

Advanced Statistical Methods for Observational Studies



LECTURE 08

class management



- If you're taking this class for three credits then you'll be giving a presentation on some related work:
 - Wednesday - June 01, 1-3PM
 - Friday - June 10, 2-5 PM
- Problem set 1 returned.
- Questions?

regression discontinuity



RD designs



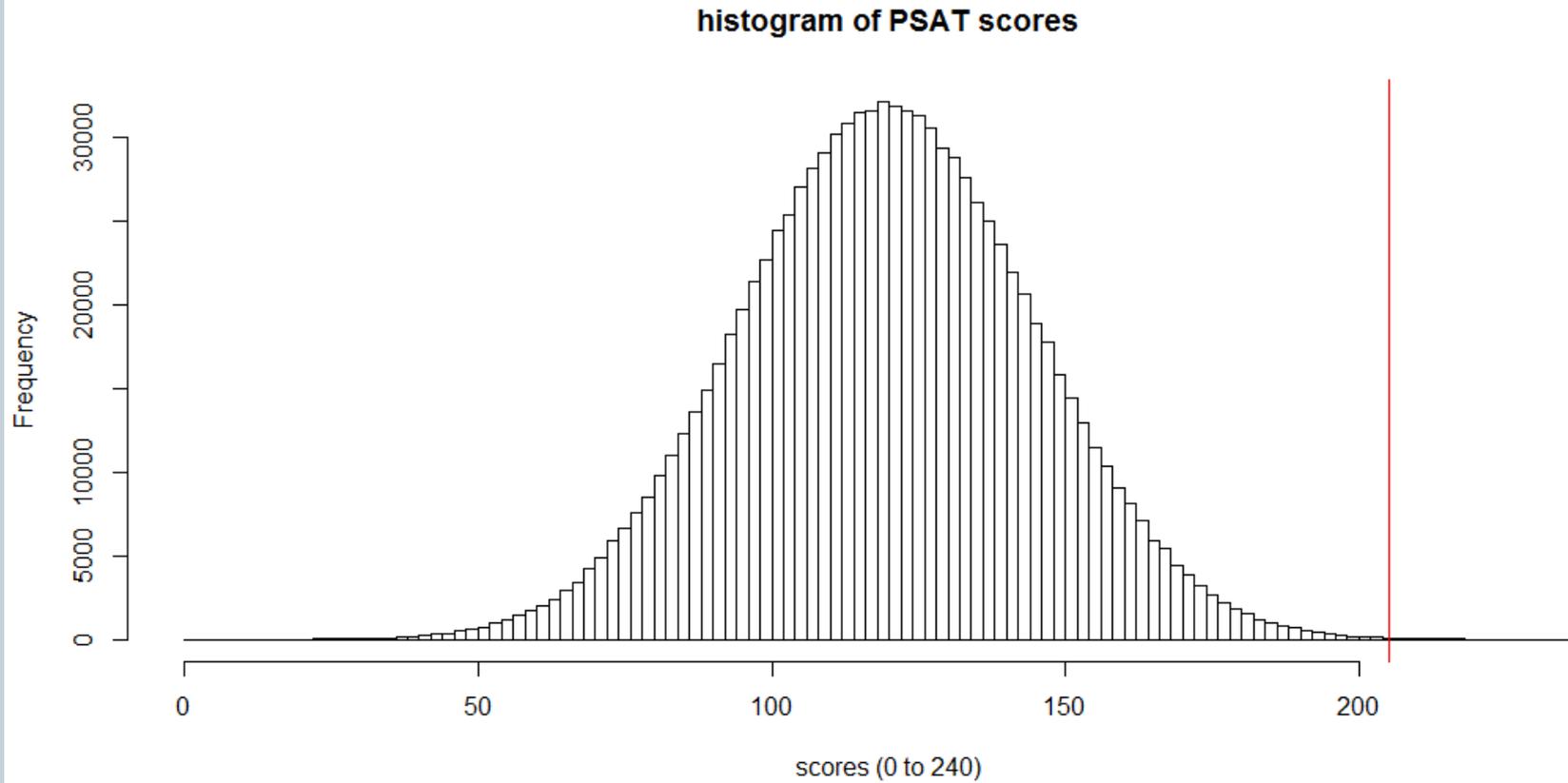
- Intuition: What if there's a known assignment mechanism and someone is just smidge on one side and someone is a smidge on the other side? How'd they end up there? Could they be the same? Couldn't I have ended up on the other side?

