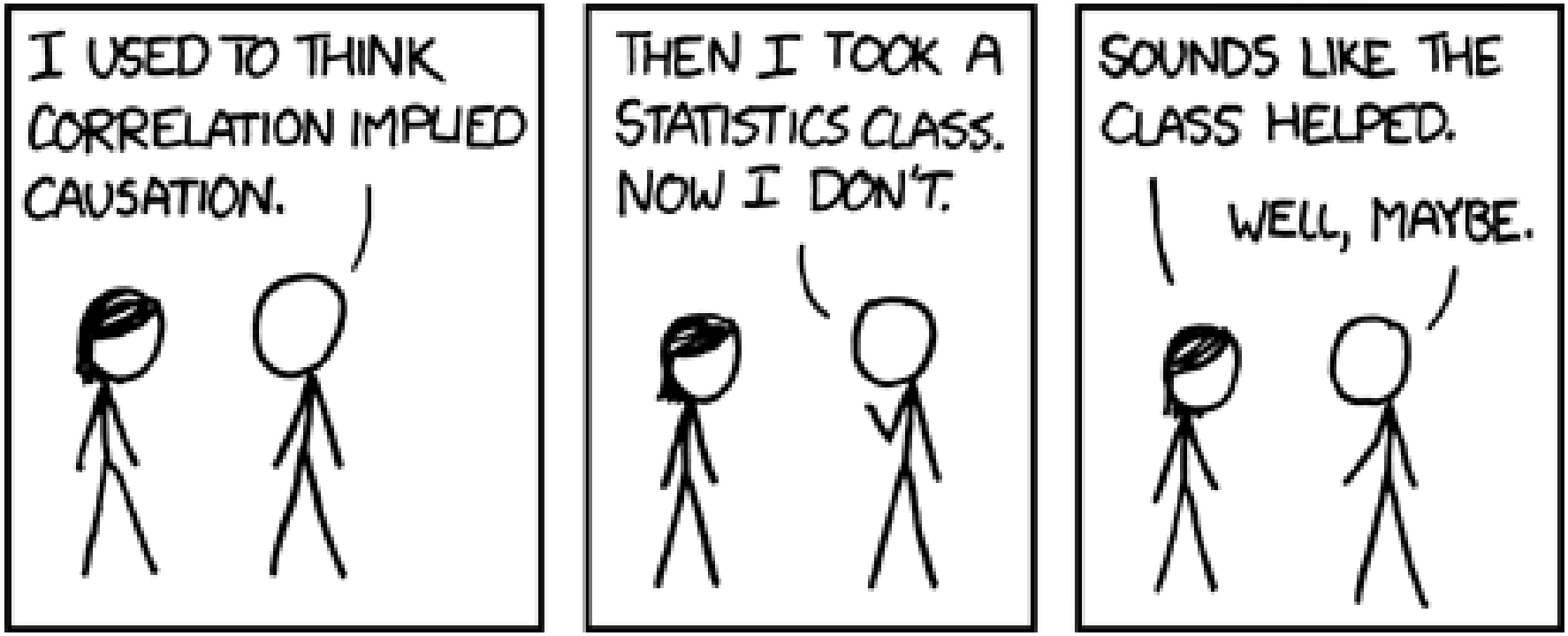
Assessing the Impact of Financial Literacy

Session 2

Nuts and Bolts of RCTs



Source: http://xkcd.com

Flipcharts in Kenya

- What was the program?
  - Flipcharts donated to village primary schools
- Where were the primary schools?
  - Western Province of Kenya
- When did this happen?
  - 1997
- Who donated the flipcharts?
  - International Child Support Africa (ICS Africa) - a local chapter of a Dutch NGO
- Who, specifically, were they meant for?
  - For students in 6^{th} - 8^{th} Grade

Source: Paul Glewwe, Michael Kremer, Sylvie Moulin and Eric Zitzewitz, "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya," *Journal of Development Economics* 74(1), June 2004, pp. 251-268.

Why were the Flipcharts Donated?

- To fulfill a resource need
  - To compensate for a shortage of textbooks
  - To compensate for a lack of teaching and learning materials in general

- To target particular students in need
  - To provide a visual supplement for students who better recall pictures than text
  - To provide an aid to students who have trouble reading/understanding English

OLS

- Say we observe that X and Y are correlated
  - e.g, X = flipcharts in schools, Y = student test scores
- Run OLS of Y on X. Find positive coefficient on X
- Can we conclude that flipcharts cause better test scores?

