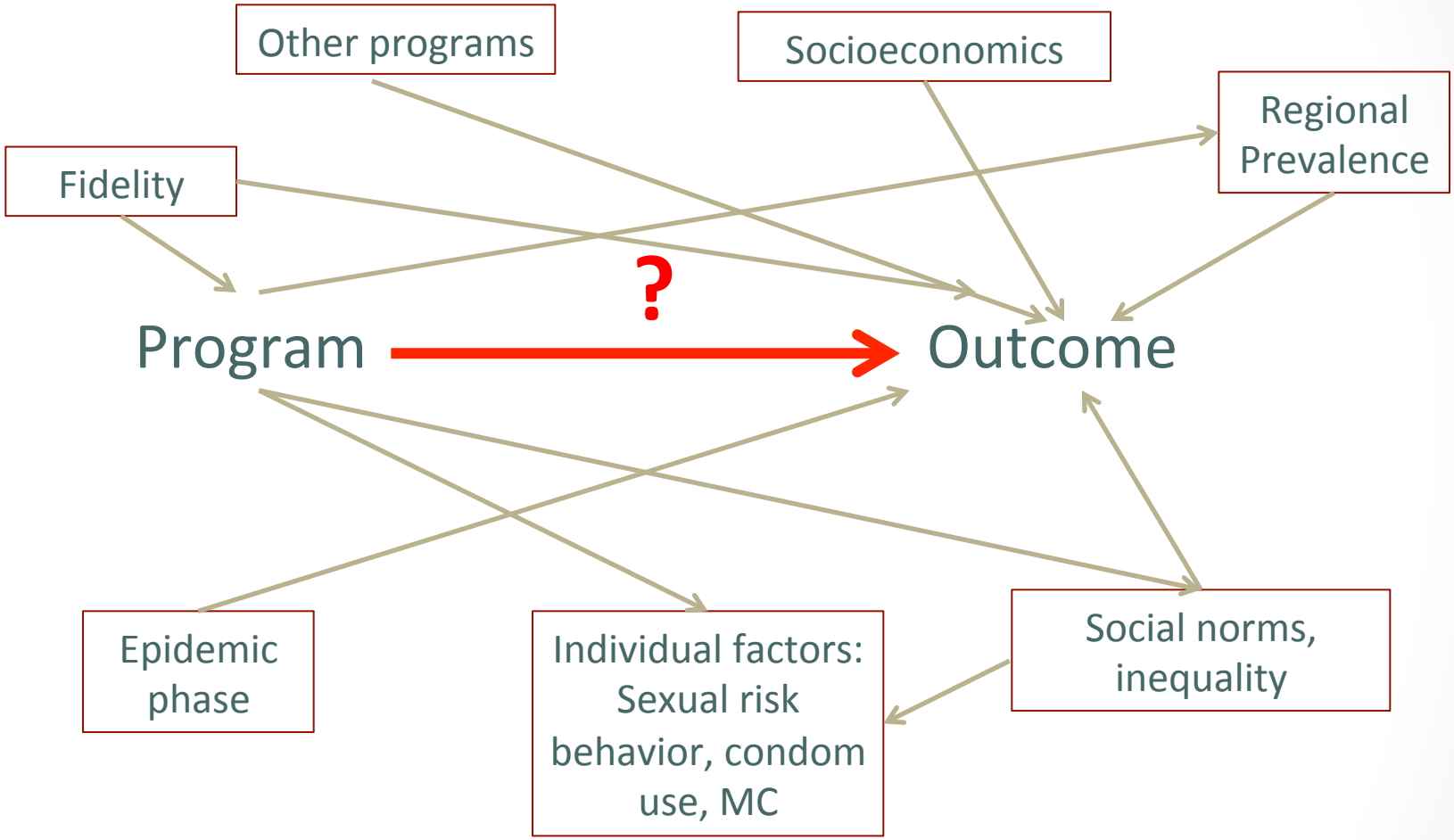
Causal Inference

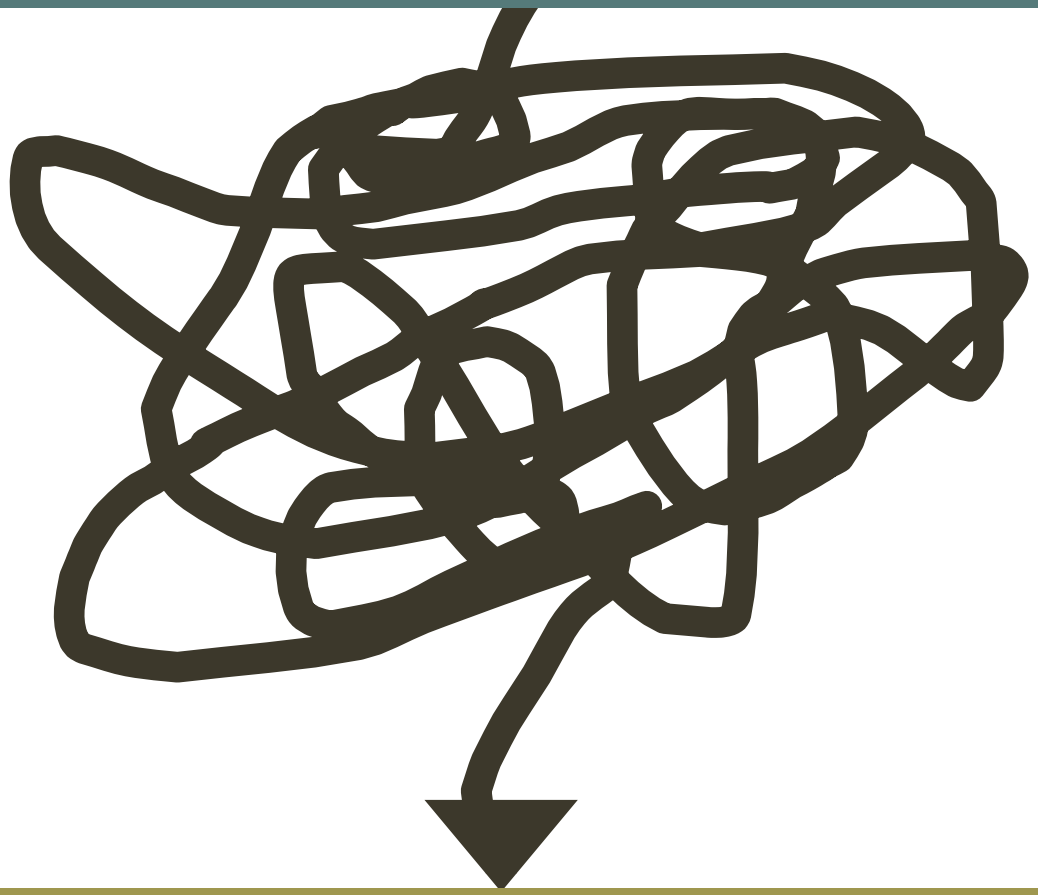
Sandi McCoy, MPH, PhD
University of California, Berkeley
July 16, 2011

Enhancing Implementation Science: Program Planning, Scale-up, and Evaluation

Program $\xrightarrow{?}$ Outcome


Other programs
Socioeconomics
Regional Prevalence
Fidelity
?
Program
Outcome
Epidemic phase
Individual factors: Sexual risk behavior, condom use, MC
Social norms, inequality