RD designs

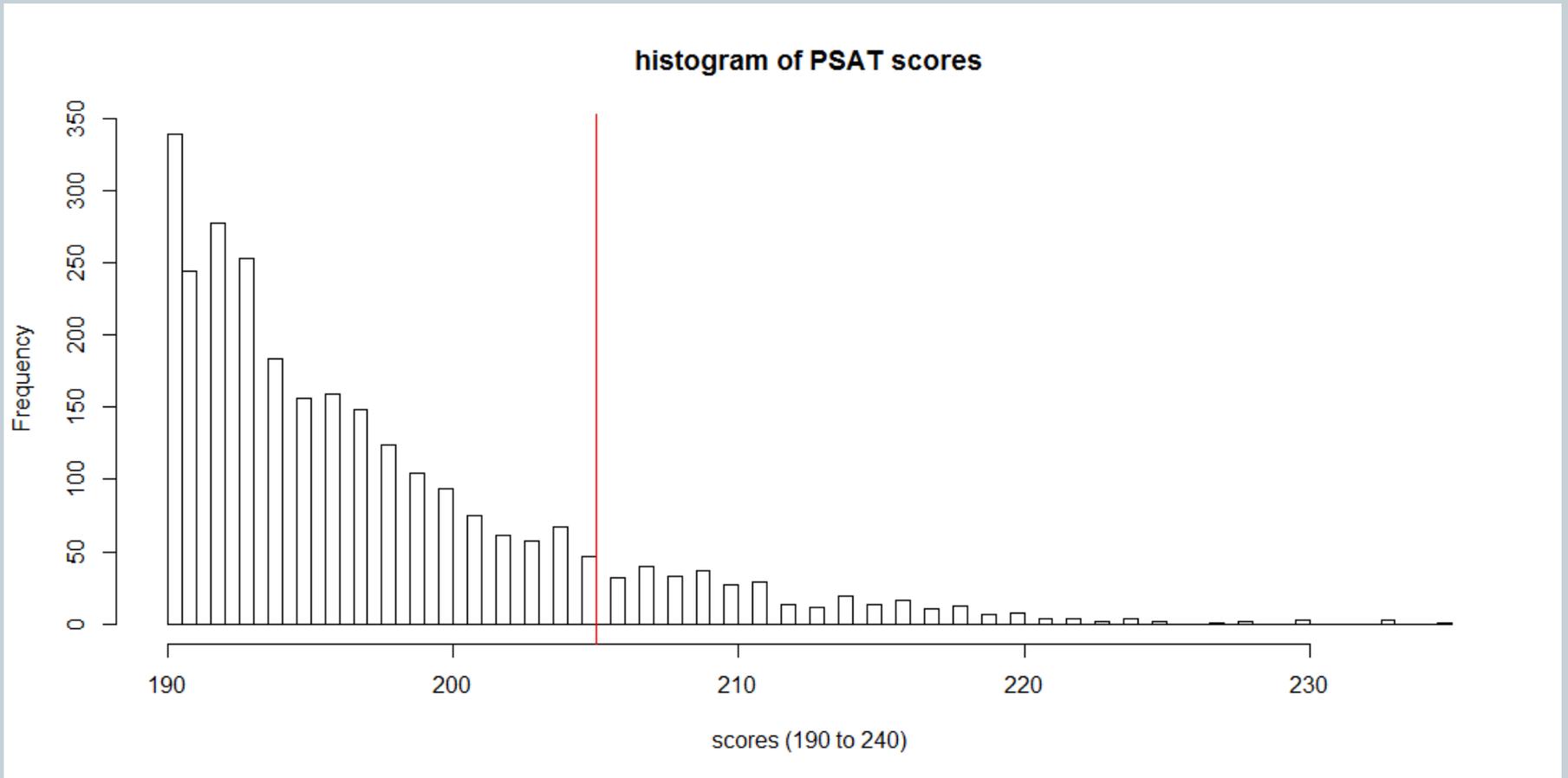


- Example: The National Merit Scholarship.
- Research question: How much benefit does the student receive from being given support for college?
- The naïve comparison is horrid: Those who work to get the NMS are outstanding and those who don't get it are a mixed bag.
- But there are millions of students who take the PSAT every year, maybe we can find a subgroup.

RD design: National Merit Scholarship



RD design: National Merit Scholarship



RD design: National Merit Scholarship

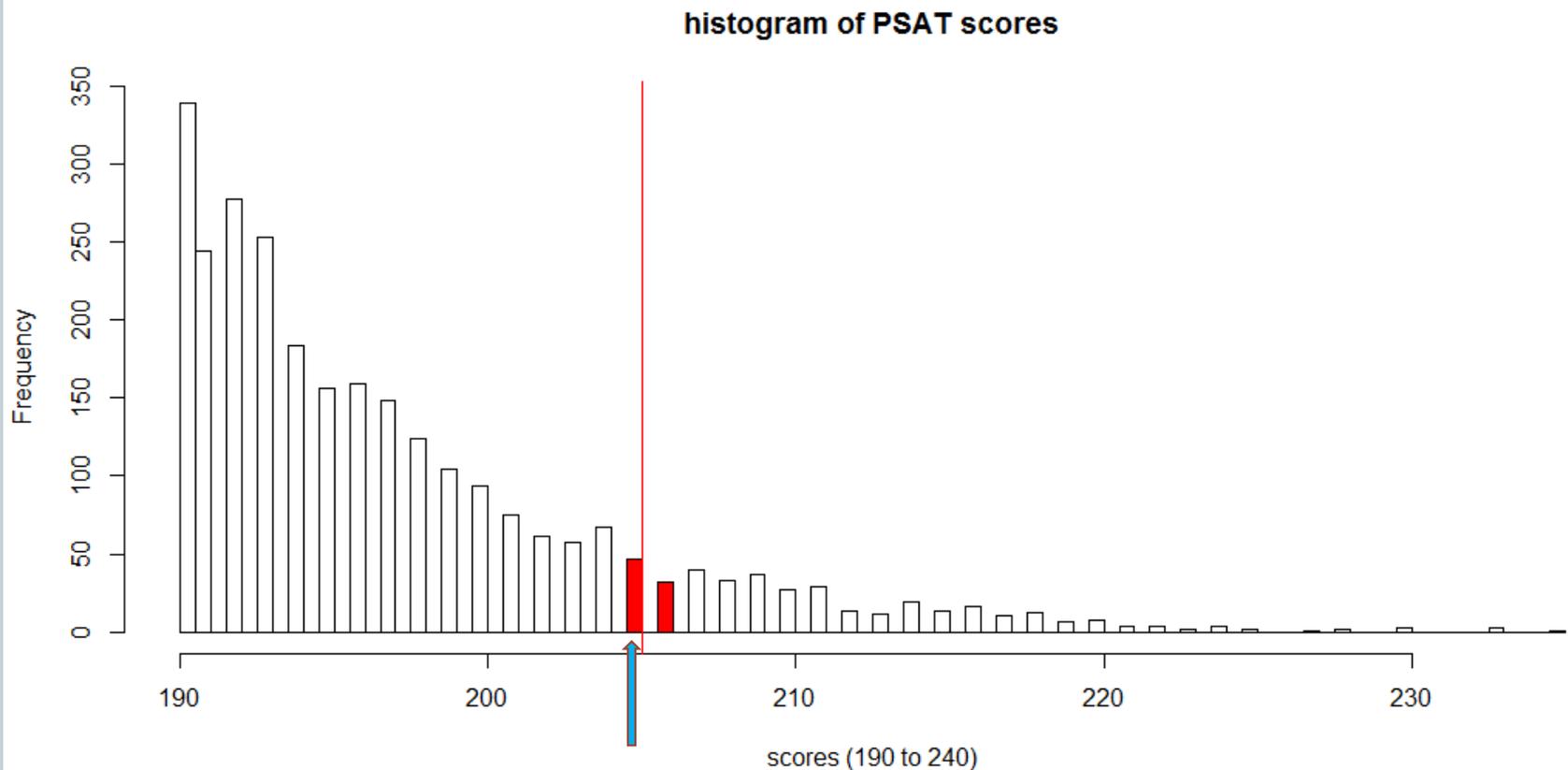


- How did they end up on one side versus the other?

RD design: National Merit Scholarship



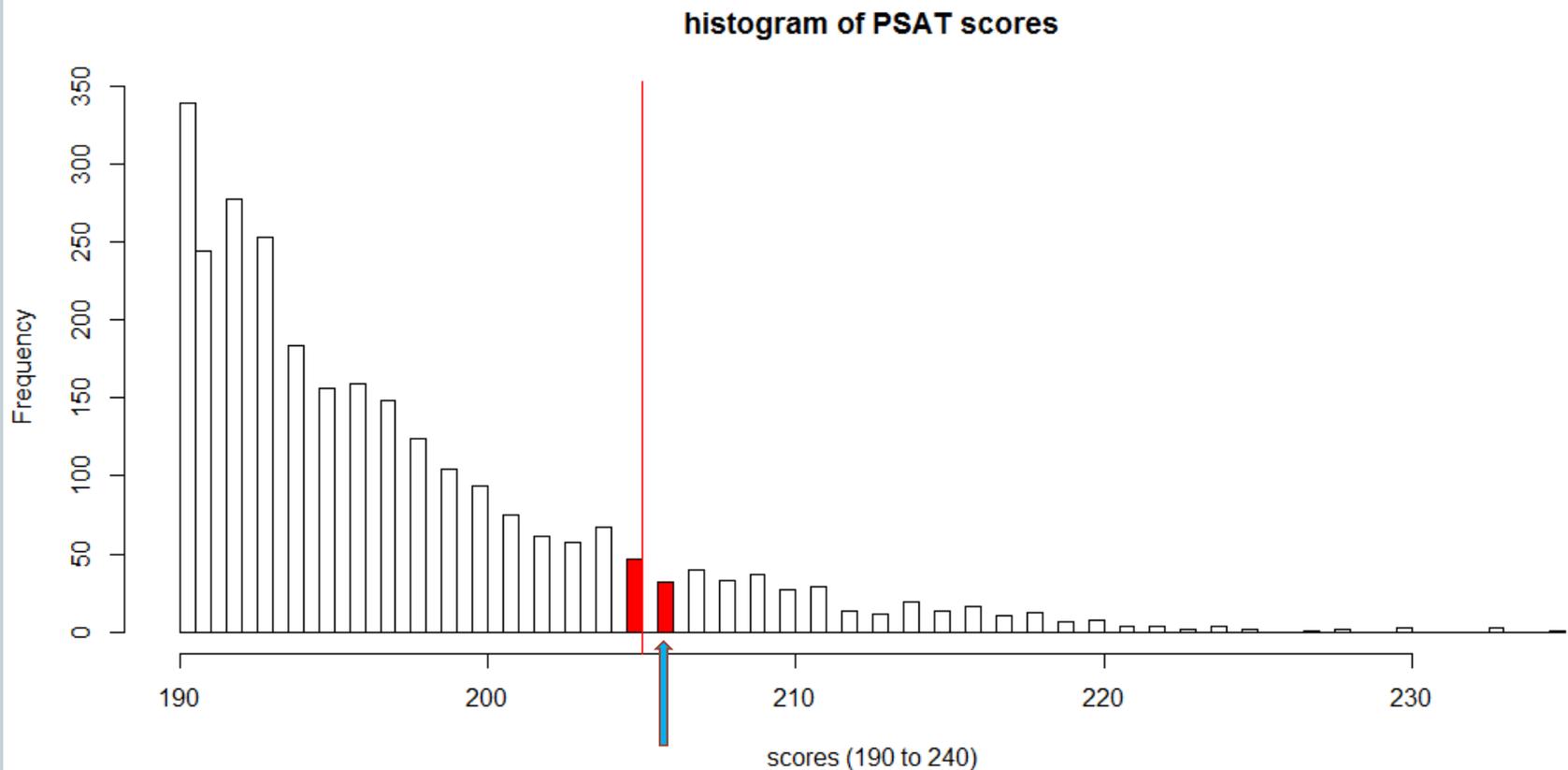
- How did they end up on one side versus the other?



