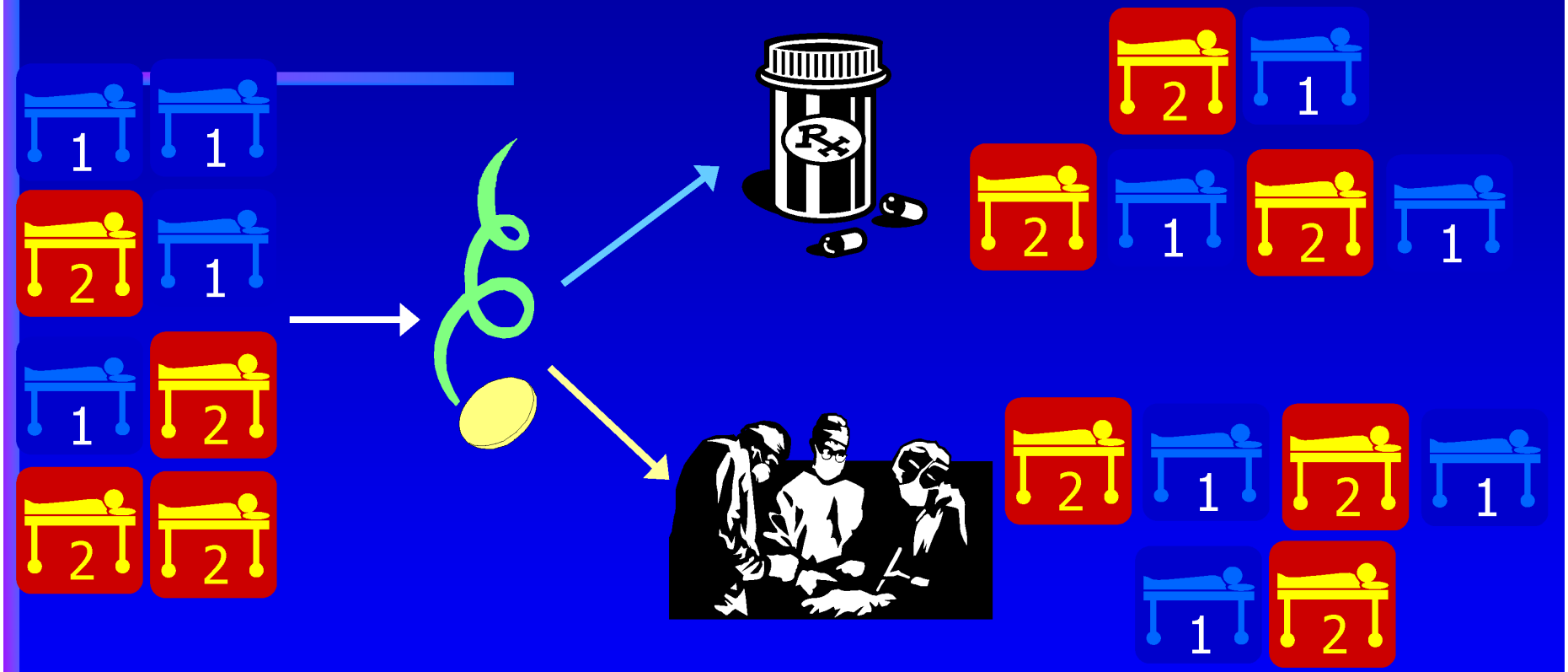
Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



Randomization ensures that subjects receiving different treatments are comparable.

Importance of Randomization



Randomization ensures that subjects receiving different treatments are comparable.

USPSTF Evidence Grades, 2000 (see Concato)

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.

Section 2

Thinking about Randomized Clinical Trials

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.

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.

Note that a 1984 NEJM follow-up (11-year survival) is available at https://pubmed.ncbi.nlm.nih.gov/6333636/.

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.

It's a false choice - it's not really possible to do "only" experiments or "only observational studies.

Why does observational inference matter?

