Introduction to Study Design

February 2025

Melissa J. Fazzari
Division of Biostatistics

ICTR Institute for Clinical and Translational Research AT EINSTEIN AND MONTEFIORE
Building Bridges in the Bronx and Beyond

Example

Suppose we want to examine the effect of sunscreen use on the risk of melanoma

- We know that in laboratory studies sunscreen blocks UV radiation (efficacy)

…..But does it reduce melanoma risk in the real world (effectiveness)?

Motivation

To translate this research idea into an actual project we need to understand study design

- A **study design** is a general plan for setting up and testing your hypothesis

- There are many types of study designs, and there is not always a single “best” study design

Scientific Rigor

Whatever study design we end up using, we want it to have strong **scientific rigor**

Scientific rigor = science done well.

The strict application of the scientific method to ensure reproducible, unbiased and well-controlled experimental design, methodology, analysis, interpretation and reporting of results

There is often a need to balance **scientific rigor** with limited time, ethical considerations, cost, and resources

Classification of Research Designs by Degree of Scientific Rigor

Most rigorous: Experimental

Has an intervention

Usually randomized

Usually has a control group

Prospective in nature

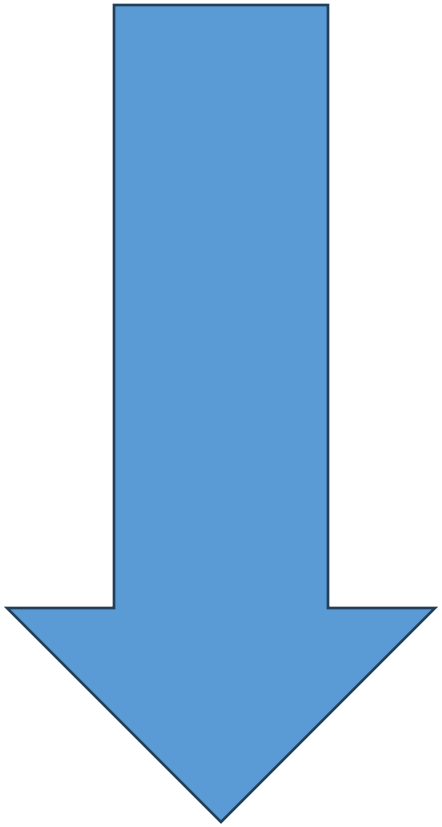
Quasi-experimental

Prospective or a mix of retrospective/prospective

Has an intervention

Non-experimental

Observational, may be prospective or retrospective



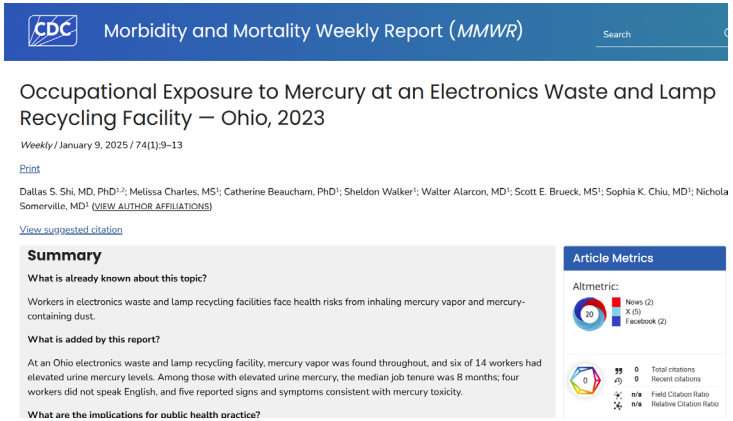
Types of Study Designs

Non-experimental/Observational Studies

What happens in the real world

Design Type 1: Case Reports and Case Series

- This is often done to alert clinicians and researchers about new, emerging issues or to describe interesting or unusual cases, treatment, or recovery.
- Hypothesis generating
- For example, we often see case reports and case series in the CDC's MMWR.



CDC Morbidity and Mortality Weekly Report (*MMWR*) Search

Occupational Exposure to Mercury at an Electronics Waste and Lamp Recycling Facility — Ohio, 2023

Weekly / January 9, 2025 / 74(1);9–13

Print

Dallas S. Shi, MD, PhD[1,2]; Melissa Charles, MS[1]; Catherine Beaucham, PhD[1]; Sheldon Walker[1]; Walter Alarcon, MD[1]; Scott E. Brueck, MS[1]; Sophia K. Chiu, MD[1]; Nichola Somerville, MD[1] (VIEW AUTHOR AFFILIATIONS)

View suggested citation

Summary

What is already known about this topic?

Workers in electronics waste and lamp recycling facilities face health risks from inhaling mercury vapor and mercury-containing dust.

What is added by this report?

At an Ohio electronics waste and lamp recycling facility, mercury vapor was found throughout, and six of 14 workers had elevated urine mercury levels. Among those with elevated urine mercury, the median job tenure was 8 months; four workers did not speak English, and five reported signs and symptoms consistent with mercury toxicity.

What are the implications for public health practice?

Article Metrics

Altmetric:

20 News (2) X (5) Facebook (2)

0 Total citations 0 Recent citations n/a Field Citation Ratio n/a Relative Citation Ratio

Design Type 1: Case Series

Case series example: A dermatology clinic publishes a report on 20 patients diagnosed with melanoma, noting that all 20 regularly used sunscreen.