Problems with OLS

- Three general reasons for observed correlation between X and Y:
  - causal effect of X on Y
  - reverse causation/endogenous selection (Y causes X)
  - omitted variable bias (Z causes X and Y)

- Example: We find that more flipcharts are associated with better student performance. What could be the explanations?
  - Flipcharts cause student performance (causal effect)
  - Student performance causes flipcharts: school system sees students doing well, so they purchase more inputs for them (reverse causation)
  - A third factor, say principal or teacher quality, causes both (omitted variables)

Bias

- Estimated effect of textbooks on test scores (e.g., in regression analysis) could be biased upwards or downwards

- Upwards: higher principal/teacher quality causes more flipcharts and higher scores

- Downwards: government targeting of educational resources to poor areas causes more flipcharts, but these areas have poor scores

Can we do better than simple OLS?

- Difference-in-difference
- Instrumental variables
- Propensity score matching
- Regression discontinuity

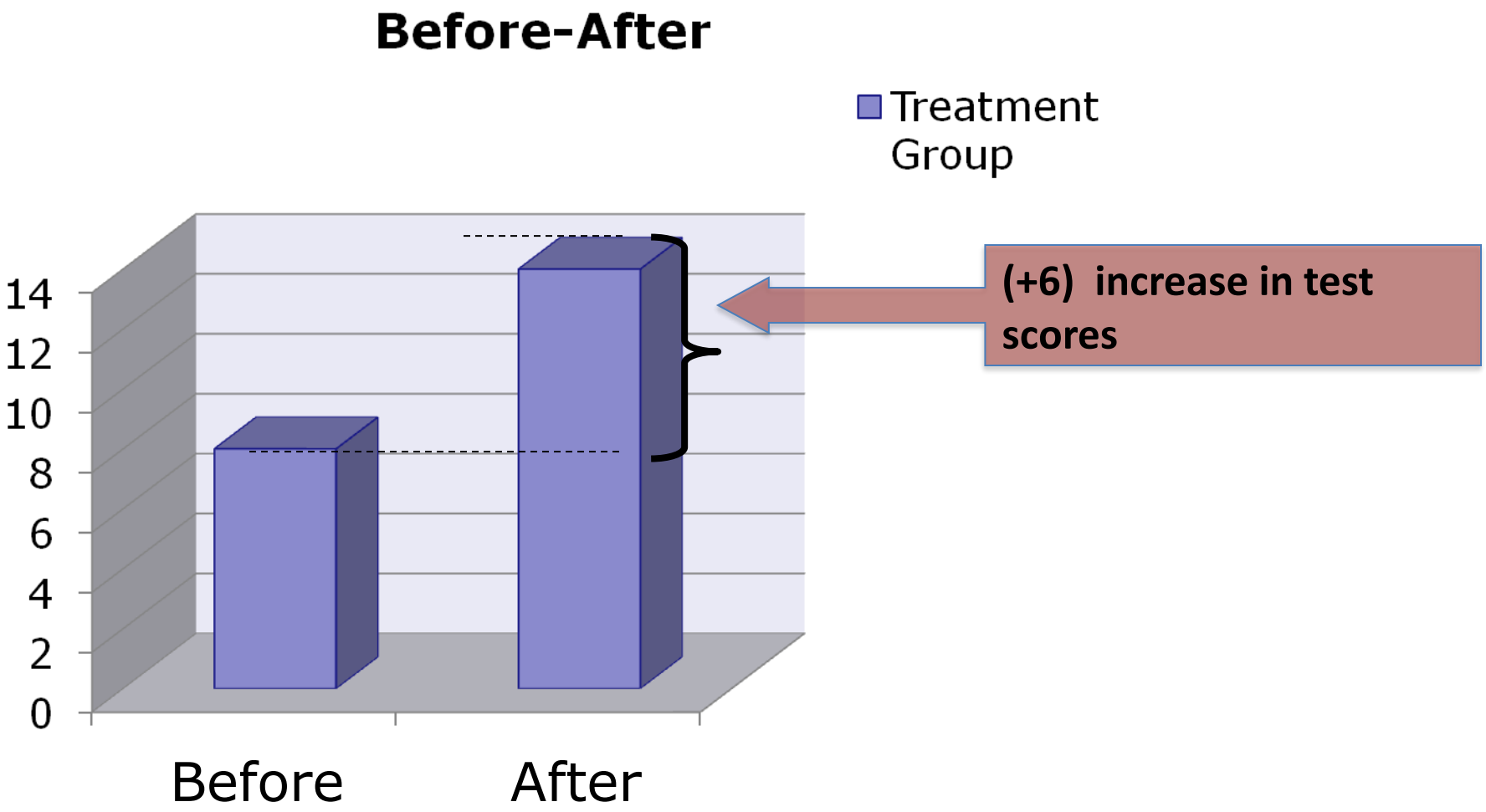
Can we do better than simple OLS?