Adapted from Emily Riederer (see Sources page)

- Even when you can experiment, understanding observational causal inference can help you better identify biases and design your experiments
- Testing can be expensive.
  - There are direct costs of instituting a policy that might not be effective, implementation costs, and opportunity costs (holding out a control group and not applying what you hope to be a good strategy as broadly as possible.)
- Randomized experimentation is harder than it sounds! Sometimes experiments may not go as planned, but treating the results as observational data may help salvage some information value.
- Data collection can take time. When we long to read an experiment that wasn't launched three years ago, historical observational data can help us get a preliminary answer sooner.

Sometimes, we have other problems...

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)[1]

> *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)

[1] Multiple risk factor intervention trial. Risk factor changes and mortality results. Multiple Risk Factor Intervention Trial Research Group. JAMA . 1982 Sep 24;248(12):1465-77. https://pubmed.ncbi.nlm.nih.gov/7050440/

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)[1]

> *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.

[1] Multiple risk factor intervention trial. Risk factor changes and mortality results. Multiple Risk Factor Intervention Trial Research Group. JAMA . 1982 Sep 24;248(12):1465-77. https://pubmed.ncbi.nlm.nih.gov/7050440/

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)[1]

> *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

[1] Multiple risk factor intervention trial. Risk factor changes and mortality results. Multiple Risk Factor Intervention Trial Research Group. JAMA . 1982 Sep 24;248(12):1465-77. https://pubmed.ncbi.nlm.nih.gov/7050440/

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)[1]

> *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.

[1] Multiple risk factor intervention trial. Risk factor changes and mortality results. Multiple Risk Factor Intervention Trial Research Group. JAMA . 1982 Sep 24;248(12):1465-77. https://pubmed.ncbi.nlm.nih.gov/7050440/

The MRFIT Trial

Multiple Risk Factor Intervention Trial (*JAMA* 1982)[1]

> *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.)

[1] Multiple risk factor intervention trial. Risk factor changes and mortality results. Multiple Risk Factor Intervention Trial Research Group. JAMA . 1982 Sep 24;248(12):1465-77. https://pubmed.ncbi.nlm.nih.gov/7050440/

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.

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

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 randomized 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 modeling 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.

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.

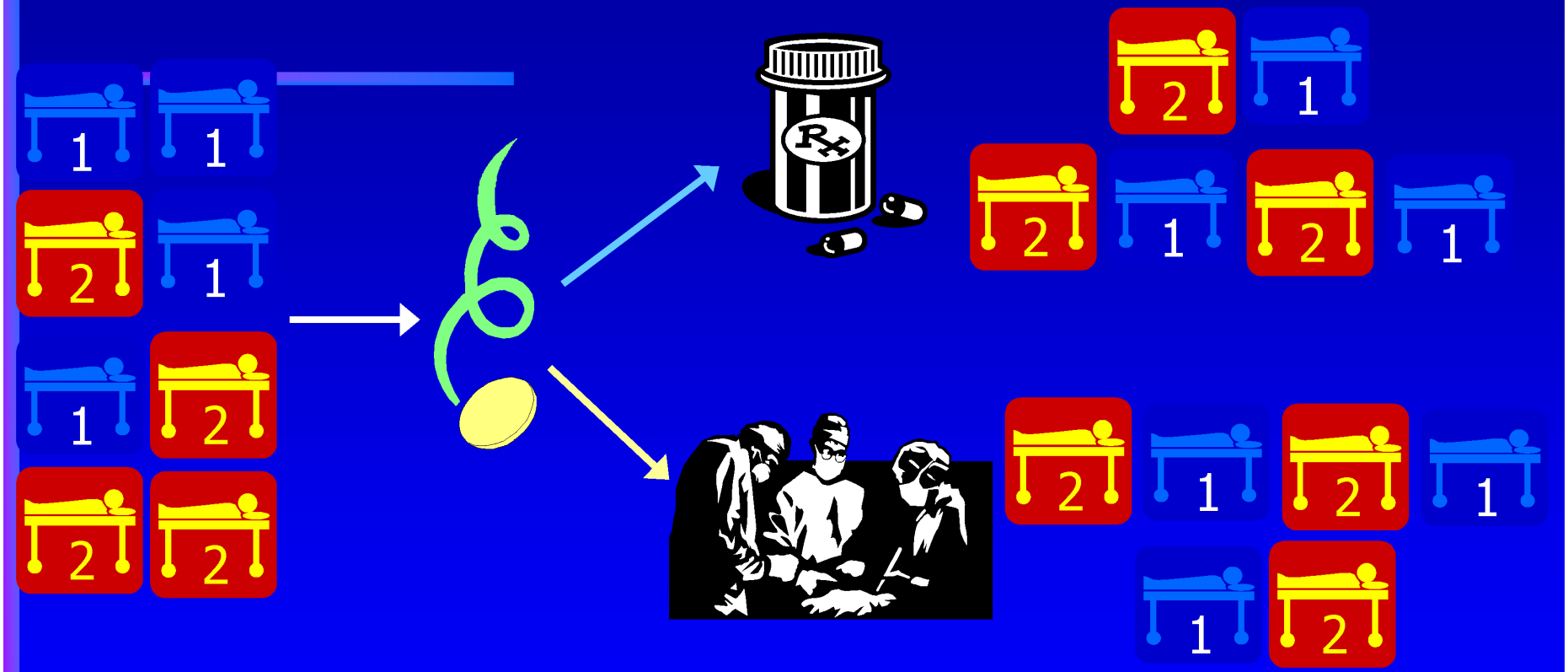
GCSS drafted the manuscript but JPP deleted all the best jokes.