RD design: National Merit Scholarship



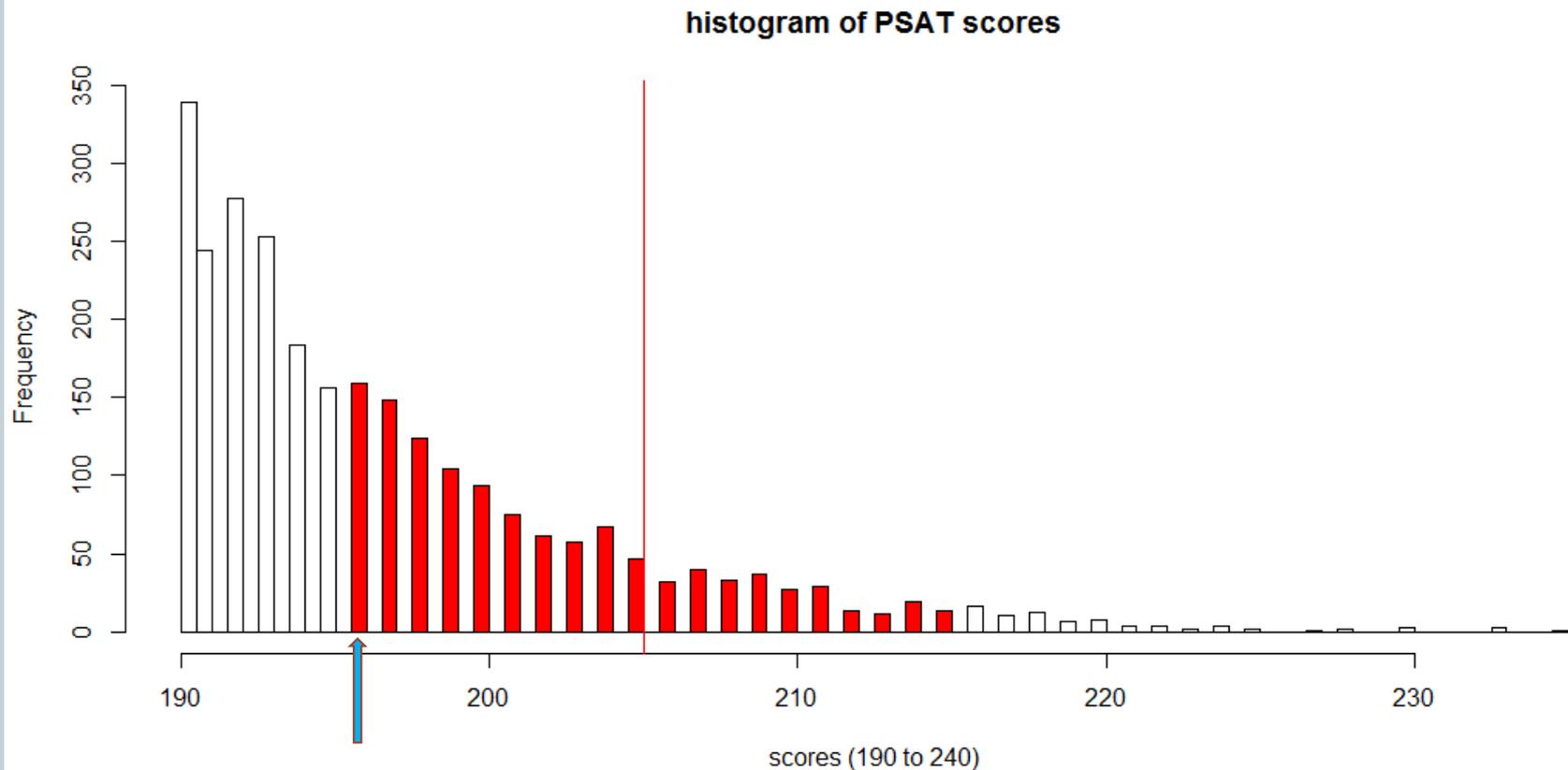
- How did they end up on one side versus the other?