- Difference-in-difference
- Instrumental variables
- Propensity score matching
- Regression discontinuity
- Or we can conduct a prospective evaluation!

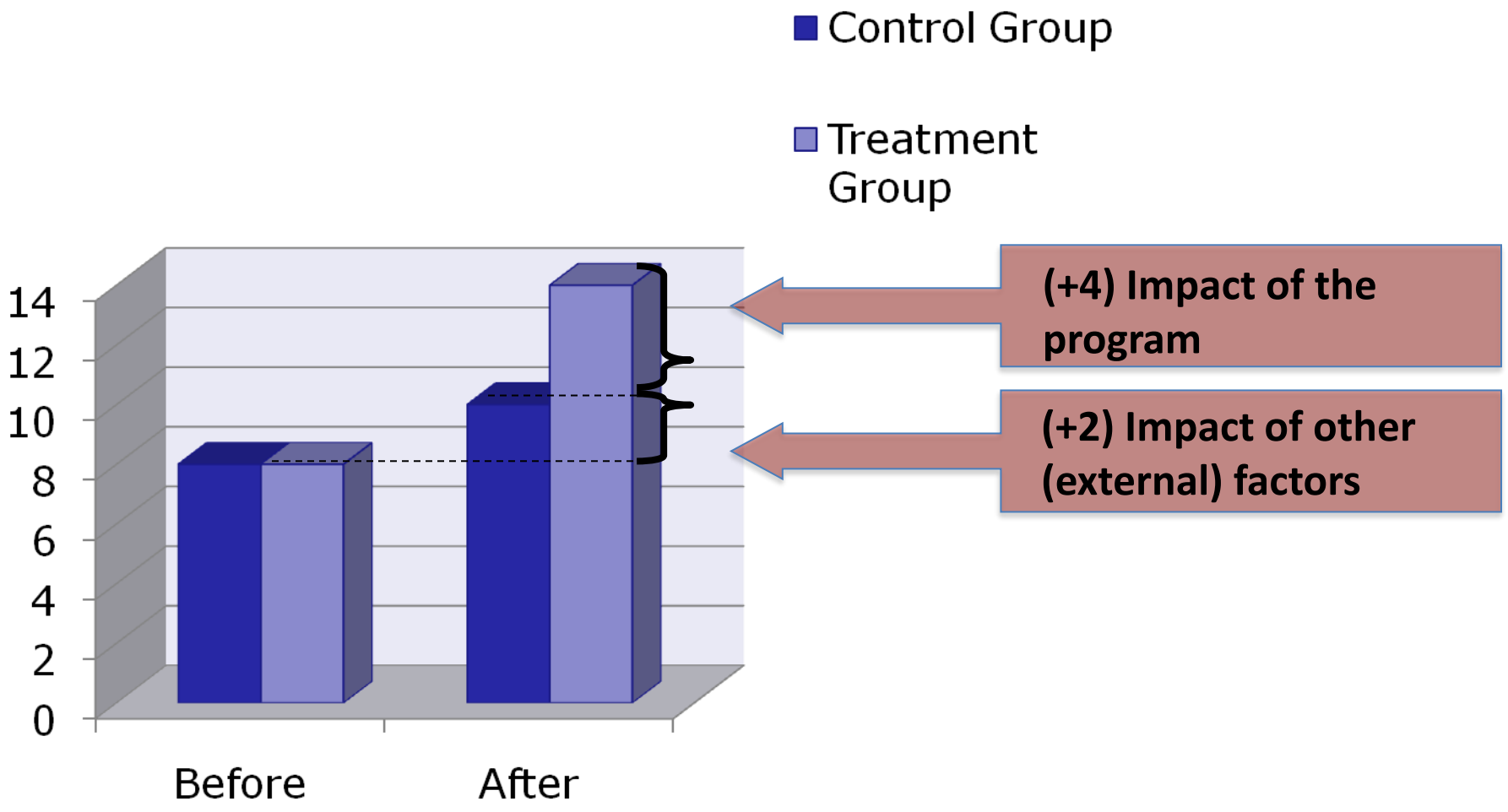
What Questions Can a Prospective Study Help Answer?