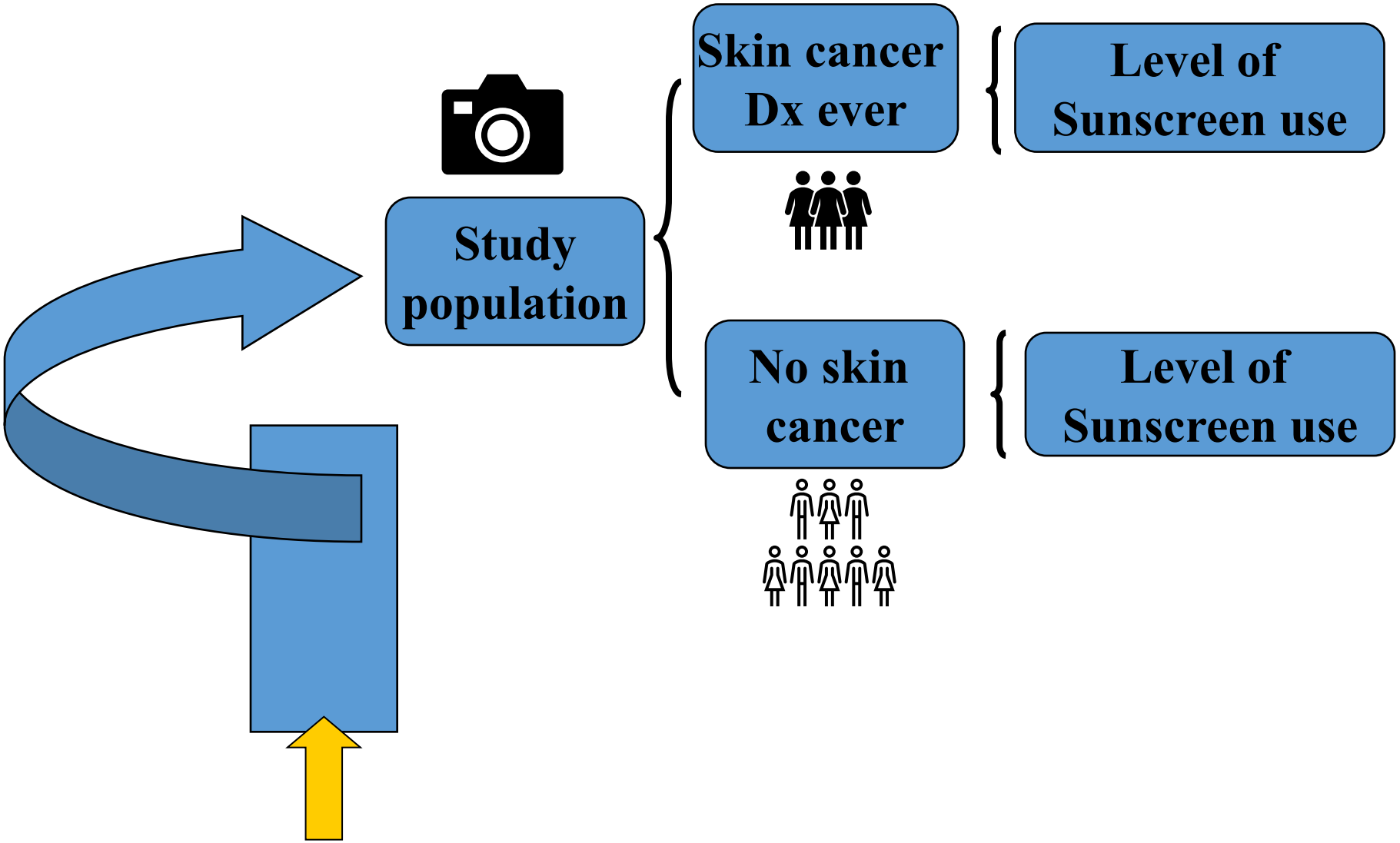
Can we conclude that sunscreen use is associated with an increased risk of melanoma on the basis of this report?

Design type 2: Cross sectional study

- A random sample from **one moment in time**
- Nothing is manipulated by the researcher
- Example: we ask 10,000 people:
  - Do you use sunscreen regularly?
  - Have you ever been diagnosed with melanoma?

We then compare melanoma rates among sunscreen users vs. non-users

Cross-sectional Design



Study only exists at this point in time = a "snapshot"

Cross sectional study

Advantages:

- Study can be conducted quickly, no waiting for outcome to develop
- Relatively inexpensive, no follow-up or loss to follow-up/dropouts
- Prevalence of risk factors and melanoma can be quickly ascertained