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

Section 3

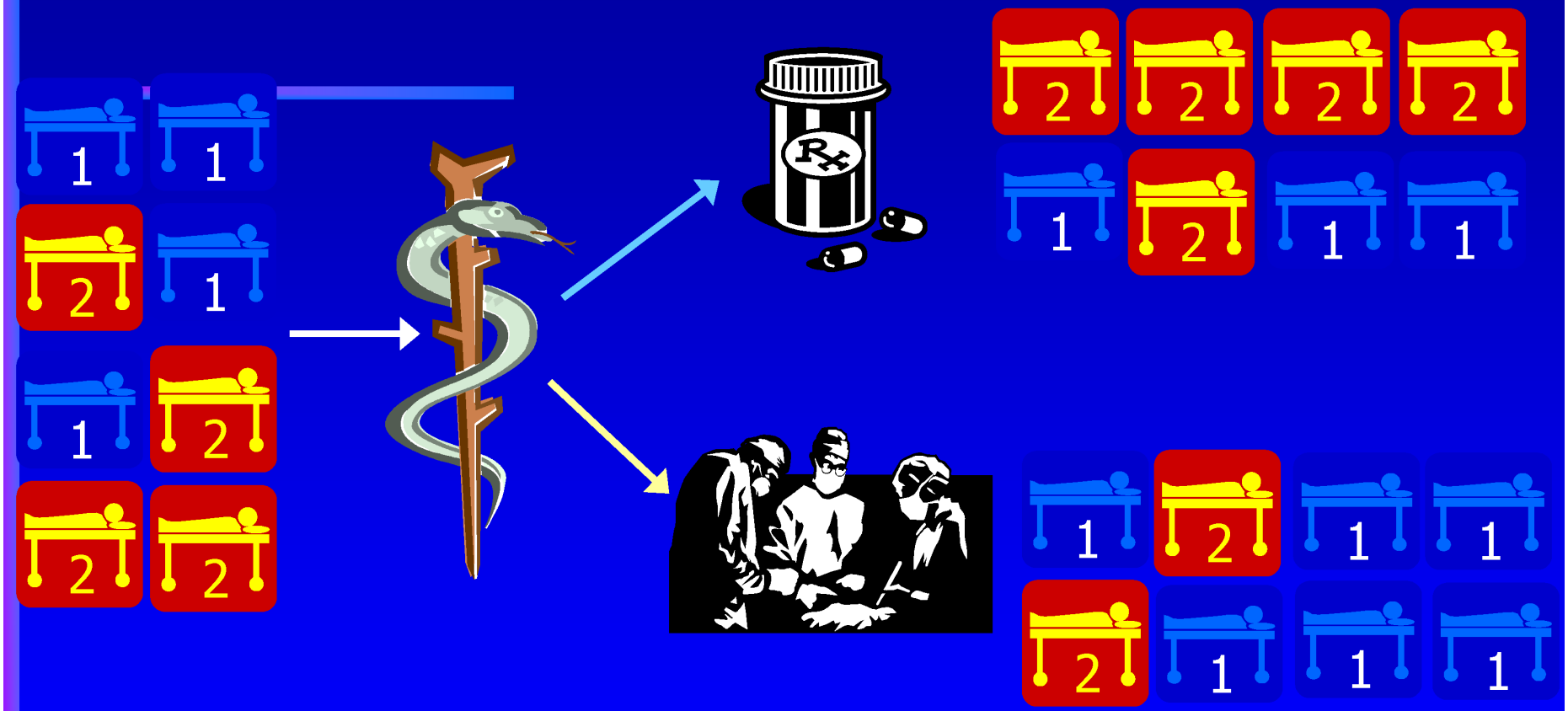
Thinking More About Observational Studies