Social, economic, cultural, political and biological milieu



HIV Incidence

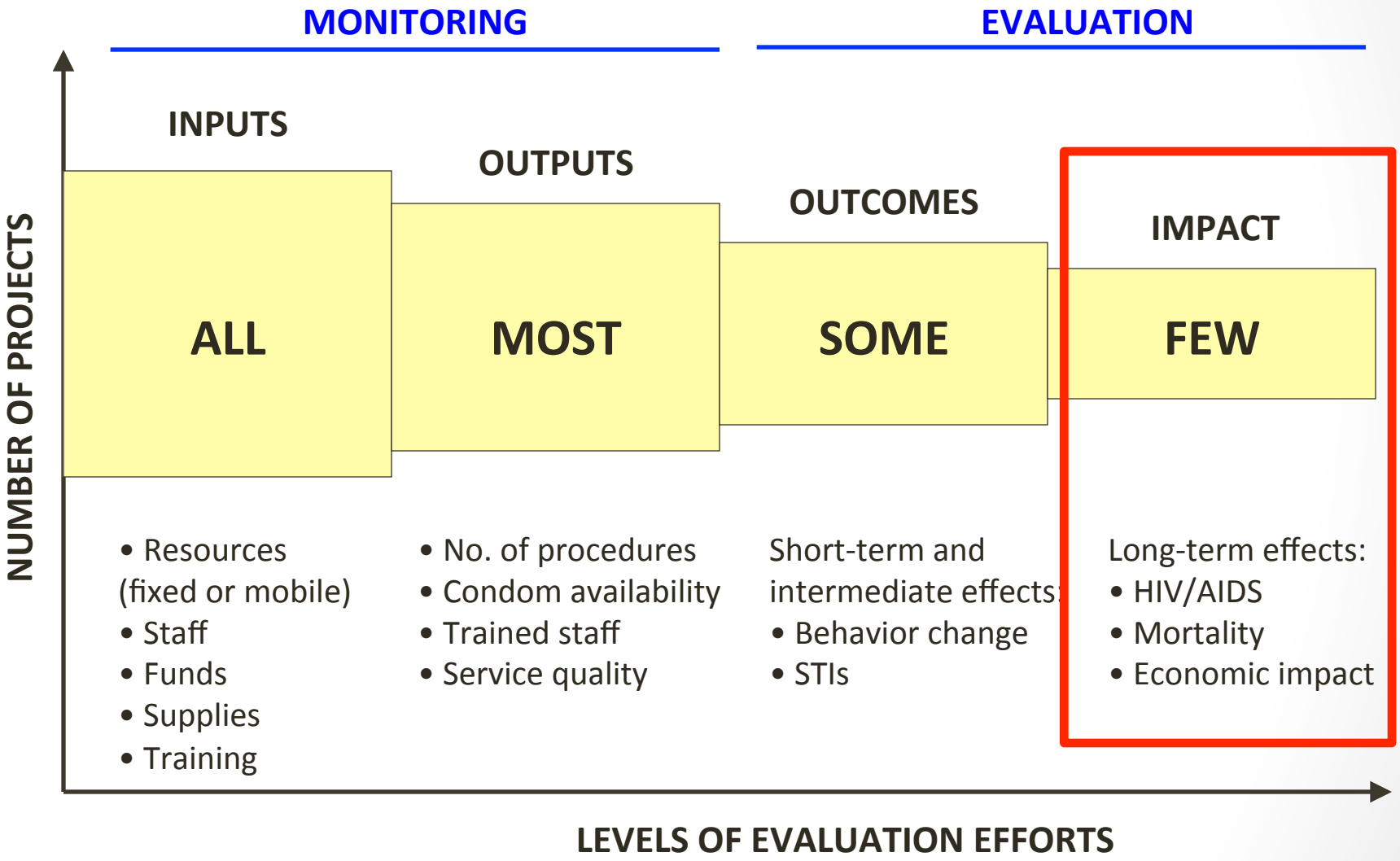
Source: Stefano Bertozzi, BMGF

Learning Objectives

Understand the concept of **counterfactuals** and how **selection bias** affects impact evaluations

1. Impact evaluation and counterfactuals
2. Importance of a valid counterfactual
3. Selection bias
4. Study designs and threats to validity
5. Quality Assessment

Monitoring and Evaluation Results Pyramid



Source: UNAIDS, The World Bank. Monitoring and Evaluation Operations Manual. 2002

Impact Evaluation Answers:

- What was the effect of the program on outcomes? (*causal attribution*)
- How much better off are the beneficiaries as a result of the program?
- What happened compared to what would have happened without the program?
- Is the program cost-effective?

Example: What is the Impact of…

giving Robert

additional pocket money

(P)

on Robert's consumption of candies?

(Y)?

The Perfect *Clone*

Robert

6 candies

Robert's Clone

4 candies

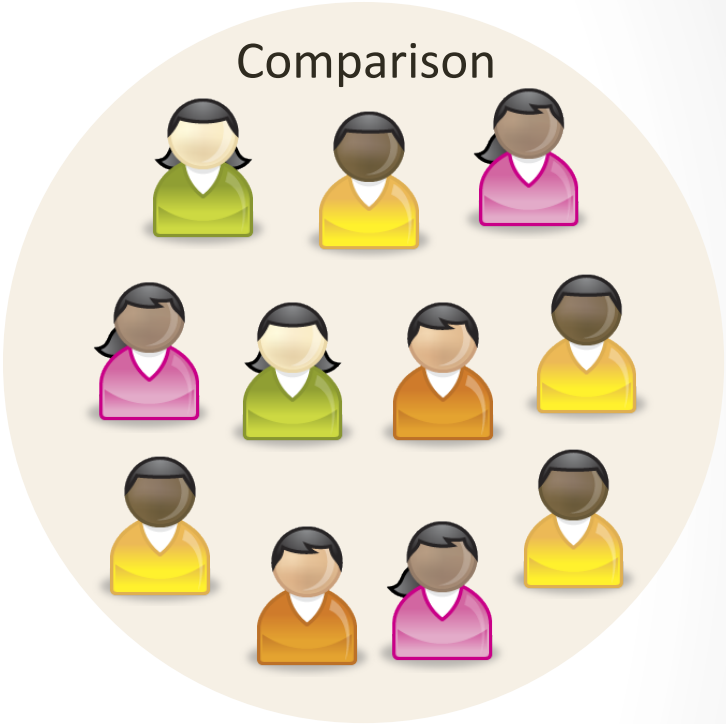
IMPACT: 6-4=2 Candies

Solving the Evaluation Problem

- Problem: we never observe the same individual with and without program at same point in time

- Need to **estimate** what would have happened to the beneficiary if he or she had not received benefits; i.e. *causal attribution*

In reality, use statistics



Average Y=6 candies

Average Y=4 Candies

IMPACT=6-4=2 Candies

Our Objective:

Estimate the *CAUSAL* effect (*impact*) of:

intervention P *(male circumcision)*

on

outcome Y *(HIV incidence)*

Solution:

- Estimate what ***would*** have happened to **Y** in the absence of **P**
- We call this the……

<u>**COUNTERFACTUAL**</u>

Solution:

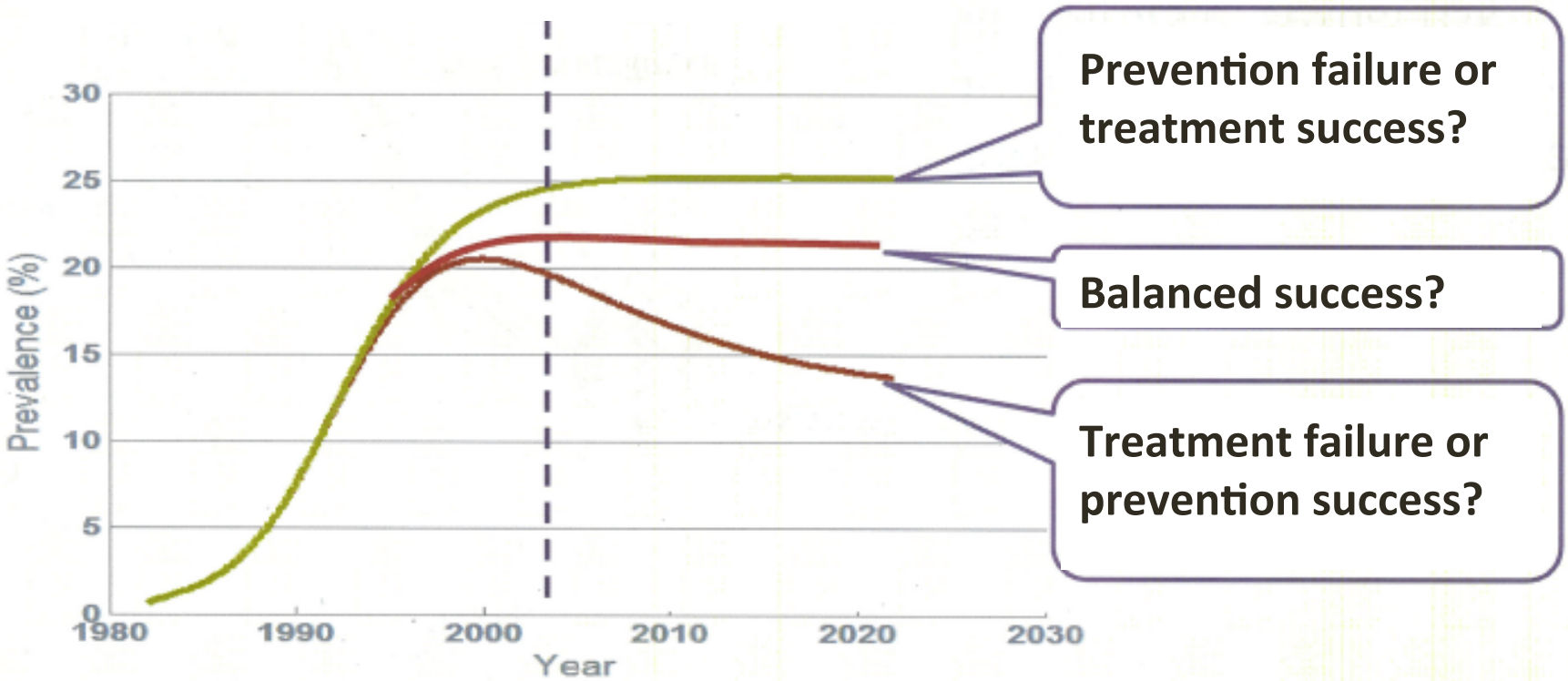
- Estimate what ***would*** have happened to **Y** in the absence of **P**
- We call this the…..

COUNTERF

The key to a good impact evaluation is a valid **counterfactual!**

Why Do We Need a Counterfactual?

At this stage of the epidemic…?



Good Counterfactuals

- Since we can never actually know what *would have happened*, **comparison groups** allow us to **estimate** the counterfactual
- A good counterfactual can help you estimate the "true" (hypothetical) causal effect

Hint: With a good counterfactual, the **only reason** for different outcomes between treatments and controls is the **intervention (P)**

Finding a Good Counterfactual

The treated group and the counterfactual group:

- Have identical factors/characteristics, <u>except for benefiting from the intervention</u>
- The only reason for the difference in outcomes is due to the intervention

Population 1 + Intervention

Vs.

Population 2

Poor Counterfactuals Can Threaten Validity

1. Selection Bias

- Volunteer participants or those not receiving program are different than those without

2. Confounding

3. Endogenous Change

- Secular changes or drift (long term trends in community, region or country)
- Maturational trends (Individual change)
- Interfering events

- **Measurement error**
- **Hawthorne/cohort effects**

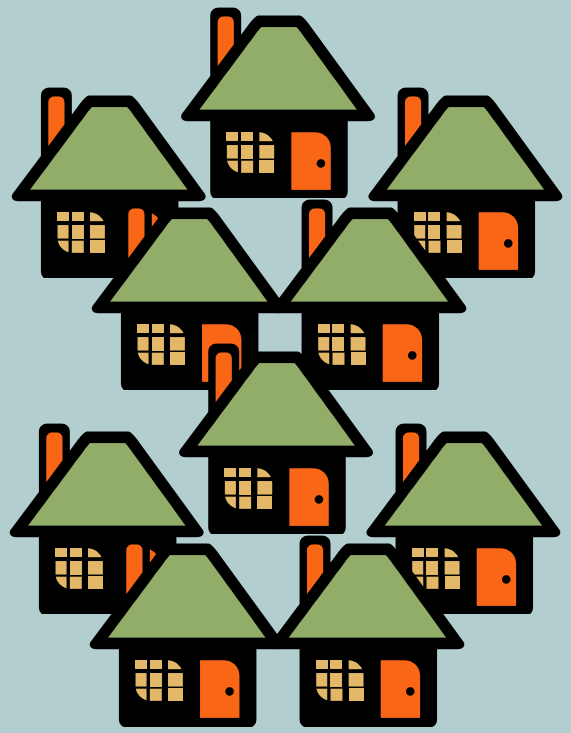
} All studies!

Study Designs
& Threats to Validity

1. Enrolled vs. Not Enrolled (Selection Bias)

Designs Leading to Biased Results: "**Enrolled versus Not Enrolled**"

Consider a school-based pregnancy prevention program



10 schools in the district are asked if they would like to participate

Designs Leading to Biased Results: "Enrolled versus Not Enrolled"

No intervention		**5 schools decline participation**
 *Pregnancy Prevention Program*		**5 schools elect to participate in the program**

Designs Leading to Biased Results: "Enrolled versus Not Enrolled"

No intervention		*Pregnancy rate = 3 per 100 student years*
 *Pregnancy Prevention Program*		*2 per 100 student years*

Designs Leading to Biased Results: "Enrolled versus Not Enrolled"

No intervention		*nancy rate er 100 ent years*
 *Pregnancy Prevention Program*		*r 100 student years*

Schools in the program had fewer adolescent pregnancies…

Can we attribute this difference to the program?

Designs Leading to Biased Results: "Enrolled versus Not Enrolled"

No intervention		*Pregnancy rate = 3 per 100 student years*	*Factor X (more conservative)*
 *Pregnancy Prevention Program*		*2 per 100 student years*	*Less conservative*

Designs Leading to Biased Results: "Enrolled versus Not Enrolled"

No intervention			*Factor X more conservative*
Pregnancy Prevention Program			*less conservative*

Observed effect might be due to differences in "Factor X" which led to differential self- selection into the program ("selection bias")

Designs Leading to Biased Results: "**Enrolled versus Not Enrolled**"

- This design compares "apples to oranges"
- The reason for not enrolling in the program might be correlated with the outcome
  - You can statistically "control" for observed factors
  - But you <u>cannot control for factors that are "unobserved"</u>

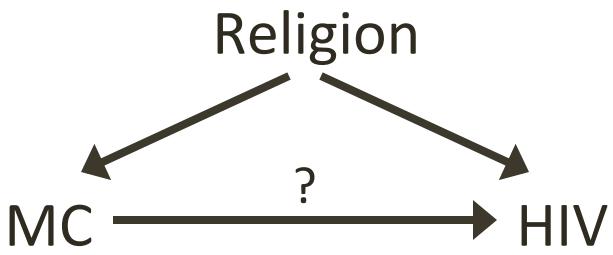
- Estimated impact erroneously mixes the effect of different factors

Study Designs & Threats to Validity

2. Confounding (Correlation or Causation?)

Confounding Example: Male Circumcision (pre-2005)

- Frequently observed that male circumcision status associated with lower HIV prevalence (since 1986)
  - Difficult to disentangle from religion



  - Was this association **correlation** or **causation**?