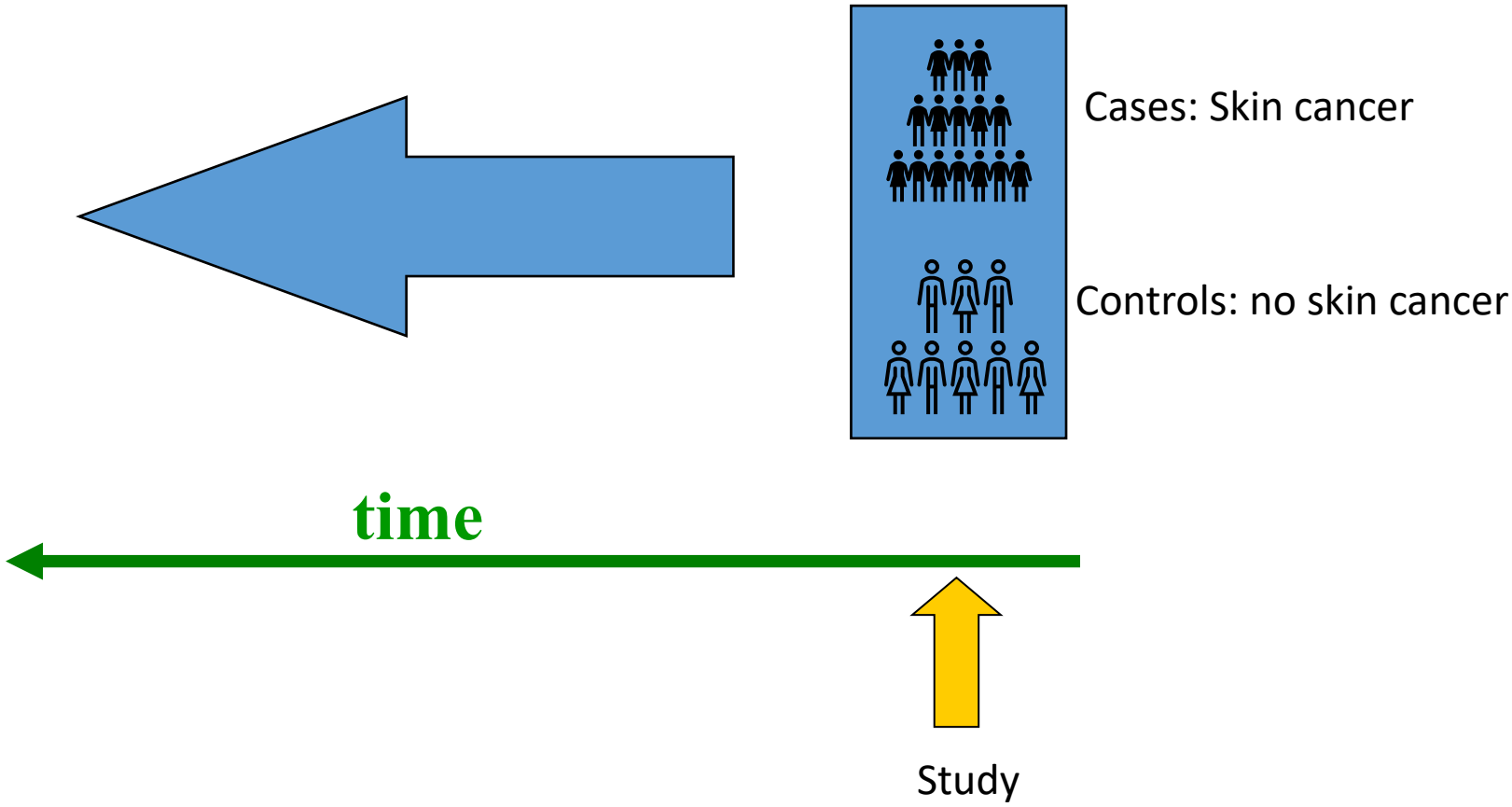
Disadvantages:

- No time element - cannot measure incidence of melanoma
- Which came first? Heavy sunscreen use may follow a skin cancer diagnosis
- Sunscreen use can vary over time, no information about long-term use
- Recall issues
- Melanoma is quite rare (~2% cumulative incidence by age 70 in White people, ~0.5% in Hispanic and Black).

Design type 3: Case-Control Studies

- Because melanoma is quite rare, a better design may be to find as many of these cases as possible in one or many clinics instead of relying on a few cases in a random sample
- Controls are sampled from the same population that gave rise to cases (perhaps people with other skin conditions going to the same clinics)
- We then look back to see the pattern of sunscreen use in both groups (assuming it was reported)

Case-Control Studies



Case-Control Study

- **Strengths**
  - Less expensive and time consuming
  - Efficient for studying rare diseases because we can analyze the data from all patients diagnosed with melanoma and a subset of patients without melanoma

- **Limitations**
  - Sunscreen use is recorded after disease occurrence = Recall bias
  - By design we are fixing the number of cases and controls, so we cannot estimate the risk of melanoma

Design Type 4: Cohort Study

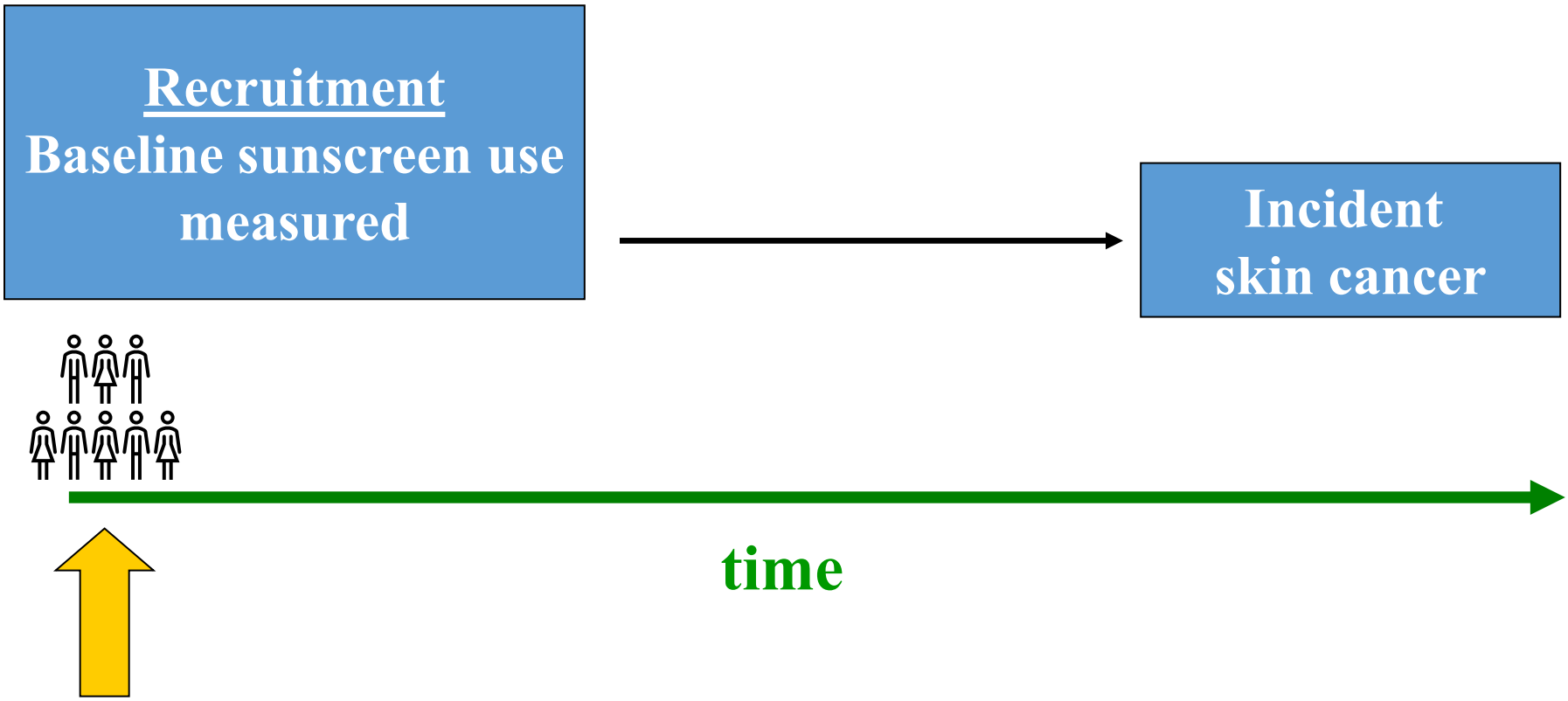
- Given the potential for recall bias, temporal issues and that we want to estimate the incidence of melanoma, a cohort study may be an even better design