Importance of Randomization



Randomization ensures that subjects receiving different treatments are comparable.

Observational Studies



In an observational study, the researcher does not randomly allocate the treatments.

Without Randomization ...

We still want 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.

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.

USPSTF Grade Definitions (2012)

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 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

- 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.

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.

- You might be interested as well in The Book of Why by Judea Pearl and Dana Mackenzie.

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.

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.

Section 4

The STROBE Guidelines for Reporting Observational Studies

STROBE Guidelines

https://www.strobe-statement.org/

> *STROBE stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of STrengthening the Reporting of OBservational studies in Epidemiology.*

STROBE Checklist

Checklist of items that should be included in reports of observational studies

- Title and Abstract (item 1)
- Introduction (items 2-3)
- Methods (items 4-12)
- Results (items 13-17)
- Discussion (items 18-21)
- Other Information (funding: item 22)

18 items are common to all three study designs and four (6, 12, 14 and 15) are specific for cohort, case-control, or cross-sectional studies.

We'll briefly discuss items 1-12 today, leaving 13-22 for next time.

STROBE Items 1-3

Title and abstract

1 (a) Indicate the study's design with a commonly used term in the title or the abstract 1 (b) Provide in the abstract an informative and balanced summary of what was done and what was found.

Background/rationale

2 Explain the scientific background and rationale for the investigation being reported

Objectives

3 State specific objectives, including any pre-specified hypotheses

STROBE Items 4-5

Study design

4 Present key elements of study design early in the paper

Setting

5 Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection

STROBE Item 6: Participants

(a) Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants.

(b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed. Case-control study—For matched studies, give matching criteria and the number of controls per case.

STROBE Items 7-9

Variables

7 Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable

Data sources/ measurement

8 For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group

Bias

9 Describe any efforts to address potential sources of bias

STROBE Items 10-11

Study size

10 Explain how the study size was arrived at

Quantitative variables

11 Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why

STROBE Item 12: Statistical Methods

(a) Describe all statistical methods, including those used to control for confounding

(b) Describe any methods used to examine subgroups and interactions

(c) Explain how missing data were addressed

(d) Cohort study—If applicable, explain how loss to follow-up was addressed. Case-control study—If applicable, explain how matching of cases and controls was addressed. Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy

(e) Describe any sensitivity analyses

STROBE Items 13-22

Items 13-17 are about **Results** (Participants, Descriptive Data, Outcome Data, Main Results, Other Analyses)

Items 18-21 are about **Discussion** (Key Results, Limitations, Interpretation, Generalizability)

Item 22 (below) is about Other Information (Funding)

> *Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based*

STROBE Articles

https://www.strobe-statement.org/strobe-publications/

An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting.

Section 5

Finishing Up

Agenda for Zoom Call

Thursday 2025-01-16 from 10 to 11 AM. Zoom details in your email and on Canvas.

- Welcome to 500
- Logistics of the Course
- A Motivating Example: Aspirin and Mortality in Heart Patients
  - How can we avoid being misled?
  - Causal Effects as comparing potential outcomes

Lecture 2 - view before 10 AM Thursday 2025-01-23

- 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?
- What is a propensity score, and how do we ...
  - estimate it,
  - see how well it's working, and
  - use it to estimate causal effects?

The slides I use will be posted to the Class 02 README.