- Are flipcharts an effective teaching tool?
- Do the kids need these flipcharts?
- Are the flipcharts well designed?
- Are they being produced and distributed in cost-effective way?
- Are they reaching the schools?
- Do the teachers know how to use them?
- Are they being used?
- Are they being resold?
- What kids are we targeting?
- Do kids learn more? By what metric?
- Do the kids come to school more?
- Are these benefits commensurate with the costs? What's the social rate of return?

How Should Outcomes be Evaluated?



Is this Better?



Basic Evaluation Problem

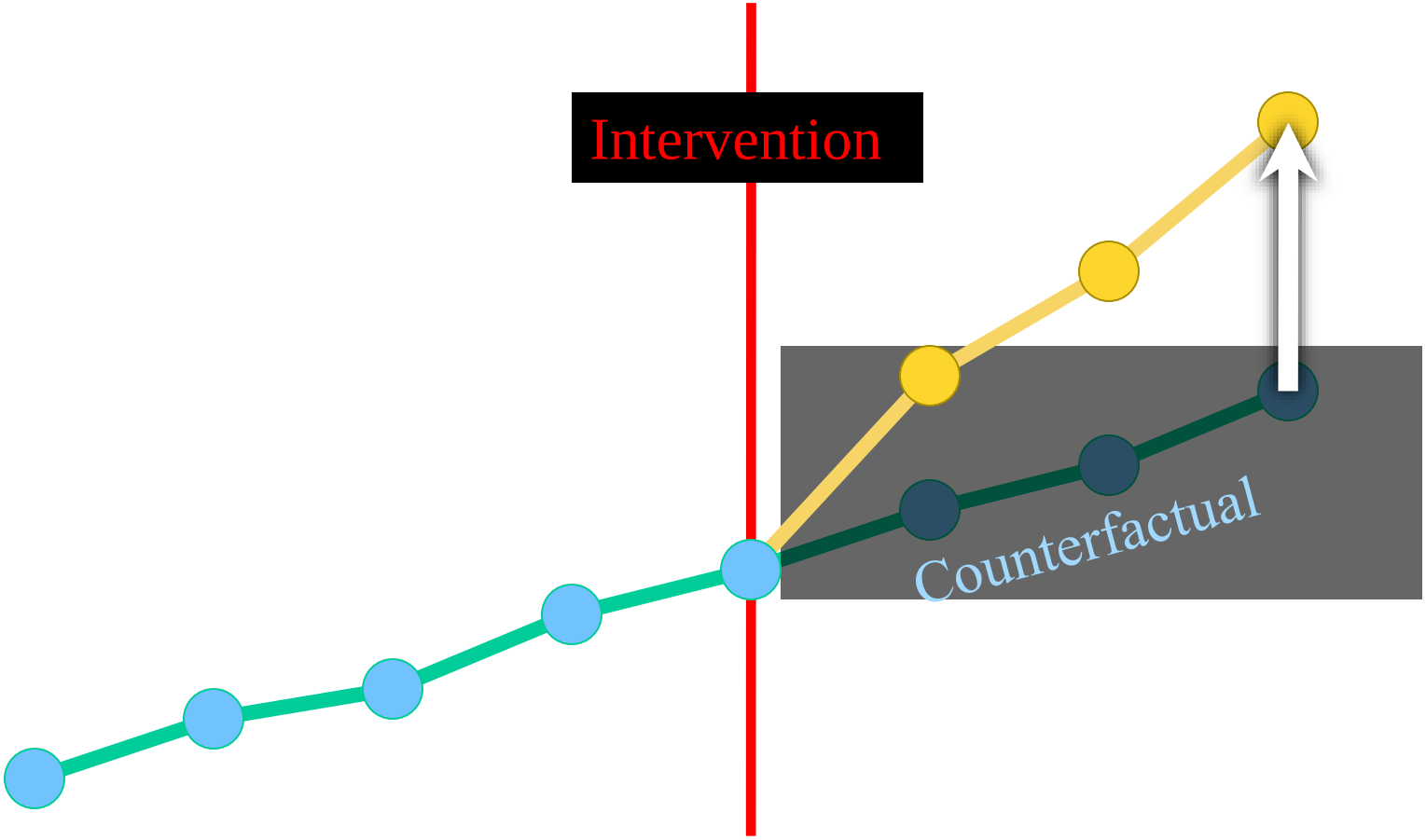
- What would have happened absent the program?
- How do we compare a world in which a program happens to a hypothetical world in which it doesn't?
  - The program either happens or it doesn't
  - We cannot create a world in which the program never happened
  - So how do we measure this "counterfactual"?
- We cannot do this exactly
- But we can identify a "similar" group, and measure outcomes in this group