- A sample of adults without skin cancer is recruited from the population of interest

- We can also make sure that we are sampling individuals at both ends of the spectrum of sunscreen use

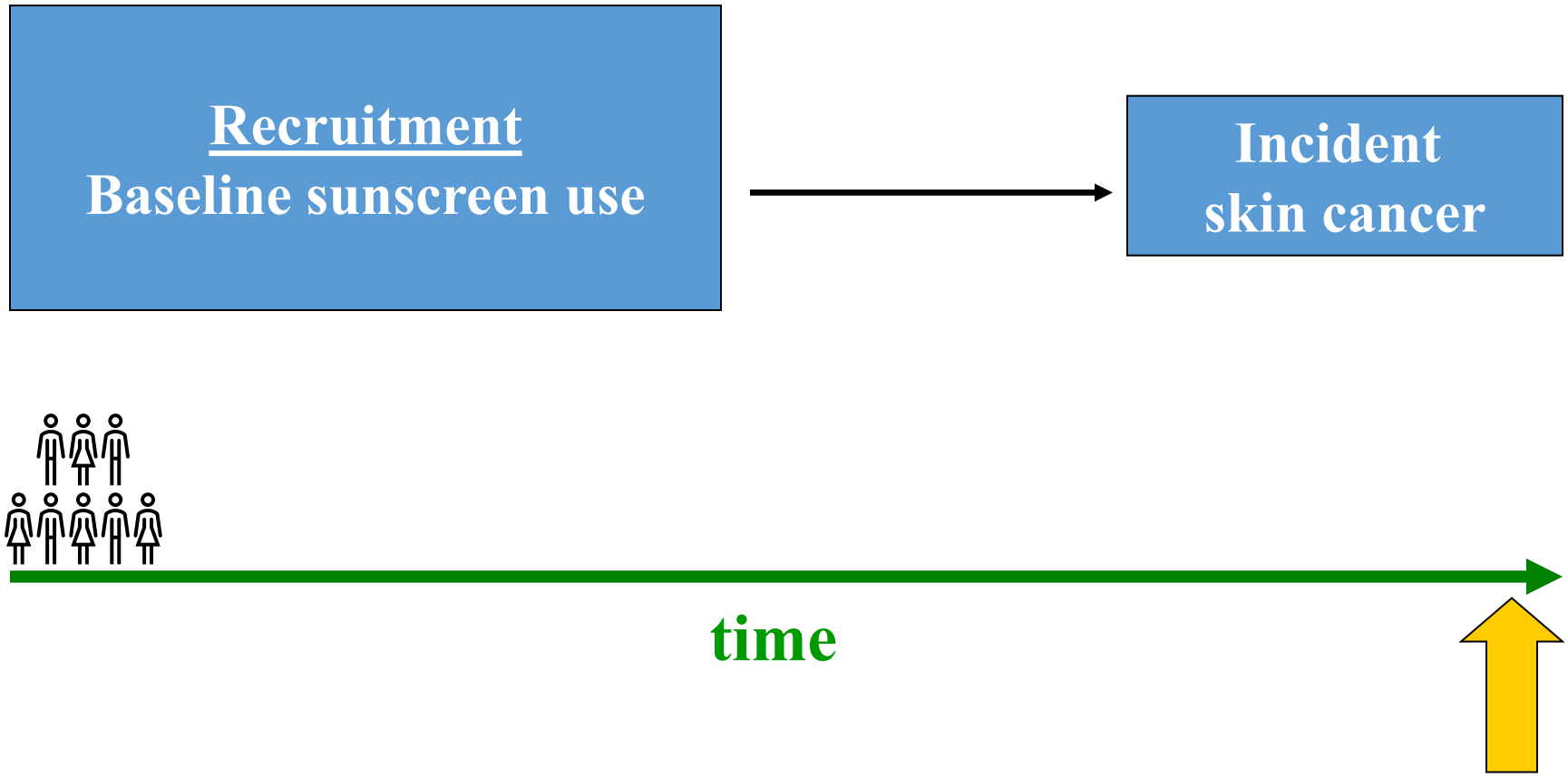
- Cohort is followed up for a long period of time until they develop skin cancer, drop out of the study, or the study ends