RCTs confirmed the observational data

The RCTs addressed this problem by randomizing men to get circumcised immediately or after a waiting period, eliminating selection bias & confounding

Study	Result
Auvert, 2005, South Africa	IRR_A = 0.39 (0.23, 0.66)
Bailey, 2007 , Kenya	RR = 0.47 (0.28, 0.78)
Gray, 2007, Uganda	IRR_A = 0.49 (0.29, 0.81)
Combined Effect	**↓58% (95% CI: 43–69%)**

Study Designs
& Threats to Validity

3. Before-After or Pre-Post (endogenous changes)

Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs

- Data on the same individuals before and after an intervention

Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs

Time

Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs

Measure HIV/AIDS knowledge at baseline

Knowledge score = 5/10

Time

Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs

HIV Educational Intervention

Time

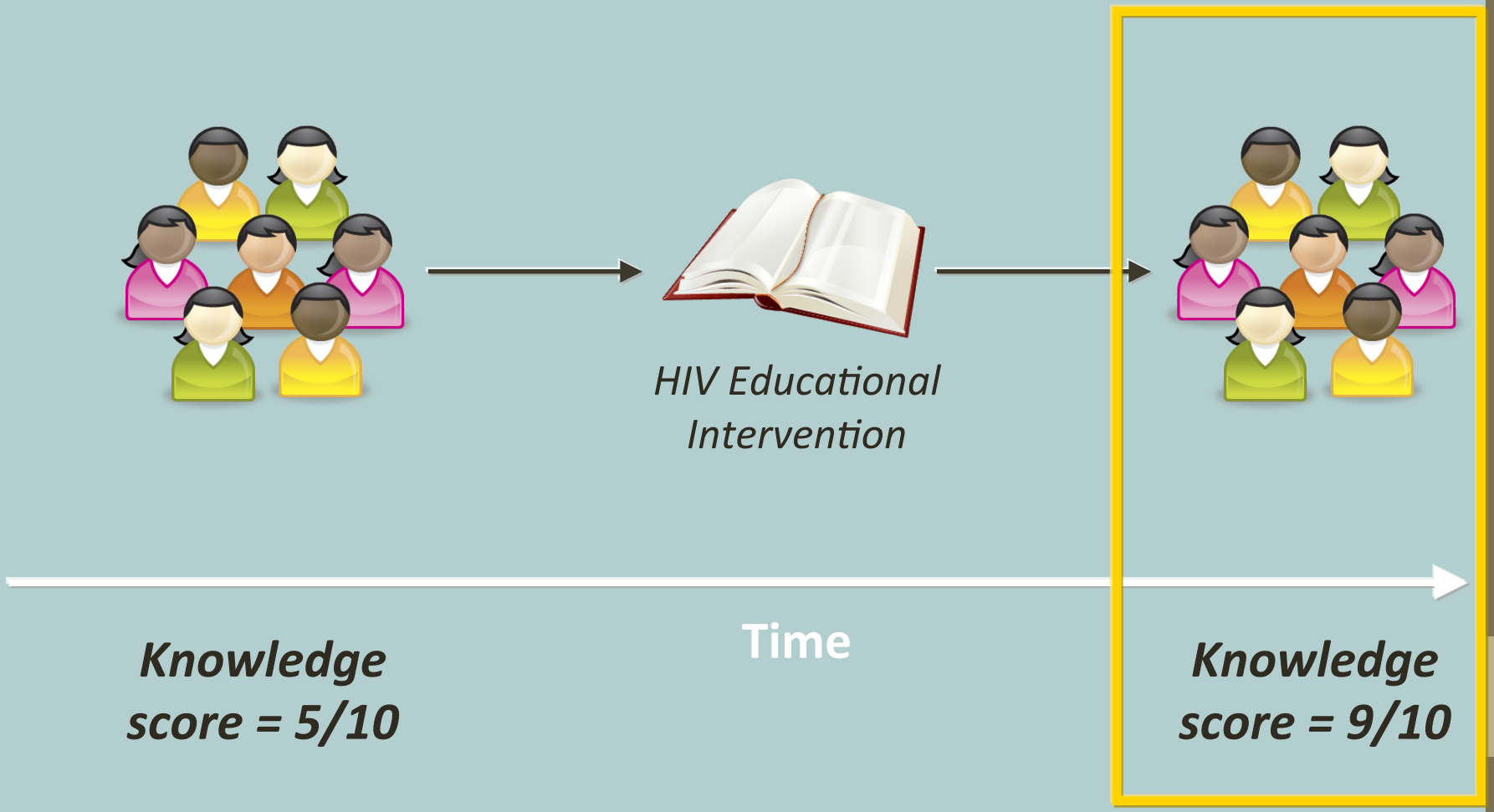
Knowledge score = 5/10

Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs

Time

Knowledge score = 5/10

Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs



Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs

4 point improvement in score after the intervention…

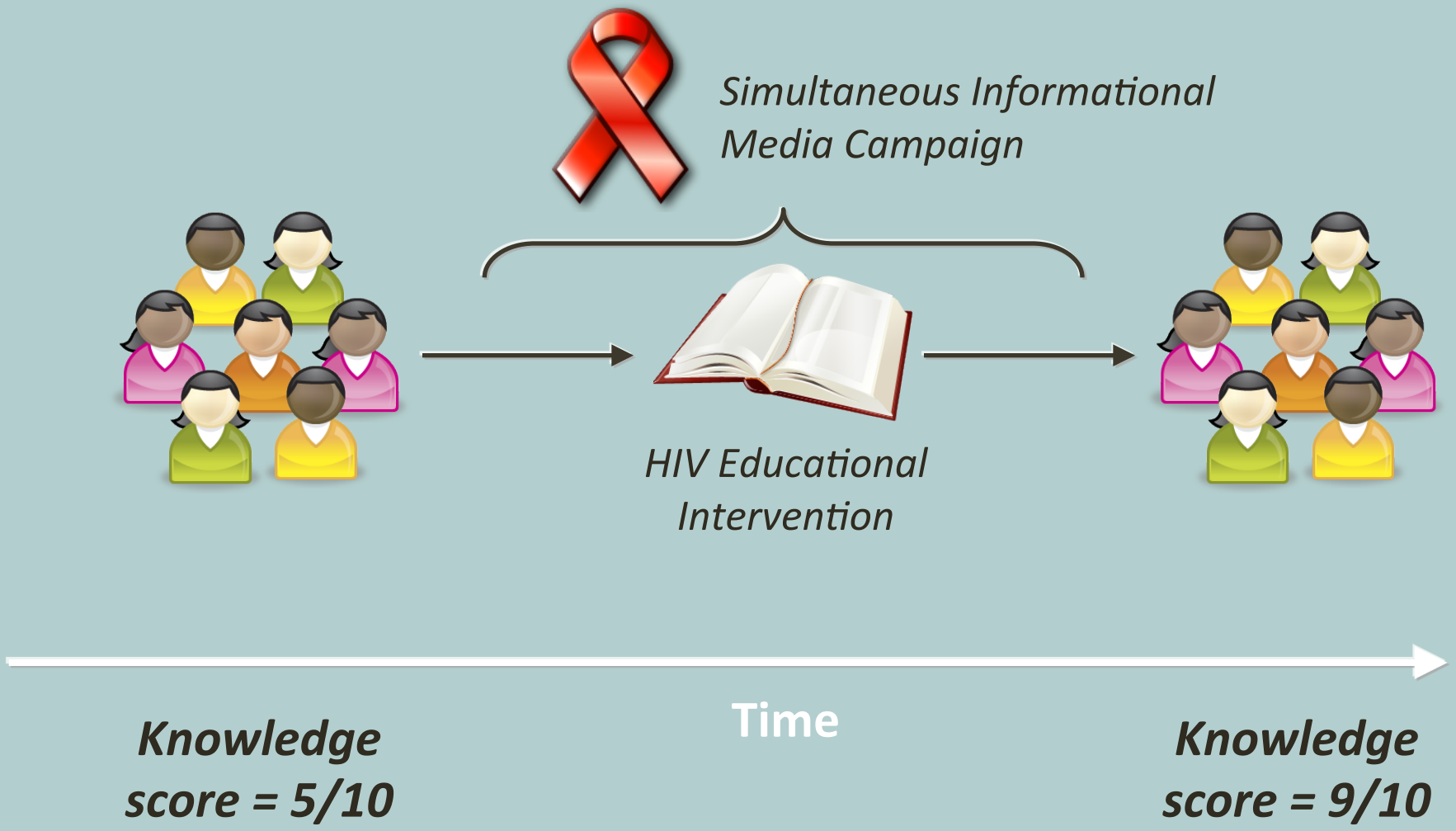
Can we attribute this improvement to the program?