RD design: National Merit Scholarship



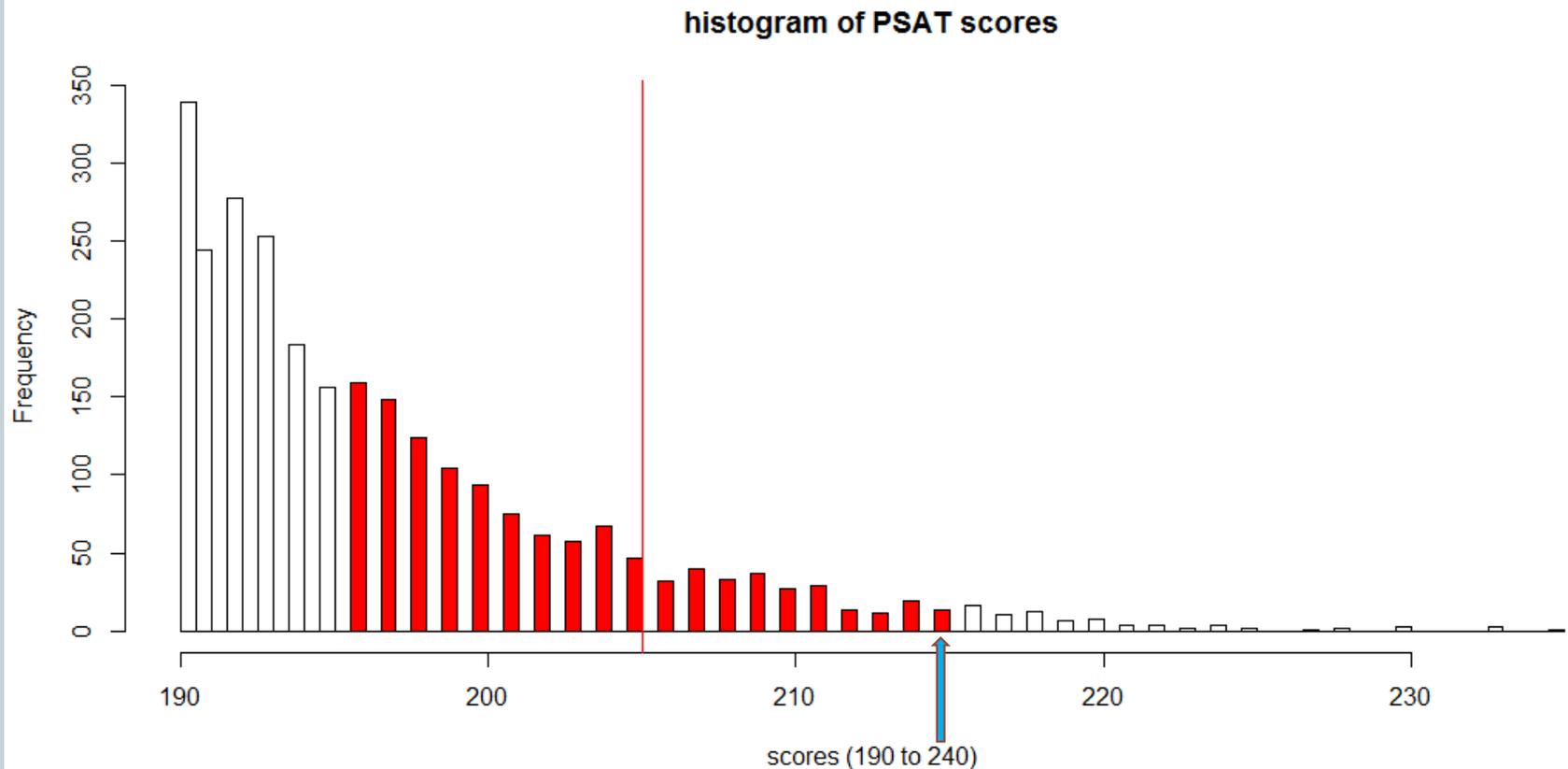
- How did they end up on one side versus the other?



RD design: National Merit Scholarship



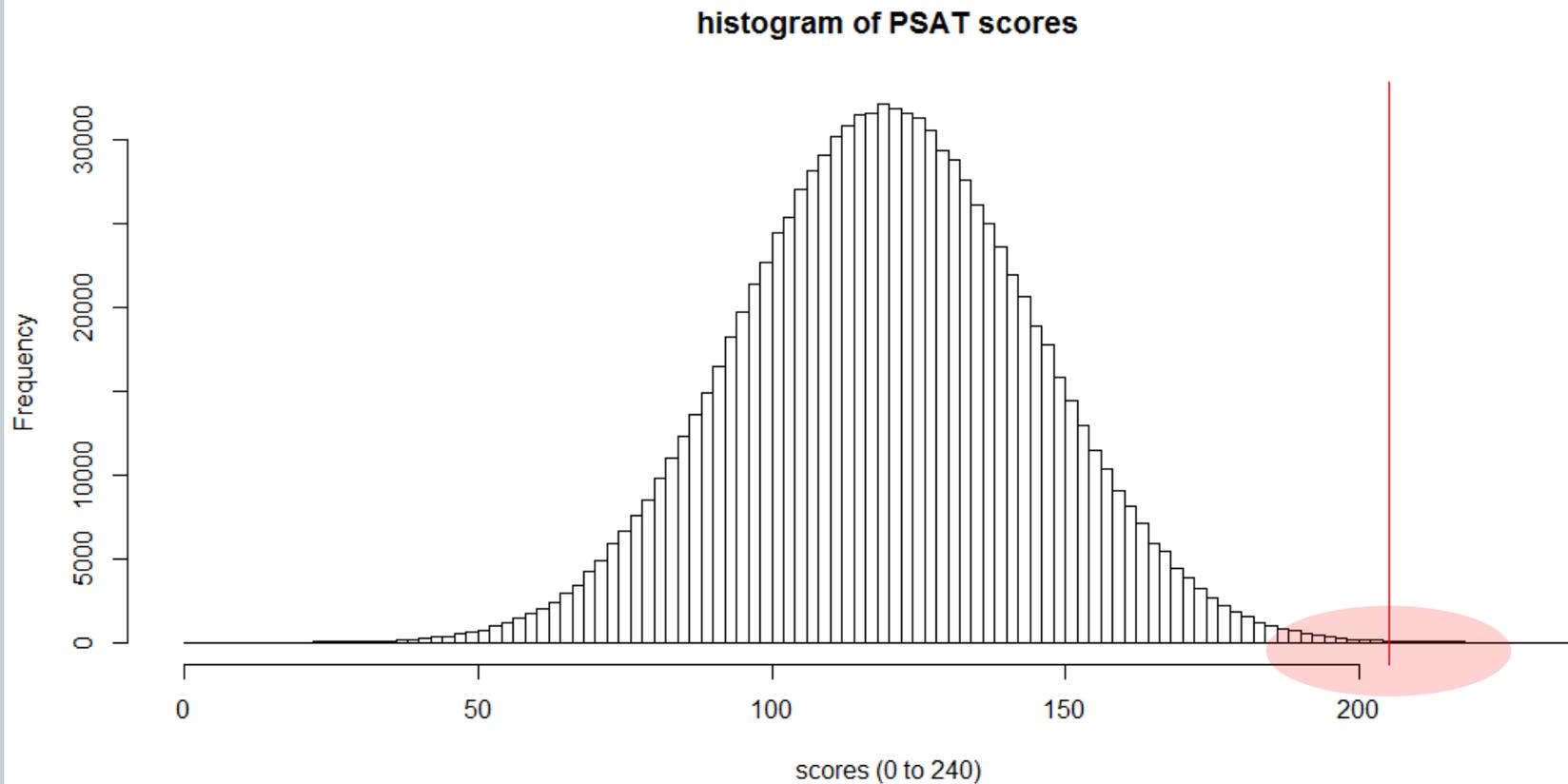
- How did they end up on one side versus the other?



RD design: National Merit Scholarship



- How did they end up on one side versus the other?



RD designs



- Features of an RD
 - RD designs can be invalid if individuals can precisely manipulate the “assignment variable”.
 - RD design is analogous to a “local” randomized experiment.
- The “randomness” comes from the lack of precise control above and below the cut off line (e.g., didn’t eat breakfast that morning)
 - Could be thought of as $y_{i,j} = \theta_i + \varepsilon_{i,j}$.
 - (observed score) – (cut off) = (randomness)
- The *localness* is really important. Think: the exact cutoff point is a bit arbitrary, but it’s being made in a covariate that is really important and meaningful for the context.

RD design: National Merit Scholarship



- Inference: RDs are usually analyzed assuming random assignment above and below the cutoff point.