Prospective Cohort study



Study begins here – person-time accumulates after study starts

Retrospective Cohort study



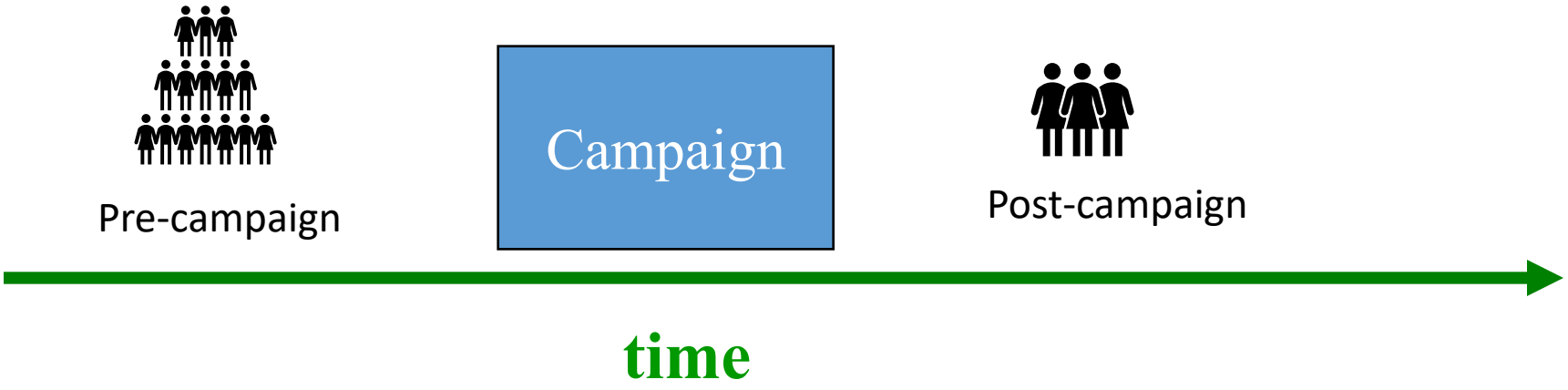
Cohort Study

- **Strengths**
  - Baseline sunscreen use determined before disease detection
  - Subjects selected before disease detection
  - Ethnically safe
  - Can establish timing and directionality of events
  - Can allow sunscreen usage to vary over time in the analysis
  - Can estimate the risk of melanoma

- **Limitations**
  - Expensive and time-consuming
  - Inefficient for rare diseases or diseases with long latency (like melanoma)
  - Loss to follow-up

Design Type 5: Quasi-experimental Studies

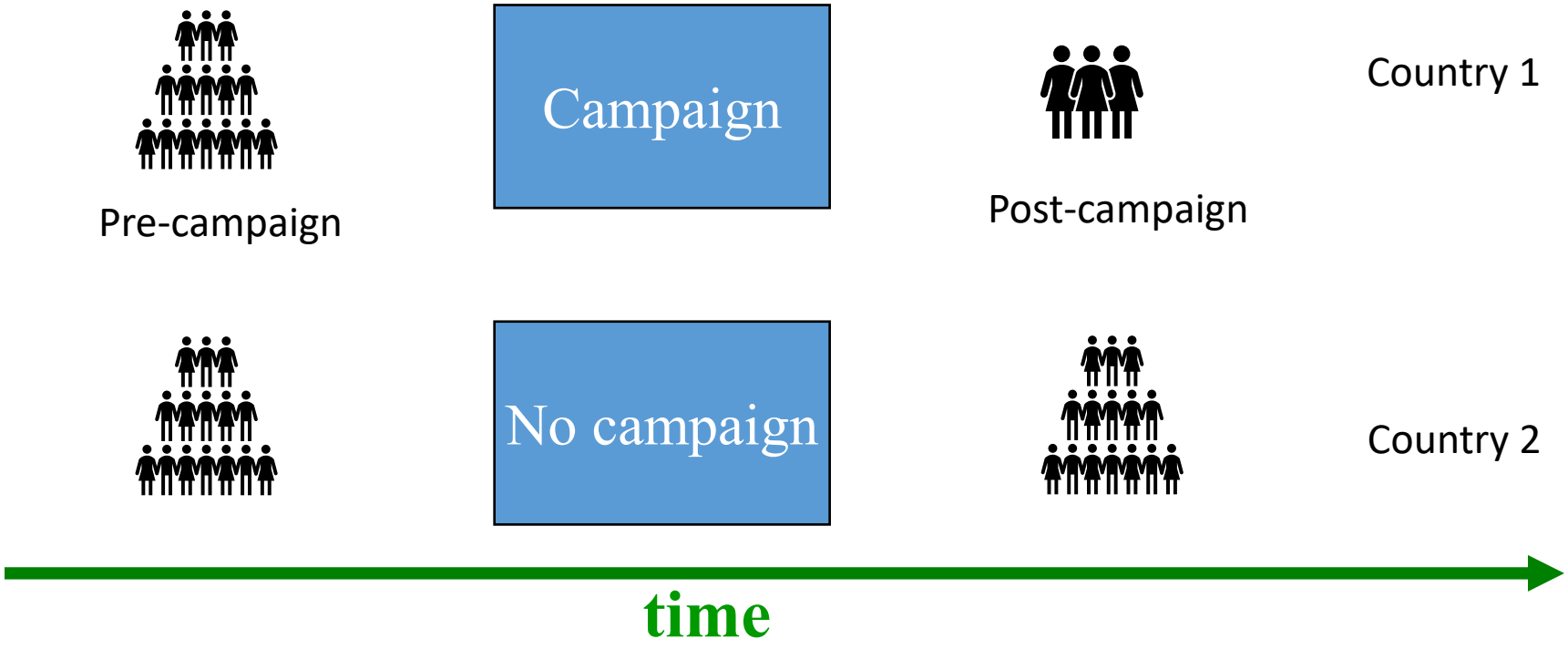
- Suppose we want to assess the effect of a sunscreen use campaign on melanoma rates