Knowledge score = 5/10

Time

Knowledge score = 9/10

Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs



Designs Leading to Biased Results: "**Before-After**" or "**Pre-Post**" Designs

Problem: Can't account for <u>time varying factors</u> that impact the outcome, such as:

- **changes in the economy**
- **natural changes in the epidemic**
- **simultaneous interventions**

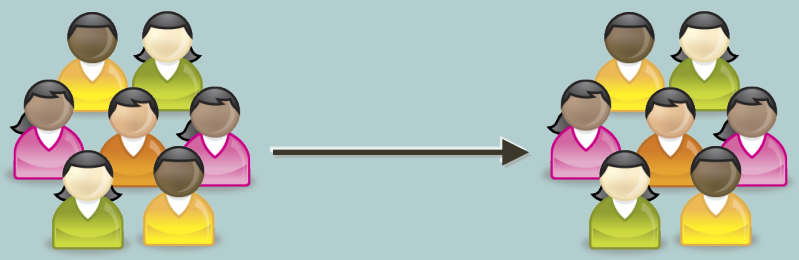
Knowledge score = 5/10

Knowledge score = 9/10

Study Designs & Threats to Validity

4. Historical Controls

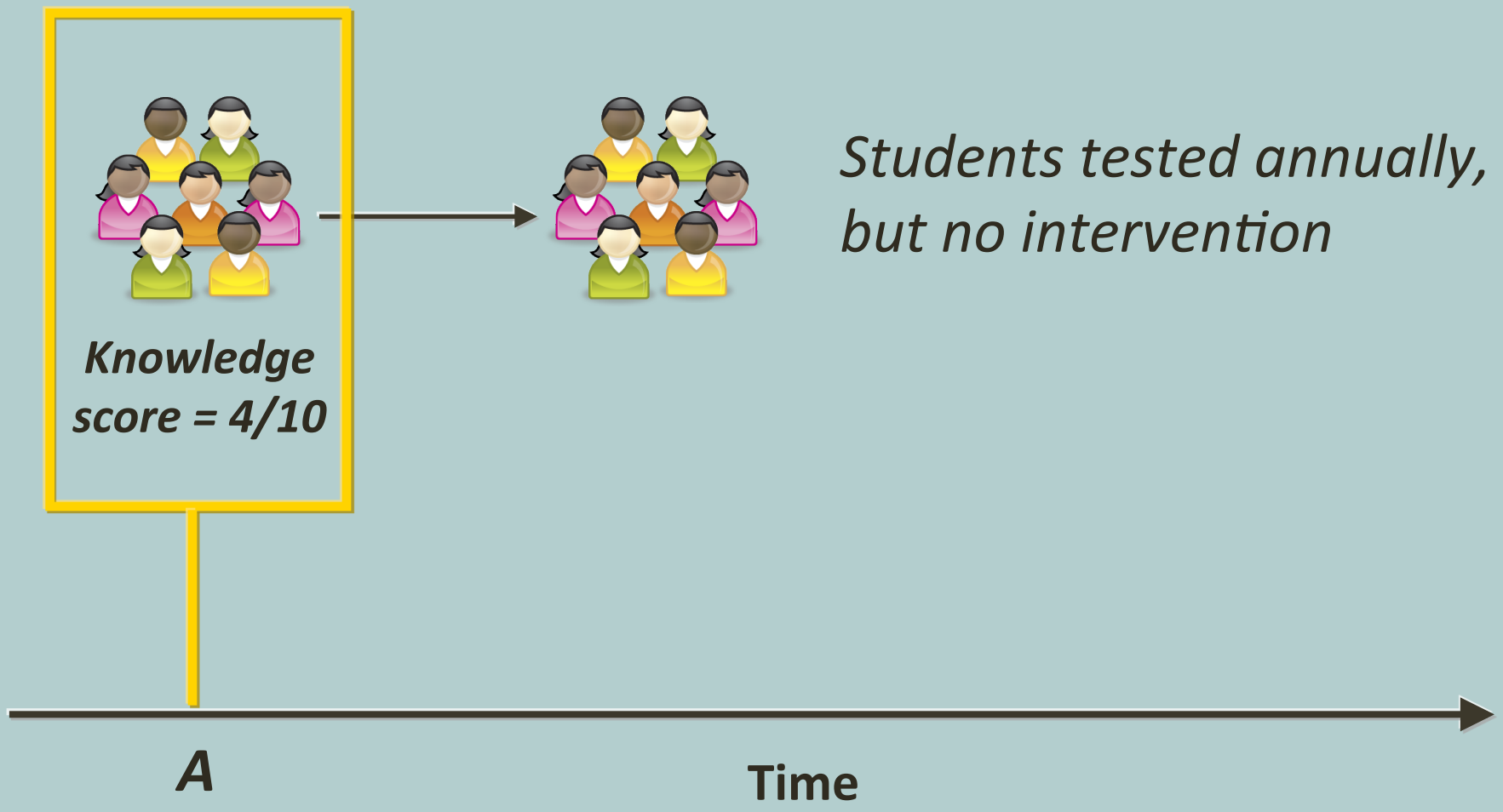
Designs Leading to Biased Results: "**Historical Controls**"