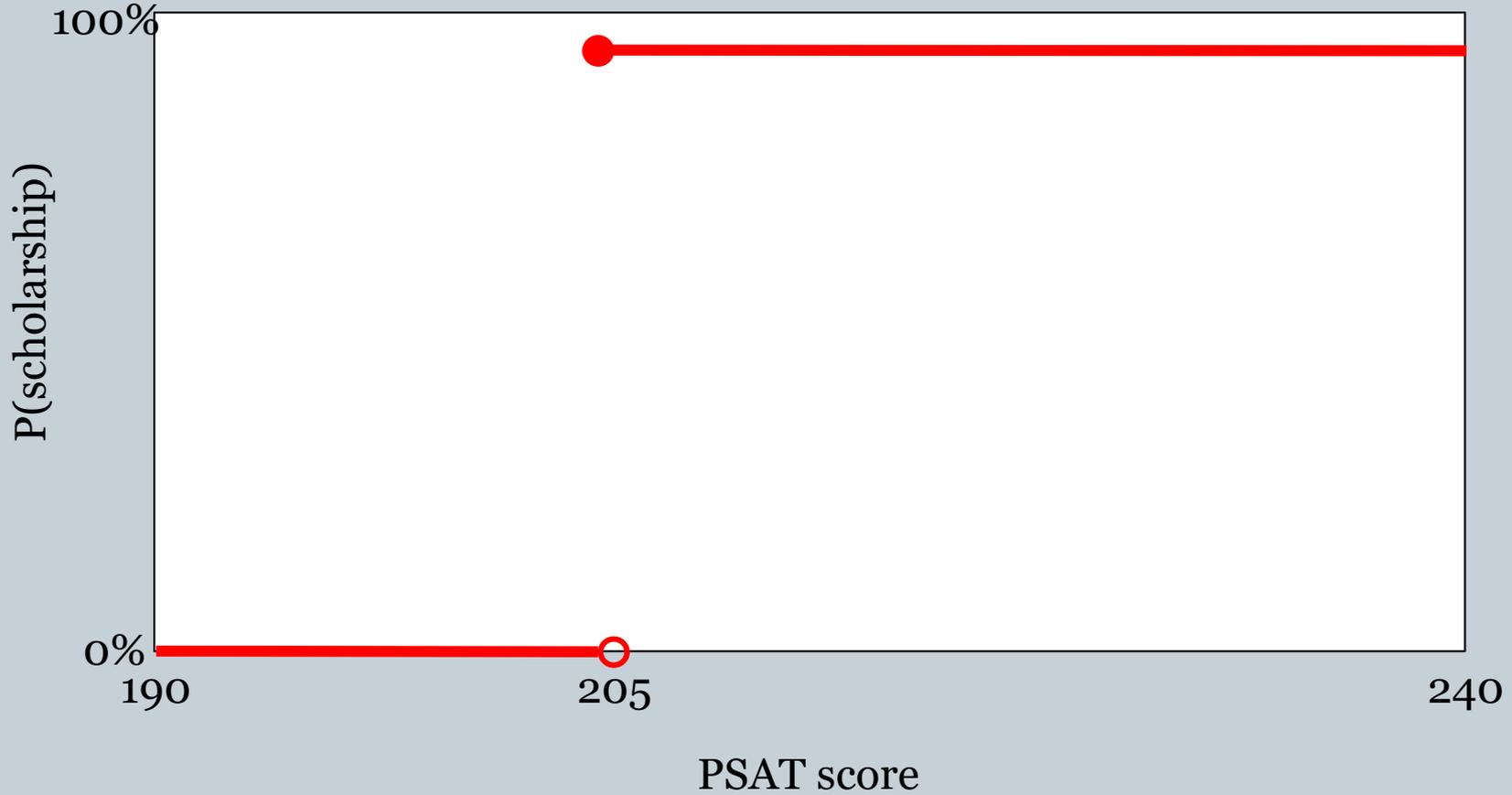
RD design: National Merit Scholarship



- Inference: RDs are usually analyzed assuming random assignment above and below the cutoff point.
- While the argument is that being above or below the cutoff is more or less random, you can enhance your argument by verifying in the covariates.
- Consider matching individuals on covariates.
 - Then you can perform a permutation based test (e.g., Wilcoxon signed rank test).
 - Perform a sensitivity analysis.
 - Looks much like what we learned in pscore.
- Many economists will use some kind of SEM:

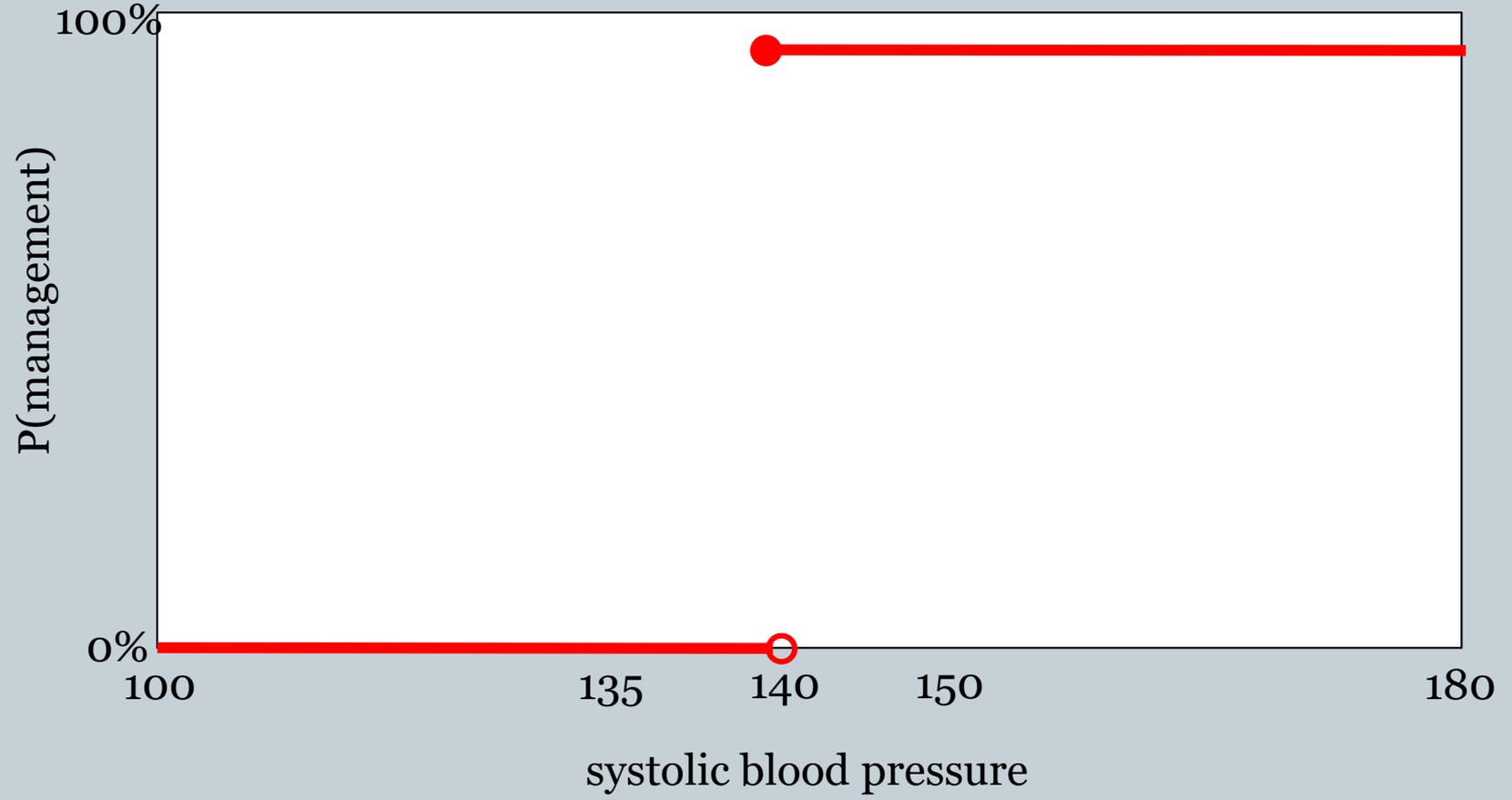
$$y_{i,j} = \theta_i + \beta * d_j + \varepsilon_{i,j}$$