We are not randomizing individuals into intervention groups, rather we are observing what happens after some sort of intervention happens, such as a sunscreen campaign.

Quasi-experimental Studies

- We can also have a control group



Confounding

While we may be able to find an association from an observational study, it may be that people who do not wear sunscreen also have other risk factors that are associated with skin cancer (confounders)

For example, people who:

- Are older may not use sunscreen as often as younger people
- Have a strong family history of melanoma or are very fair may use sunscreen all the time given this higher risk
- Do not use sunscreen may have other lifestyle behaviors that increase melanoma risk
- Exercise outdoors a lot may apply sunscreen regularly and exercise may also lower the risk of melanoma through other mechanisms, such as vitamin D exposure

- We cannot determine whether no sunscreen use *causes* skin cancer using an observational study, even a good one, unless **all possible confounding is known and precisely measured and controlled for**

Confounding

A lack of confounding means that the people who wear sunscreen are *interchangeable* with the people who do not wear sunscreen.

They are not more likely to have other sun-protection behaviors, not more healthy, not younger, do not have a different skin type,…..

In observational studies, **investigators do not control assignment to high sunscreen use vs low sunscreen use**. This means that they cannot always effectively control for the differences between the two groups.

Because of this, **confounding is often a very real and important consideration in observational studies**.

Experimental Studies

Design type 6: Randomized Controlled Trials (RCTs)

- Researchers control the exposure. Participants are randomly assigned to "treatment" and "comparison" groups
- Gold standard - provides most convincing evidence of relationship between exposure and effect
- Not possible to use RCTs to test effects of exposures that are known beneficial or expected to be harmful, for ethical reasons (equipoise)

Randomized Controlled Trials (RCTs)

Randomization plays a key role:

> Randomization produces groups that are not systematically different with respect to known and unknown prognostic factors

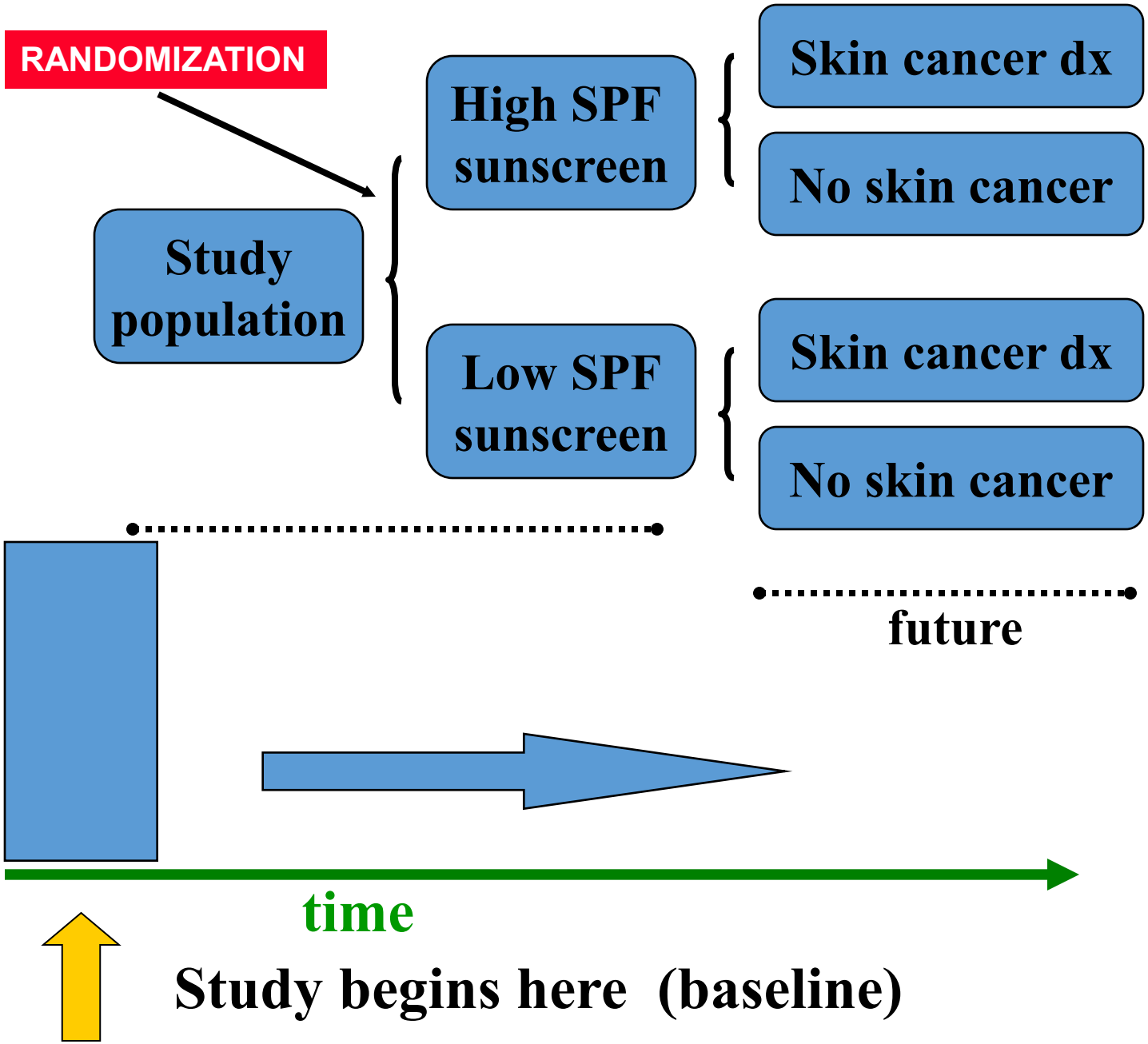
Recall, we said that we wanted the two groups to be *interchangeable.* Randomization allows this to happen.