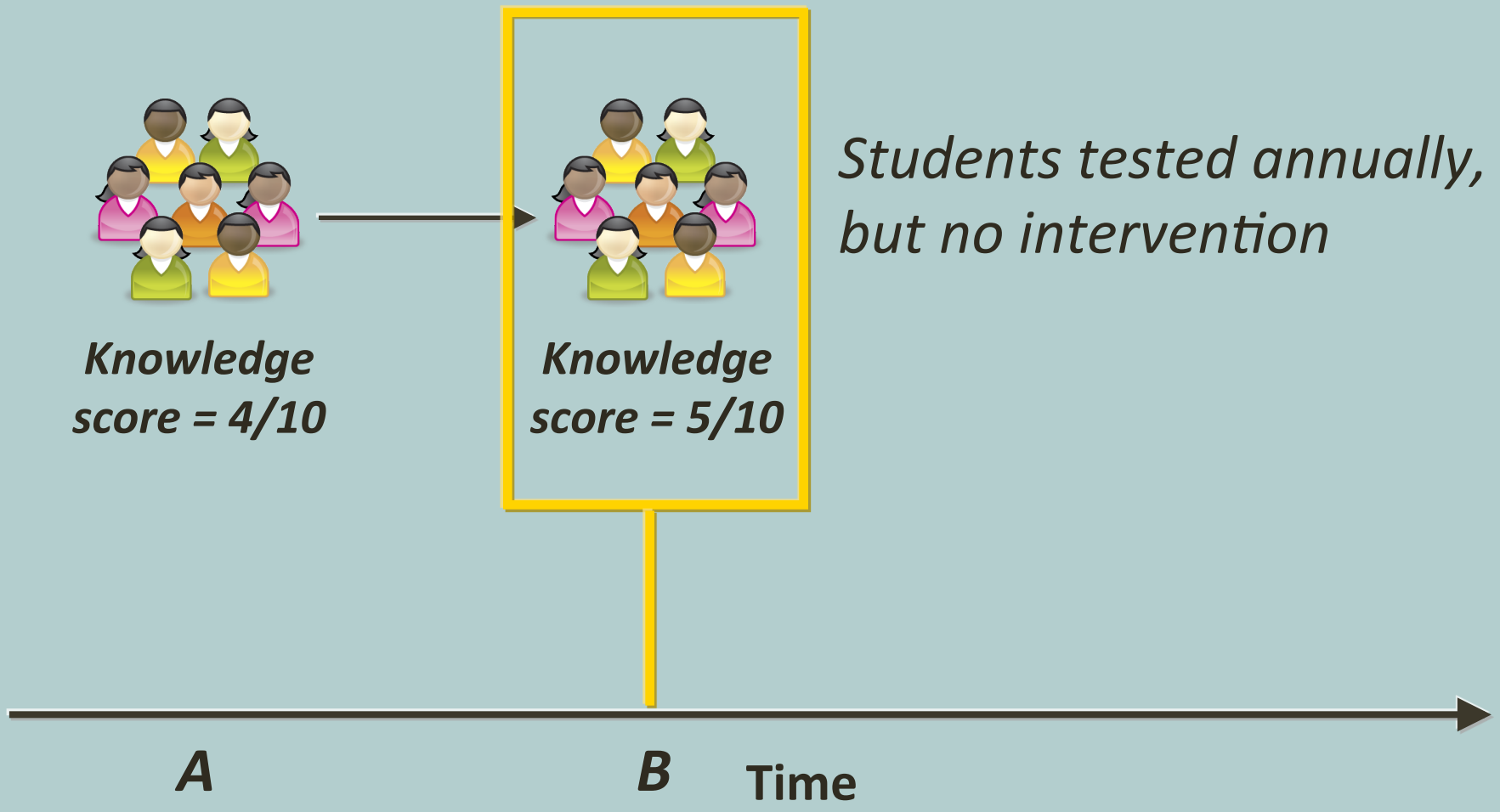
Students tested annually, but no intervention

Time

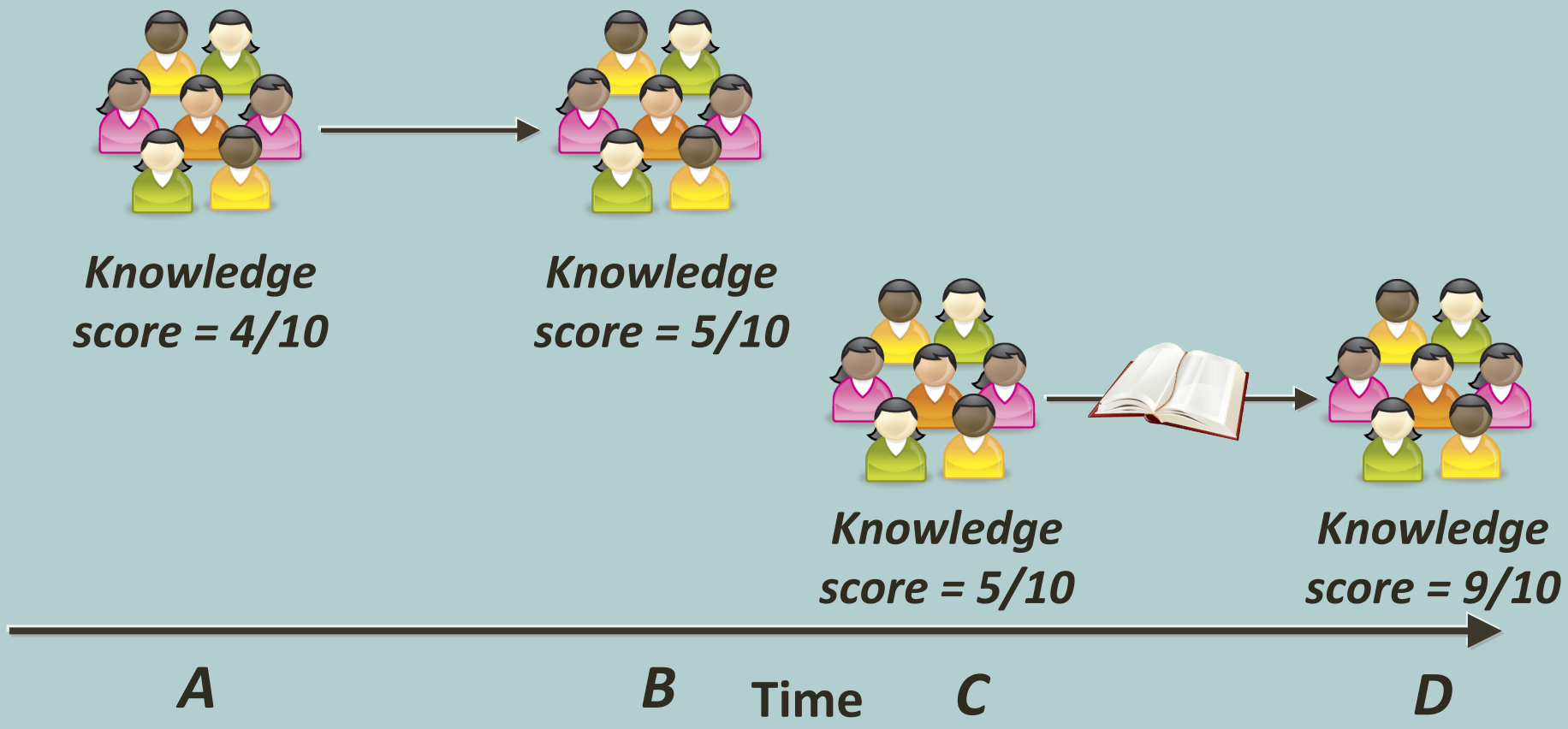
Designs Leading to Biased Results: "Historical Controls"



Designs Leading to Biased Results: "Historical Controls"



Designs Leading to Biased Results: "Historical Controls"



Designs Leading to Biased Results: "Historical Controls"

Knowledge score = 4/10

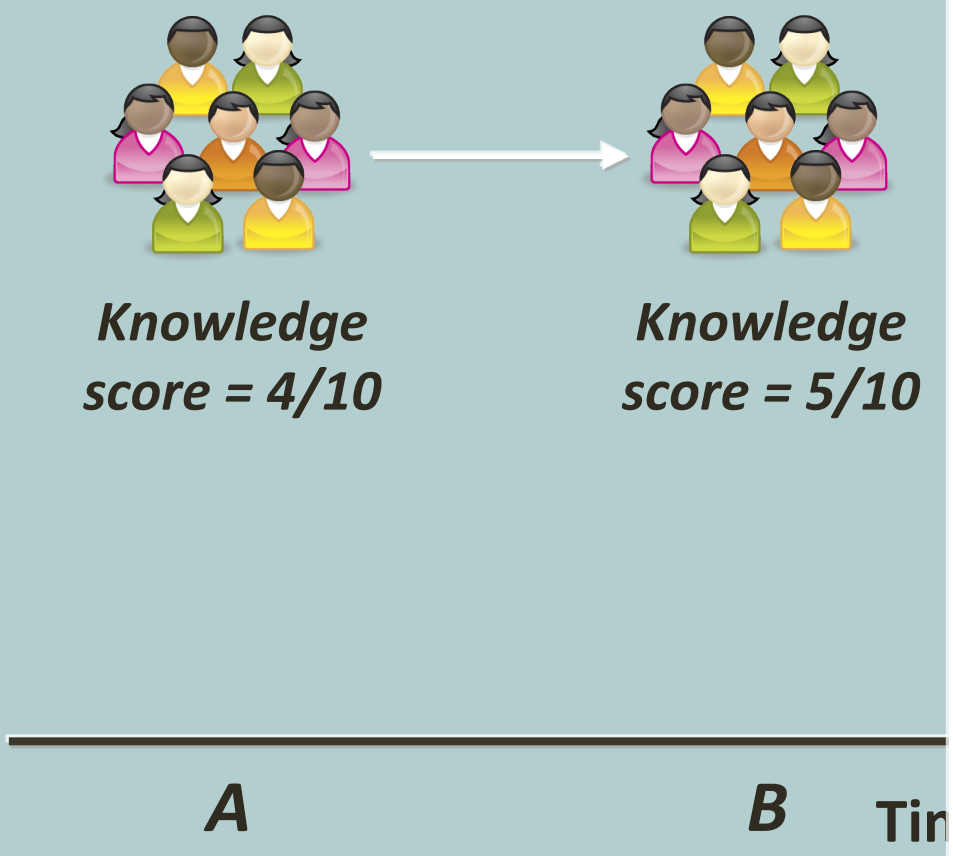
The controls improved by 1 point whereas the intervention group improved by 4 points…