RD designs



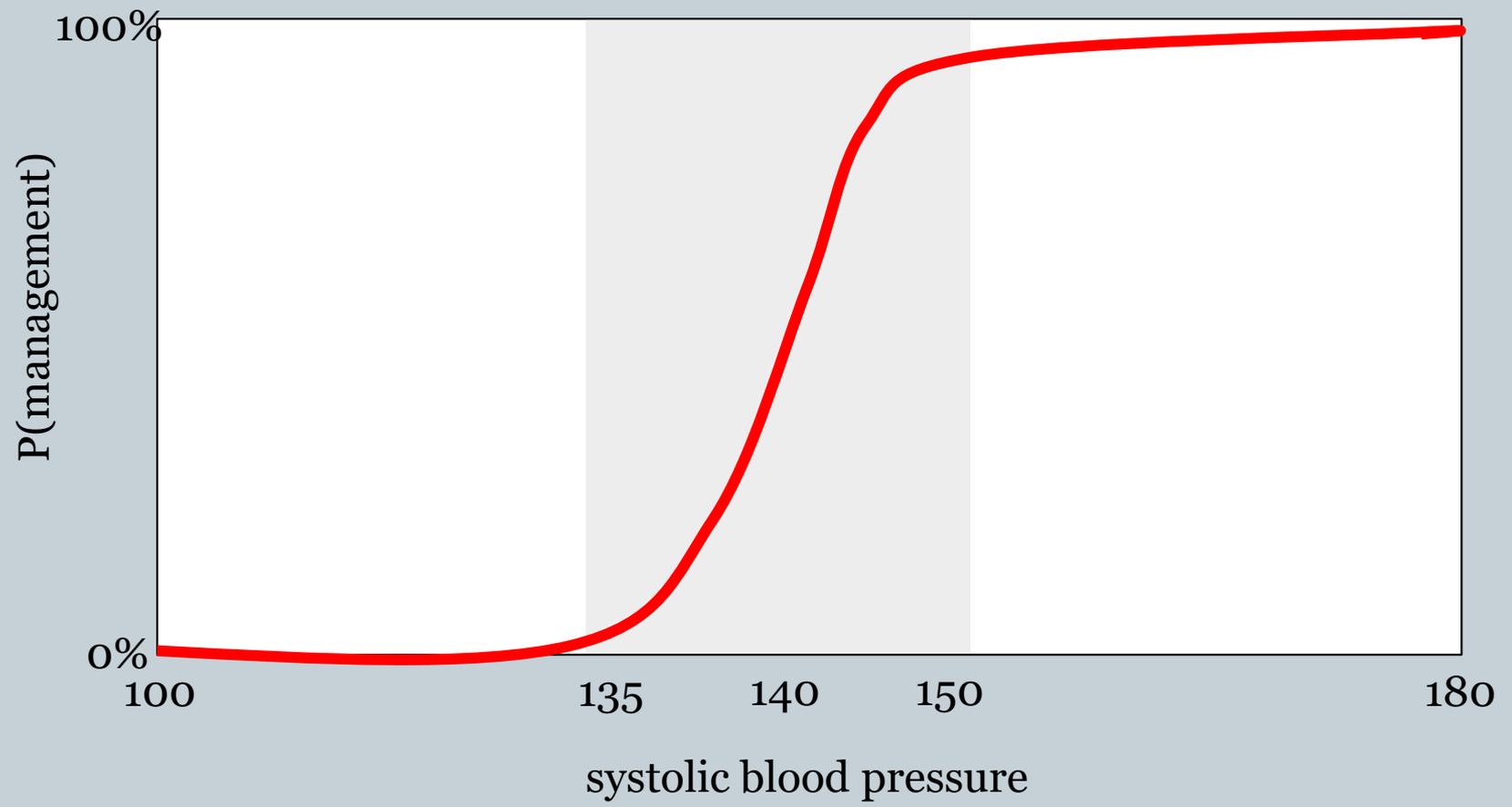
fuzzy

RD designs: Blood Pressure



fuzzy

RD designs: Blood Pressure



RD designs



- Connection to IVs: in the fuzzy regression discontinuity design you can see the connection to an encouragement design and to IVs.

RD designs



- Discuss the randomness
 - Comes from lack of precise control around the cutoff point.
 - Is the randomness really unconnected with the variables you are concerned may be causing confounding?
- Consider how far from the cutoff point to use.
- Do a Table 1 of above and below cutoff.
 - Consider matching on covariates to improve balance.
- Bottom line: RD is not radically different from IVs. Only really different in terms of the presentation of the randomness.
- Inference can be done like the pscore set up (sharp RD) or like an IV (fuzzy RD).

difference-in-differences



diff-in-diff



- Intuition: difference-in-difference approaches try to reduce the variation in your estimate due to unit-to-unit differences. It does this by focusing on differencing a pre-treatment and post-post measurement.
- Keep in mind: diff-in-diff is an approach that can often be paired with other aspects of design (e.g., pscore matching, IVs, RCTs).

diff-in-diff



an example



- In November of 2014 The Lancet ran a special issue devoted to understanding violence against women and girls

Series

Violence against women and girls 1



Prevention of violence against women and girls: what does the evidence say?

Mary Ellsberg, Diana J Arango, Matthew Morton, Floriza Gennari, Sveinung Kiplesund, Manuel Contreras, Charlotte Watts

In this Series paper, we review evidence for interventions to reduce the prevalence and incidence of violence against women and girls. Our reviewed studies cover a broad range of intervention models, and many forms of violence—ie, intimate partner violence, non-partner sexual assault, female genital mutilation, and child marriage. Evidence is highly skewed towards that from studies from high-income countries, with these evaluations mainly focusing on responses to violence. This evidence suggests that women-centred, advocacy, and home-visitation programmes can reduce a woman's risk of further victimisation, with less conclusive evidence for the preventive effect of programmes for perpetrators. In low-income and middle-income countries, there is a greater research focus on violence prevention, with promising evidence on the effect of group training for women and men, community mobilisation interventions, and combined livelihood and training interventions for women. Despite shortcomings in the evidence base, several studies show large effects in programmatic timeframes. Across different forms of violence, effective programmes are commonly participatory, engage multiple stakeholders, support critical discussion about gender relationships and the acceptability of violence, and support greater communication and shared decision making among family members, as well as non-violent behaviour. Further investment in intervention design and assessment is needed to address evidence gaps.