And it is this randomization that allows us to infer causality

Randomized Controlled Trials (RCTs)

- Suppose we randomize participants to daily sunscreen use vs. no daily sunscreen use (or a low SPF sunscreen) and then observe melanoma rates over a period of time

- If an eventual difference in melanoma rates is observed, we can infer that it was **caused** by differences in sunscreen use, because the two groups were interchangeable at baseline and only differed in their sunscreen use


RANDOMIZATION
Study population
High SPF sunscreen
Low SPF sunscreen
Skin cancer dx
No skin cancer
Skin cancer dx
No skin cancer
future
time
Study begins here (baseline)

Randomized Controlled Trials

Disadvantages

- Very expensive, may take a lot of time to conduct
- May not reflect real-world behavior
- Not appropriate to answer certain types of questions
  - it may be unethical, for example, to assign persons to certain treatment or comparison groups

Design Type 7: Meta Analysis

- A meta-analysis is part of a systematic review
- We examine the results of many different studies and combine them by calculating an *average* or *common* effect
- Improves the precision of an estimate by using all valid available studies and data

When can we do a meta-analysis?

- When there are no differences in the study characteristics that are likely to substantially affect outcome
- When the outcome has been measured in similar ways
- When the data are reported and easily extracted
- When more than one study has estimated an effect

Example

A 2020 meta-analysis reviewed 20 studies and calculated an overall risk estimate

Some studies showed a protective effect of sunscreen, others found no effect, and a few observed a higher melanoma risk with sunscreen

The “average” effect of sunscreen by combining the evidence from these 20 studies showed that sunscreen had a subtle protective effect.

Issues:

- Different studies may have used different definitions of sunscreen use
- We want all included studies to be high quality
- Different populations studied may produce different estimates of risk. Should we even be averaging in the first place?

Some statistical considerations

Measuring association

Typical measures:

- Risk ratios
- Odds ratios
- Absolute differences

1. Risk

- 240 people in a study, and 6 cases of melanoma reported

- risk

  = 6 cases/240 who could have had melanoma

  = 6/240 = = 0.025 = 2.5%

$$\text{risk} = \frac{\text{number of events of interest}}{\text{total number of observations}}$$

2. Odds

- 240 people in the sample, and 6 with melanoma

- odds of melanoma
  = 6 cases of melanoma/234 with no melanoma
  = 6/234 = 0.0256

$$\text{odds} = \frac{\text{number of events of interest}}{\text{number without the event}}$$

Expressing it in words

- Risk
  - the probability of melanoma in the sample was 0.025
  - One person had melanoma for every forty
- Odds
  - the odds of having melanoma were 0.0256
  - one person had melanoma for every 39 people who didn’t

Risk and Odds ratios

Now suppose we examine another group of 240 people who do not use sunscreen and 10 have had melanoma.

- Risk of melanoma is 10/240 = 0.04167

- Odds of melanoma are 10/230=0.0435

Risk and Odds ratios

- Risk ratio 0.04167/0.025 = 1.66
  - the risk of having melanoma in the high-risk group was 1.66 times the risk compared to the original group

- Odds ratio = 0.0435/0.0256 = 1.70
  - The odds of melanoma in the high-risk group was 1.70 times higher compared to the original group

- **When OR or RR = 1, this implies no difference in the effect of sunscreen**

3. (Absolute) Risk difference

- Risk of melanoma in no sunscreen group minus risk in sunscreen group

=0.04167 – 0.025 = 0.0167

- The absolute risk for melanoma is increased by 1.7% in the no sunscreen group

When risk difference = 0, this implies no difference in the risk of melanoma based on sunscreen use

Relative vs absolute measures of association

- RRs and ORs tend to look much more dramatic compared to absolute differences
  - ~ 70% increase in the risk of melanoma
  - ~ 1.7% increase in absolute risk

Confounding

- We observed that no sunscreen use is associated with higher risk of melanoma

- **Our estimate of association (OR=1.70) may be misleading.**

- Since age is associated with overall cancer risk and lower use of sunscreen, we could have overestimated the effect of sunscreen because we did not separate the strong effect of age from the effect of sunscreen on melanoma risk.

How can we minimize confounding by age?

- **Randomization** – aims for an equal age distribution in both treatment arms

- **Restriction/stratification** – a variable cannot produce confounding if it is prohibited from varying. For example, restrict the analysis to people who are > 65 years old. For stratification, we are **estimating the effect of sunscreen for each age level**.

- **Adjustment** – using statistical models, adjust for the effects of confounders. By adjusting for age, **we are holding it fixed at each level** while looking at the effect of sunscreen

How can we minimize confounding by age?