Can we attribute this improvement to the program?

Knowledge score = 9/10

A *B* Time *C* *D*

Designs Leading to Biased Results: "Historical Controls"



Problem: Control group may not be comparable

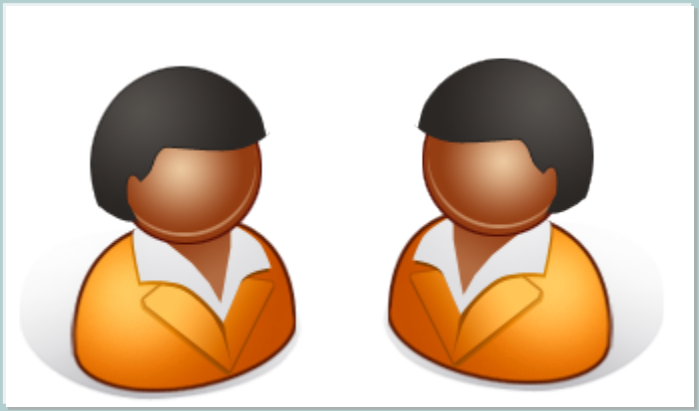
Factors (other than the intervention) may differ:

- **teachers**
- **teacher: student ratio**
- **funding**
- **textbooks**

Study Designs & Threats to Validity

5. Matched Designs

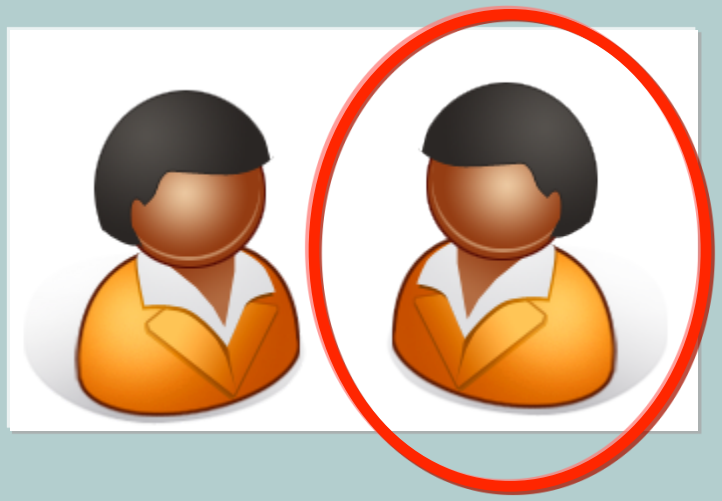
Designs Leading to Biased Results: "**Matched Designs**"



Individuals, groups, or communities are matched based on known characteristics to improve comparability:

- Age, race, sex
- Region, poverty

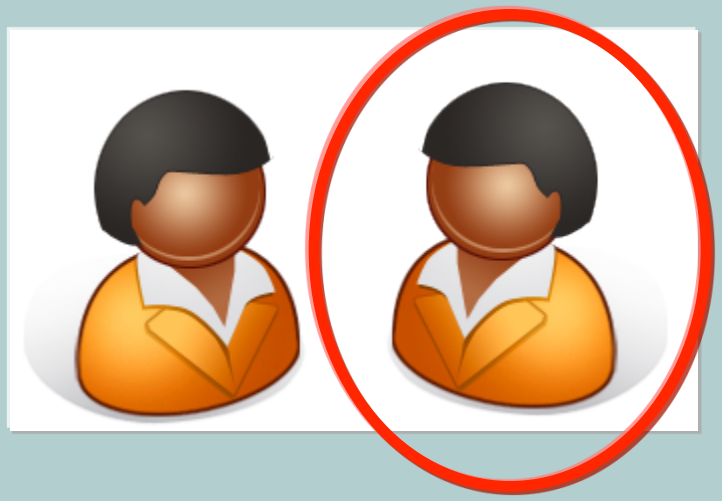
Designs Leading to Biased Results: "**Matched Designs**"



From each pair, one receives the intervention

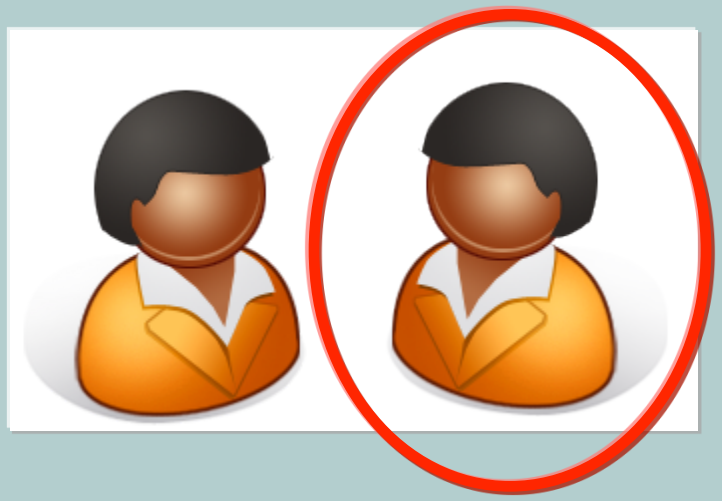
Differences in outcomes are compared within the pair

Designs Leading to Biased Results: "**Matched Designs**"



Does this design ensure that the matched pairs are comparable on all factors *except the intervention*?

Designs Leading to Biased Results: "**Matched Designs**"



Does this design ensure that the matched pairs are comparable on all factors *except the intervention*?