Introduction
Violence against women and girls is a global human rights violation and a substantial development challenge. It affects women throughout the world, and crosses cultural and economic boundaries. WHO estimates that

Key messages

- Evidence for interventions is highly skewed towards high-income countries, and response, rather than

Published Online
November 21, 2014
[http://dx.doi.org/10.1016/S0140-6736\(14\)61703-7](http://dx.doi.org/10.1016/S0140-6736(14)61703-7)
[http://dx.doi.org/10.1016/S0140-6736\(14\)61775-X](http://dx.doi.org/10.1016/S0140-6736(14)61775-X) and
[http://dx.doi.org/10.1016/S0140-6736\(14\)61840-7](http://dx.doi.org/10.1016/S0140-6736(14)61840-7)
Global Women's Institute (Prof M Ellsberg PhD, D J Arango MSc, M Contreras PhD), George Washington University, Washington, DC, USA; Department of Global Health (Prof M Ellsberg), and Department of Prevention and Community Health (F Gennari MPH), Milken Institute School of Public

Evidence Levels

- Promising – 6
- Conflicting – 4
- Insufficient evidence – 18
- No evidence – 9
- Ineffective – 7

Example	Type of violence	Evidence level		
		High-income countries	Low-income and middle-income countries	
Response to violence against women				
Women-centred programmes for survivors*	Psychosocial counselling, post-exposure prophylaxis and emergency contraception as needed, risk assessment, referrals, safety planning	IPV, NPSA	Conflicting	Insufficient evidence
Perpetrators programmes*	Interventions for men who assault their female partners	IPV	Conflicting	Insufficient evidence
One-stop crisis centres	Multidisciplinary crisis centres (community or hospital based)	IPV, NPSA	Not applicable or no evidence	Insufficient evidence
Shelters	Safe accommodations that provide short-term refuge and other services	IPV	Insufficient evidence	Insufficient evidence
Women's police stations	Specialised police services for survivors of violence against women, can include psychosocial counselling and referrals	IPV, NPSA	Not applicable or no evidence	Insufficient evidence
Victim Advocacy*	Case management, connection to legal services and information	IPV	Promising	Insufficient evidence
ICT services	National emergency hotlines or mobile applications	IPV, NPSA	Insufficient evidence	Insufficient evidence
Population-based prevention				
Community mobilisation*	Participatory projects, community-driven development engaging multiple stakeholders and addressing gender norms	IPV, NPSA, FGM, CM	Not applicable or no evidence	Promising
Awareness-raising campaigns*	One-off information or media efforts, billboards, radio programmes, posters, television advertisements	IPV, NPSA, FGM, CM	Ineffective	Ineffective
Social marketing campaigns or edutainment plus group education*	Long-term programmes engaging social media, social media, mobile applications, thematic television series, posters, together with interpersonal communication activities	IPV, NPSA, FGM, CM	Insufficient evidence	Insufficient evidence
Group-based training or workshops for prevention of violence against women and girls				
Empowerment training for women and girls*	School or community programmes to improve women's agency. Can include other components such as safe spaces, mentoring, life skills, or self-defence training	IPV, NPSA, FGM, CM	Insufficient evidence	Promising
Men and boys norms programming*	School programmes and community workshops to promote changes in social norms and behaviour that encourage violence against women and girls and gender inequality	IPV, NPSA	Insufficient evidence	Conflicting
Women and men*	School or community workshops to promote changes in norms and behaviour that encourage violence against women and girls and gender inequality	IPV, NPSA	Insufficient evidence	Promising

(Table 3 continues on next page)

IPV=Intimate partner violence
NPSA=Non-partner sexual assault
FGM=Female Genital Mutilation
CM=Child Marriage

kenyan slums



- In Kenya, 11.3% to 46% of women report childhood sexual assault.

- In our study, *annual* incidence of sexual assault was 18-25% among girls 13-19 (6% according to the police).

background



- In to this crisis stepped a wife and husband team, who created and implemented an empowerment and physical defense program.

the founders



the intervention: classroom-based



the intervention: classroom-based



the intervention: classroom-based



the intervention



- Delivered in a classroom environment
- Six sessions, two hours for each session
- Group-based, roleplaying and situational practice
- The girls' program is (roughly speaking) a three-pronged approach:
 - situational awareness,
 - verbal and social intervention, and
 - and self-defense.