- **Match -**
  - In a case-control study, match each case (skin cancer) with a control (no skin cancer) with respect to age.
  - In a cohort study, match people with high baseline sunscreen use to people with low sunscreen use by age

- **Propensity scores** – estimate the propensity for low sunscreen use using all available predictors, and then adjust, weight, or match for propensity score

Power Considerations

Part of a good study design is having an appropriately powered study

Power = *the probability of finding a statistically significant effect given a true effect exists*

If the true OR=2.0 (sunscreen use lowers risk), and I conduct this exact study a thousand times, what proportion of these 1000 studies will have a statistically significant result?

Of course, we only do the study once.

Power considerations

Most statistical testing basically boils down to a ratio:

$$\frac{Signal}{noise}$$

If the signal is strong relative to the noise, we may declare it “statistically significant”

In planning our study, we focus on having adequate power. This means that if there is truly a signal, we should be able to detect it with high probability.

Power considerations

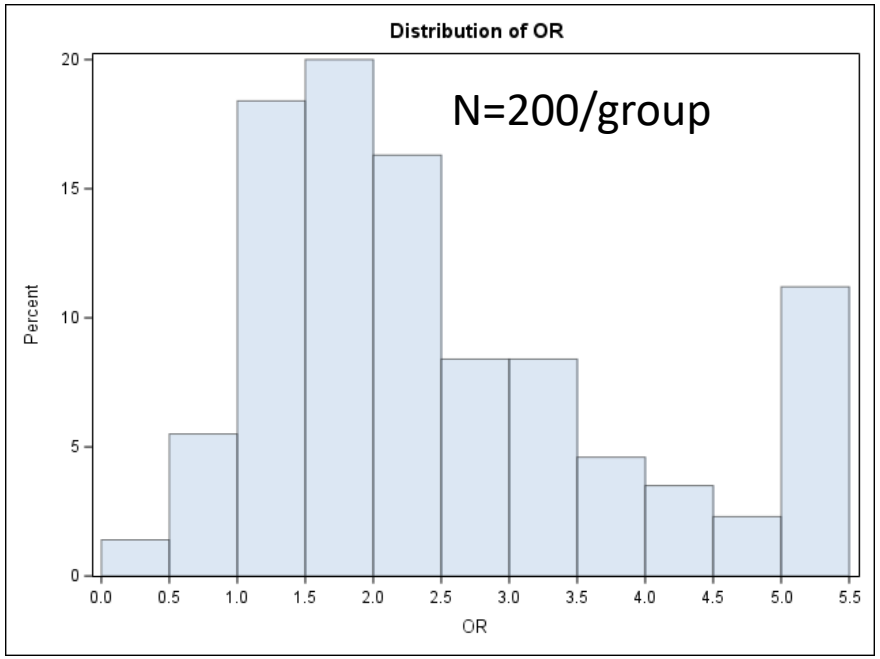
Power is higher when the true difference between groups (signal) is large compared to the noise

Increasing sample sizes reduces the noise/increases the precision of our effect estimate, making it easier to detect the signal

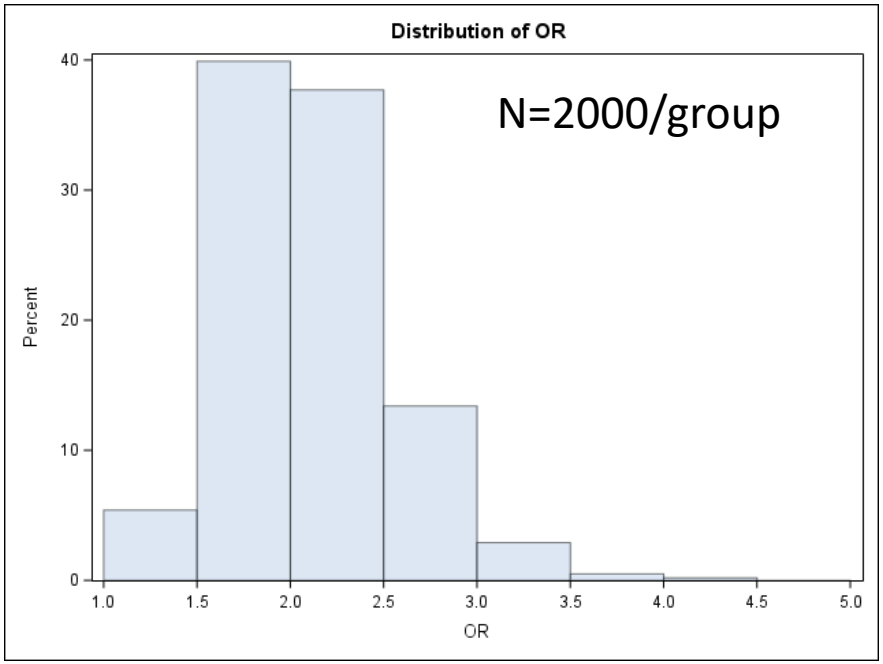
Suppose we assume that the risk of melanoma is 2% in people who use sunscreen, and 4% in those who do not.

If we only compare melanoma proportions in 200 people who use sunscreen and 200 people who do not (exp: 12 melanoma cases), our observed odds ratio will fluctuate a lot more across studies than if we looked at 2000 sunscreen users and 2000 non-users (exp: 120 melanoma cases).

Power considerations



87% of time,
not statistically significant



95% of time statistically significant

Summary

- There is no single perfect study design.
- The biggest issue for non-experimental studies is how to control for confounders and infer causality
- The biggest issues for RCTs are typically time and expense
- The measure of association used may depend on the type of study as well as what is clinically meaningful.
- In all studies, we need to carefully evaluate statistical power to avoid being underpowered

Thank you