Counterfactual

- We compare a group that is given the program to a group that is *not* given the program

- Terminology:
  - Treated: Group affected by program
  - Untreated: Group unaffected by program (a.k.a. "control group")

- Every impact evaluation establishes an untreated group, both retrospective and prospective.


Intervention
Counterfactual

Prospective Evaluation

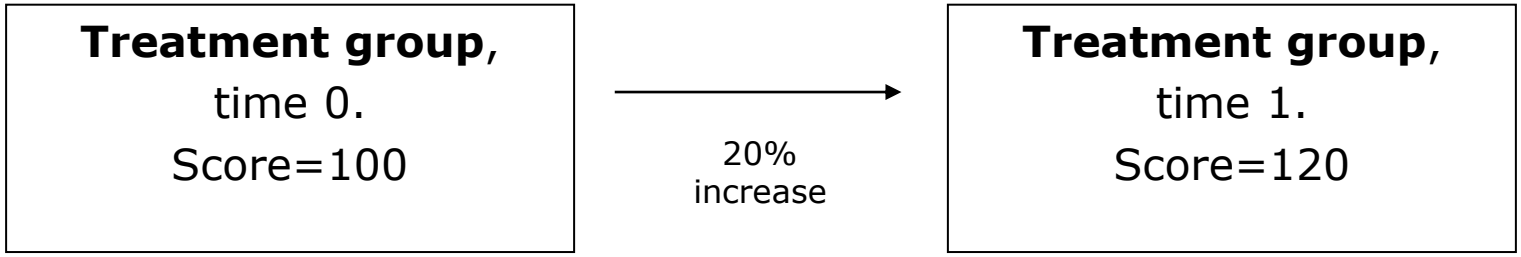
- Selection is, by definition, random and hence exogenous → BIG problem solved!

- But, can we simply look at before-after once random assignment is done?

- Before-and-after comparisons: treated individuals are simply followed over time
  - treatment schools prior to treatment used as counterfactual

- Example:
    - Provide flipcharts to schools
    - Observe that test scores in schools given the flipcharts rise by 20% over 12 months
      - Is this the *treatment effect?*