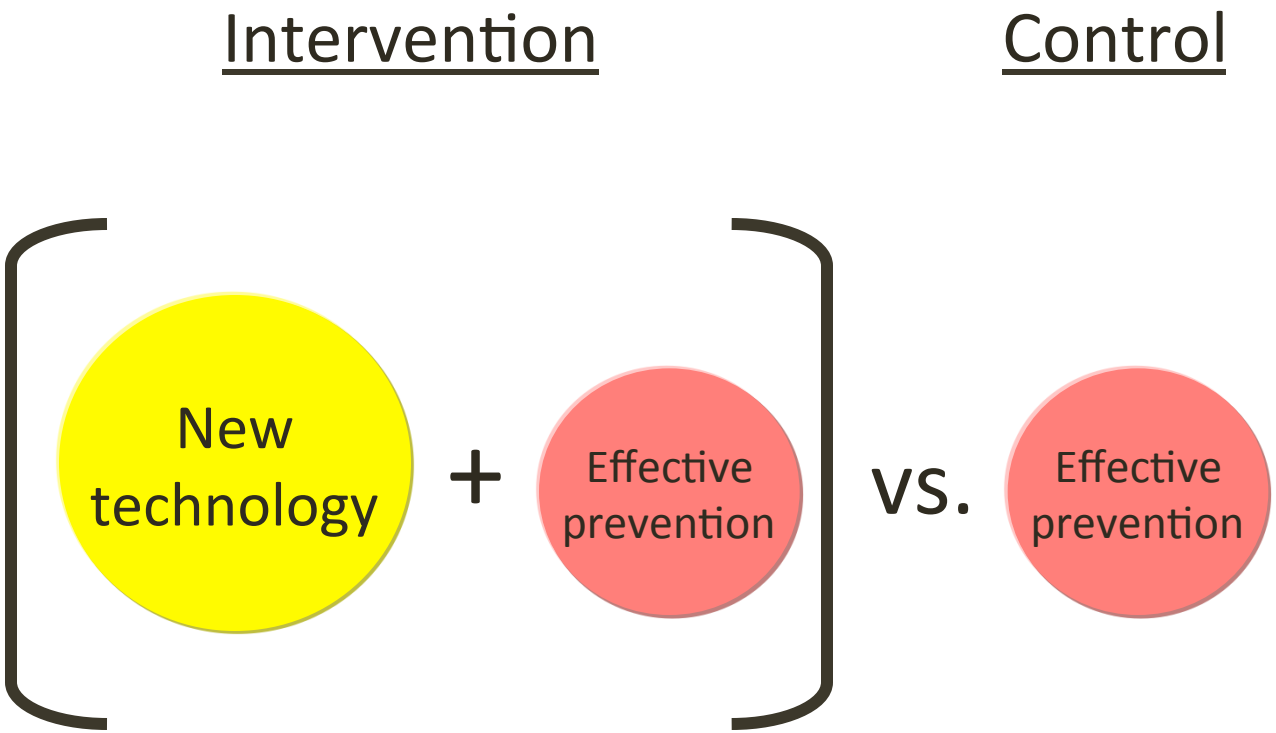
No: Only **observed** factors are used for matching

Unobserved factors may differ

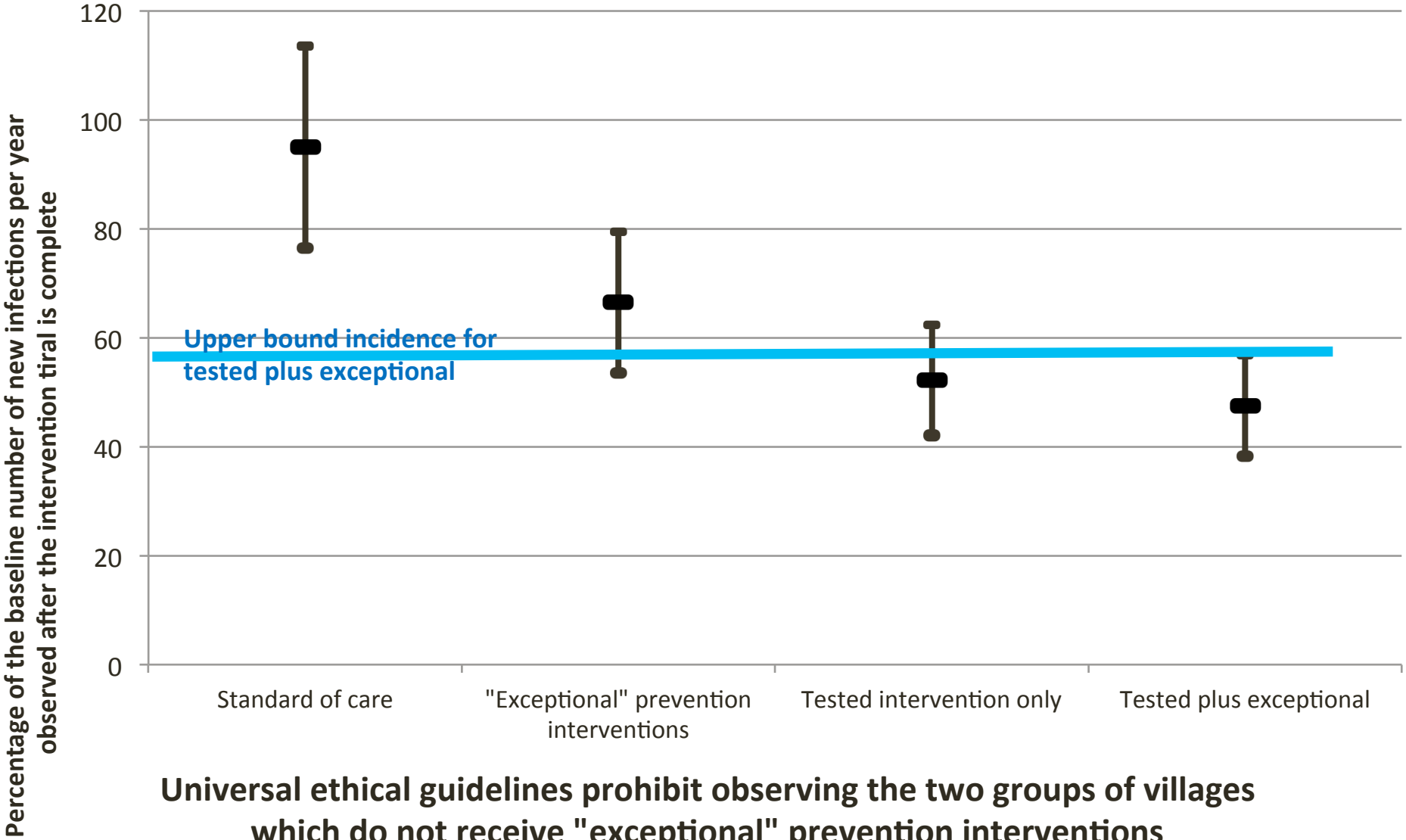
Counterfactuals & Ethical Conduct of Research

- Often our choice of counterfactuals is limited by ethics
- Lack of naive comparison groups
- When we might want to compare "something" to "nothing" we have to instead compare to "something" to "something else"

Prevention Services in Control Arm



Hypothetical results of measuring new HIV infections in four groups of villages receiving four prevention options



Universal ethical guidelines prohibit observing the two groups of villages which do not receive "exceptional" prevention interventions

Impact Evaluation Quality Assessment

What is the question of interest?

- *"What is the impact of **P** on **Y** among population X?"*
- What is the program, intervention, or treatment, **P**?
- What is the outcome, ***Y,*** of interest?

What is the unit of analysis?

- Is impact being measured in individuals or clusters (e.g., facilities, schools)?
- Was clustering accounted for in the analysis?

Impact Evaluation Quality Assessment

What is the counterfactual?

- Is there an observed comparison group?
- If not, what are observed data being compared to (e.g., historical controls, pre-intervention baseline, modeled counterfactual)?
- How is membership in the treatment and comparison group determined?

Impact Evaluation Quality Assessment

Quality of the counterfactual

- Does the treatment and comparison groups differ only with respect to the program?
  If randomized assignment → Yes
  If non-randomized assignment → No

- What are potential sources of bias?

- What are strengths and weaknesses of this counterfactual?

During IAS sessions….

Session Title	Research Question	Exposure/Treatment	Outcome	Unit of Anlysis (Community, individual, district etc)	If cluster, did they account for clustering in analysis?	Comparison/ Counterfactual	Strengths	Weaknesses

Questions?

Many thanks to those whose slides I've borrowed:
Nancy Padian, Stefano Bertozzi, Sebastian Martinez, Paul Gertler, Mead Over

Several slides adapted from:

Impact Evaluation in Practice

by Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch

http://publications.worldbank.org/index.php?main_page=product_info&products_id=23915