Simple Before-and-After Comparison

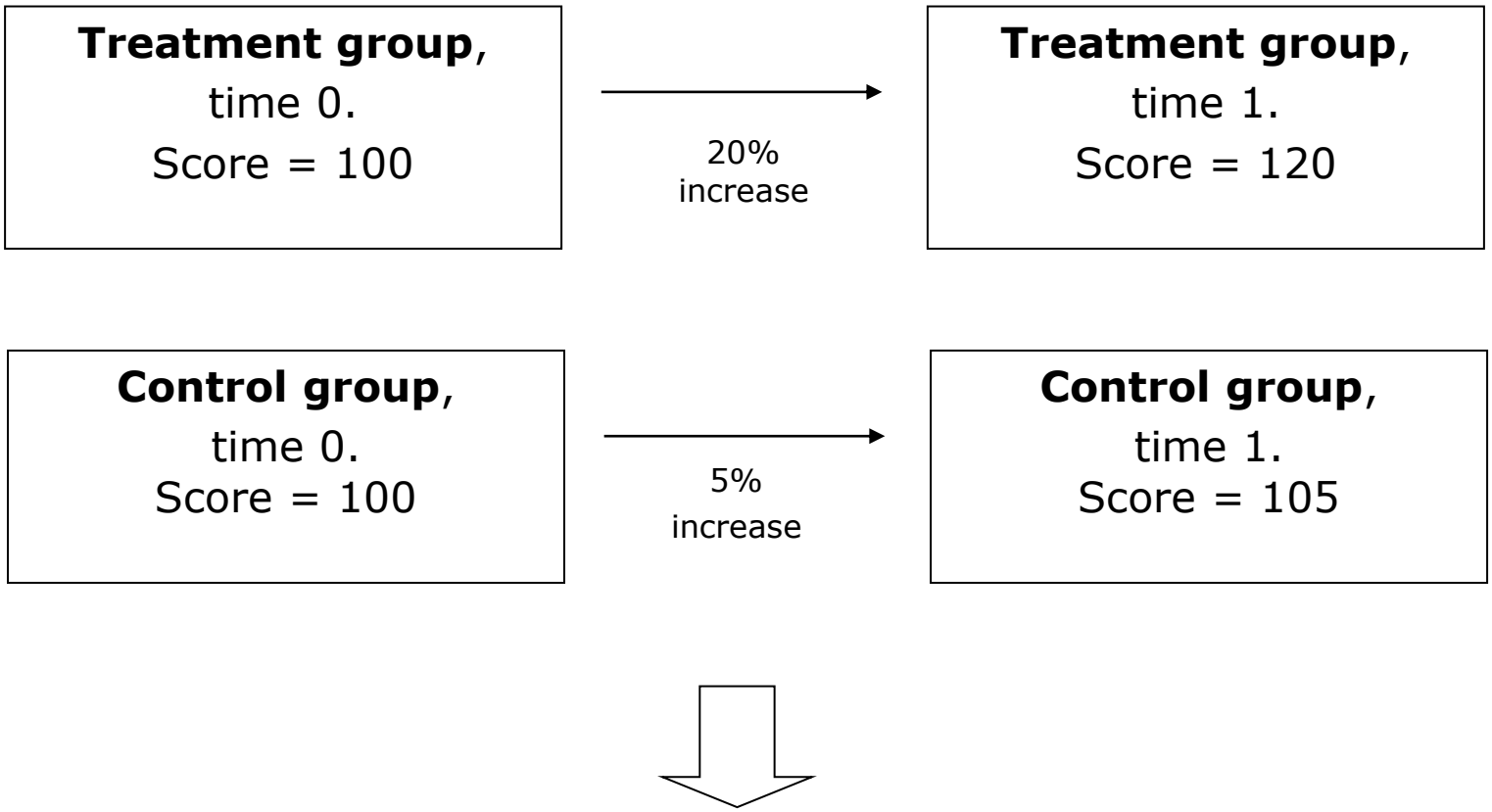


Problems with Interpretation

- What is the counterfactual? How would scores have changed *without* the treatment?

- What else happened over the 12 month study period that would have affected scores?
    - Changes in other school inputs
    - Changes in economic conditions

Treatment effect with control group



Treatment effect: 20% - 5%
= 15% improvement

Establishing the Counterfactual: Experimental Protocols

- Start with a group of eligible participants
  - Survey to establish baseline conditions
- Randomize into *treatment* and *control* groups
- Only the treatment group experiences the intervention
- Re-survey both treatment and control groups at some later date
- Treatment effect: the change for those *assigned to* treatment group *minus* the change for those assigned to control group
- Key benefit: control group serves as counterfactual
  - Represents what would have happened to treatment group *in absence* of treatment

Key Advantage of Experiments

- Random assignment
  - The control group in an experiment is meant to mimic the counterfactual
  - With random assignment, baseline characteristics (e.g. wealth, health, literacy of parents, etc.) are identical between the treatment and control groups
  - Analogous to medical drug trials

- Because members of the groups (treatment and control) do not differ systematically at the outset of the experiment, any difference that subsequently arises between them can be attributed to the treatment rather than to other factors

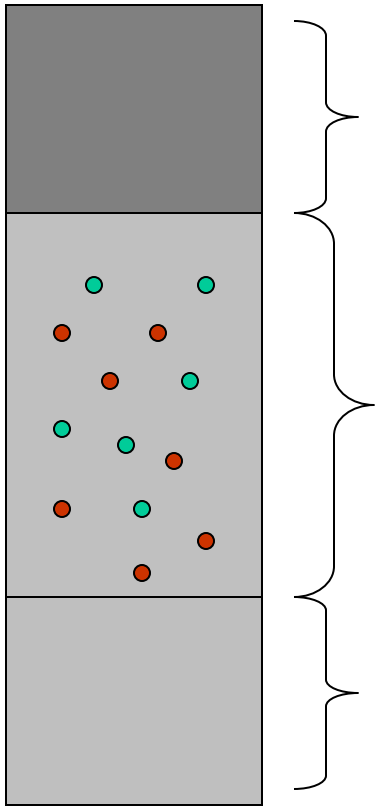
Other Advantages of Experiments

- Relative to results from non-experimental studies, results from experiments are:
  - Less subject to methodological debates
  - Easier to convey
  - More likely to be convincing to program funders and/or policymakers
  - Most "fair" way of allocating a scarce facility, product, or service

Options for Randomization

- Lottery (only some schools get the flipcharts)
  - Lottery to receive new product/service
- Random phase-in (everyone gets the flipcharts eventually)
  - Some groups or individuals get service each year
- Variation in treatment
  - More than one treatment compared to a control
  - Trt 1: Flipcharts; Trt2: Flipcharts + extra tutoring; Control
- Encouragement design
  - When "treatment" is non-excludable (such as a retirement product), the treatment group can be given extra encouragement to take up the product
  - Can separately estimate effect of encouragement (did encouragement work?) and effect of treatment

Fairly Flexible Methodology



Must receive the program

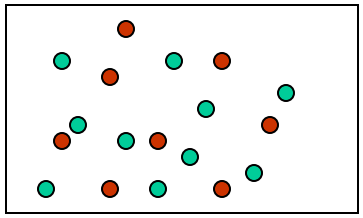
Randomize who gets the program

Not suitable for the program

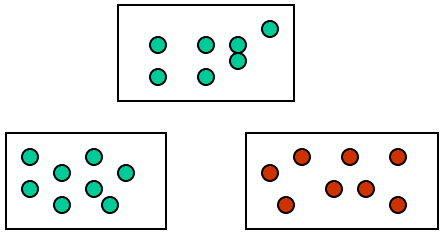
Group or Individual Randomization?

- If a program impacts a whole group-- usually randomize whole community to treatment or comparison
- Easier to get big enough sample if randomize individuals

Individual randomization



Group randomization



Unit of Randomization

- Randomizing at higher level sometimes necessary:
  - Political constraints on differential treatment within community
  - Practical constraints—confusing to implement different versions
  - Spillover effects may require higher level randomization

- Randomizing at group level requires many groups because of within community correlation
  - Micro-credit program to treat 100,000 people.
    Choose Senegal and Gambia, and randomly offer program in one country.
  - What do we learn?
  - Similar problem if choose only 4 or only 10 districts

Common Pitfalls to Avoid

- Calculating sample size incorrectly
  - Randomizing one district to treatment and one district to control and calculating sample size on number of people you interview
- Collecting data in treatment and control differently
- Counting those assigned to treatment who do not take up program as control—don't undo your randomization!!
- Not accounting for spillovers between treatment and control
- Not accounting for attrition that may be correlated with treatment status

Threats to Validity

- Despite great methodological advantages of experiments, they are also potentially subject to threats to their validity

- Internal validity
  - Hawthorne effects (behavior changes because subjects are being observed)
  - survey non-response (attrition)
  - no-shows
  - crossovers (contamination)

- External validity
  - Are the results generalizable to the population of interest?

- It is important to realize that some of these threats also affect the validity of non-experimental studies

Other Limitations of Experiments

- Can only test a subset of hypotheses (e.g., micro-level interventions)
- Low take-up (threat to power)
- Costs (?)
- Ethical issues
- Partial equilibrium

When NOT to do a Randomized Evaluation

- When the program is already over
- When the program is premature and still requires considerable "tinkering" to work well
- When the project is on too small a scale to randomize into two "representative groups"
- If a positive impact has been proven using rigorous methodology and resources are sufficient to cover everyone
- After the program has already begun and you are not expanding elsewhere
- When non-universal access is damaging/unethical/impossible (freedom of press, exchange rate changes)