variation in outcome

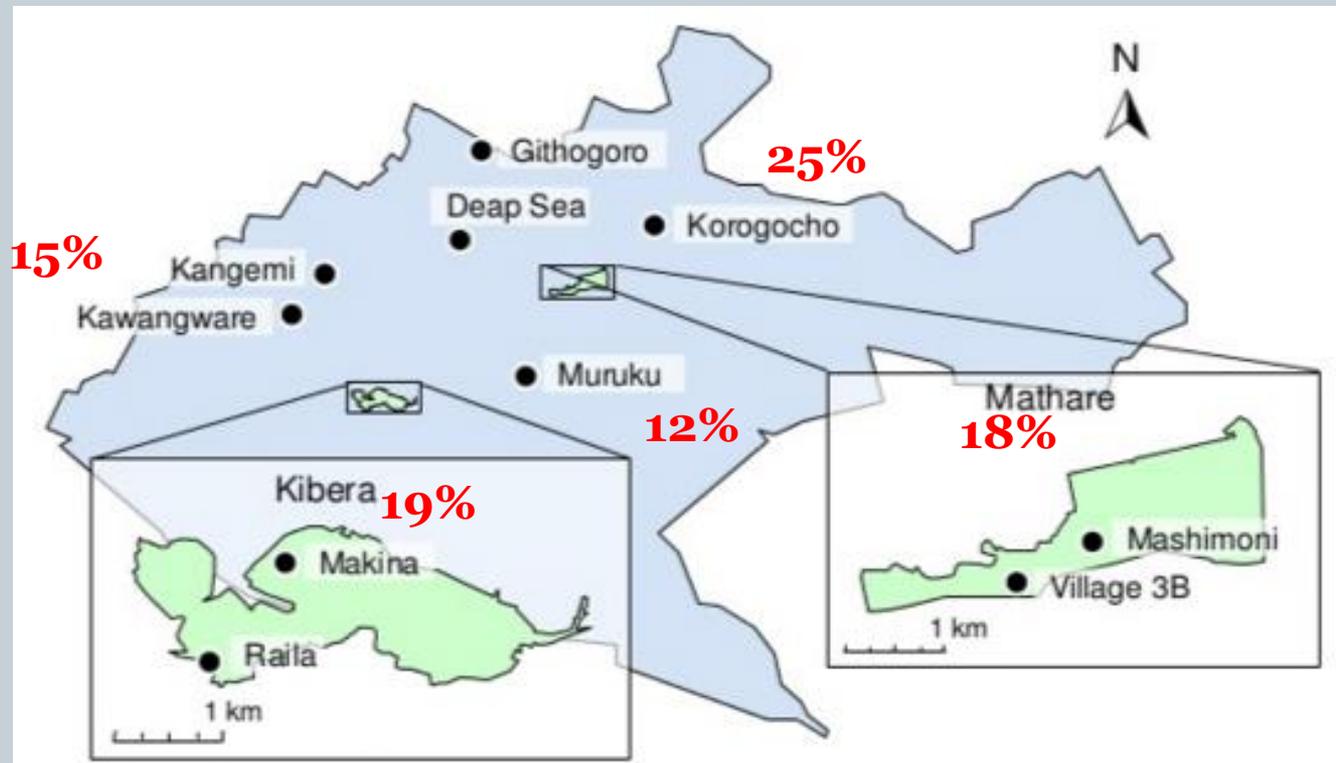


- Difference by slum

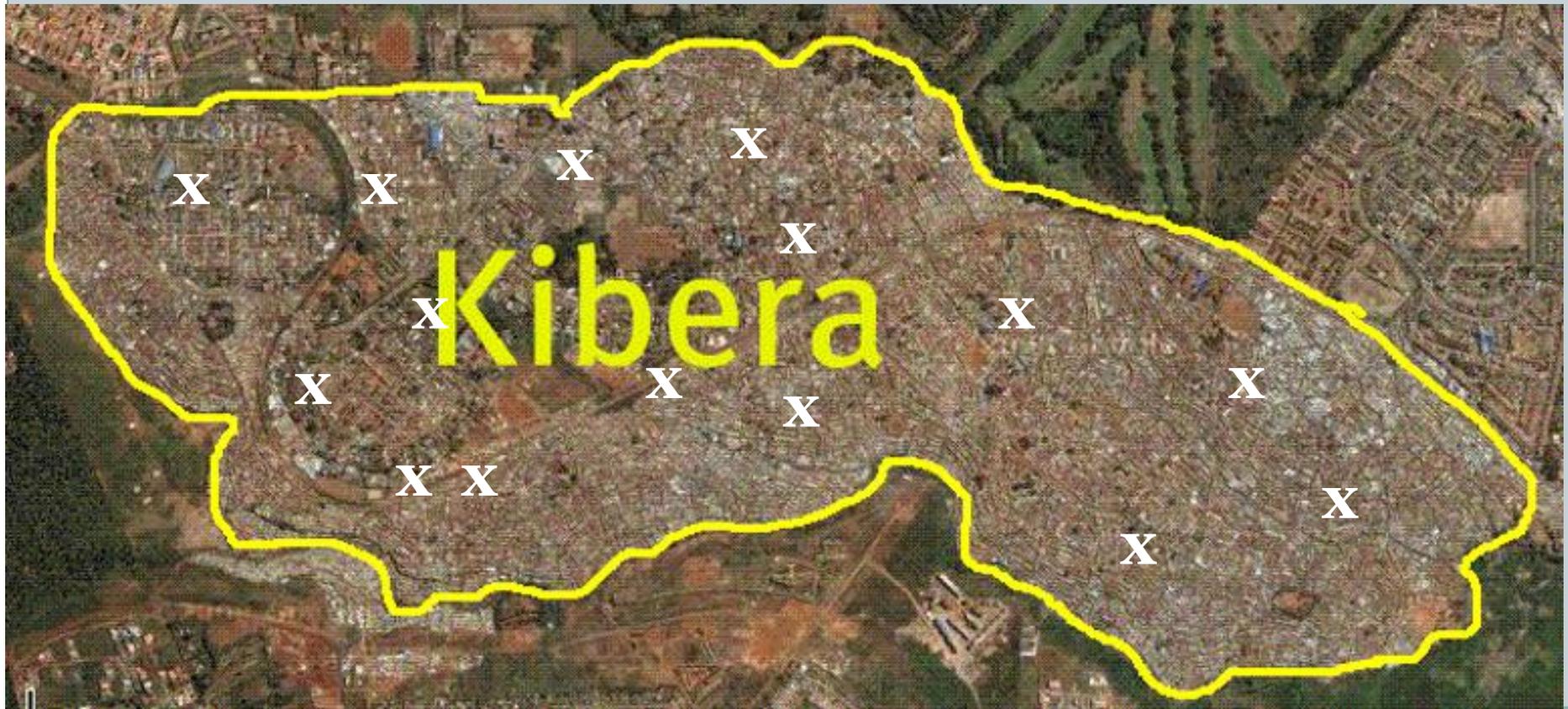
variation in outcome



- Difference by slum



mitigating school-level variation

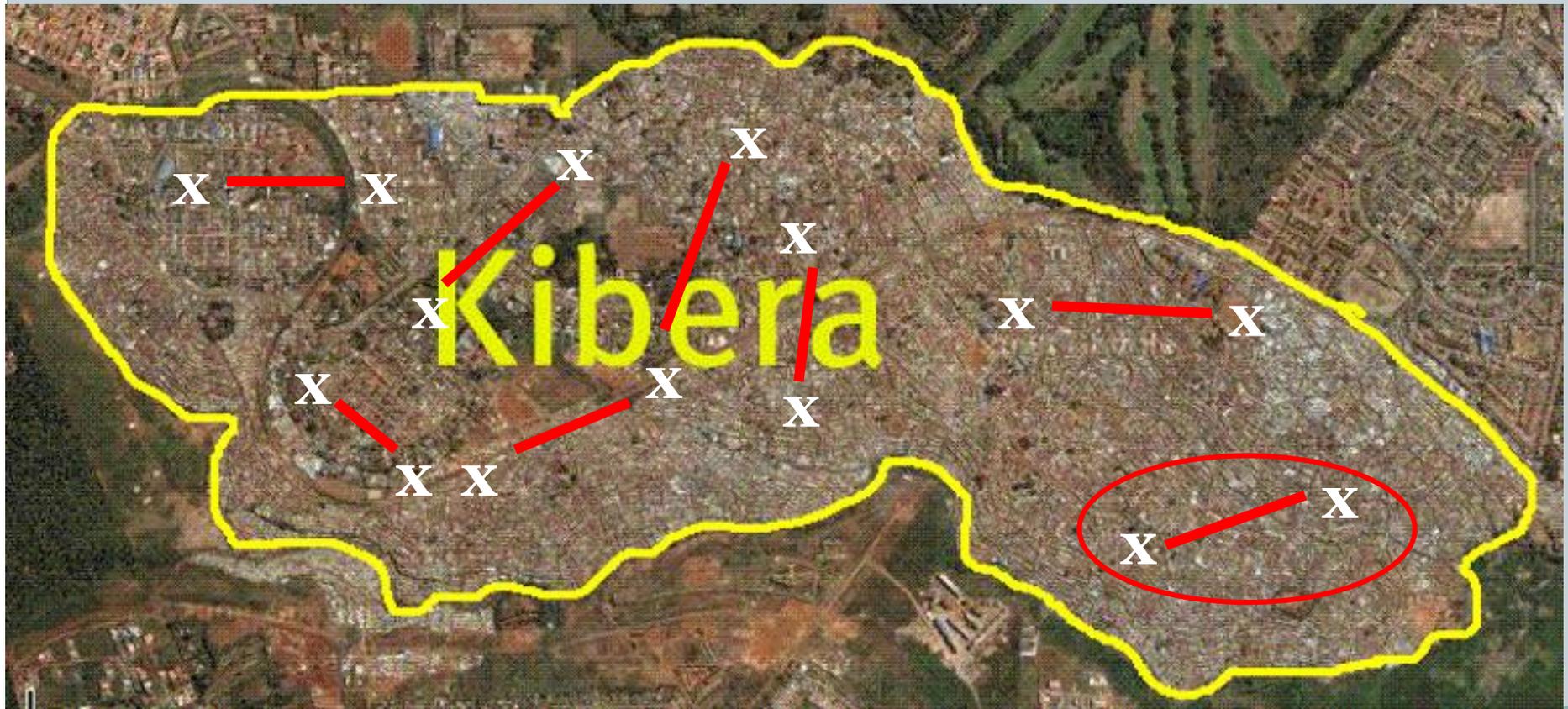


constrained randomization

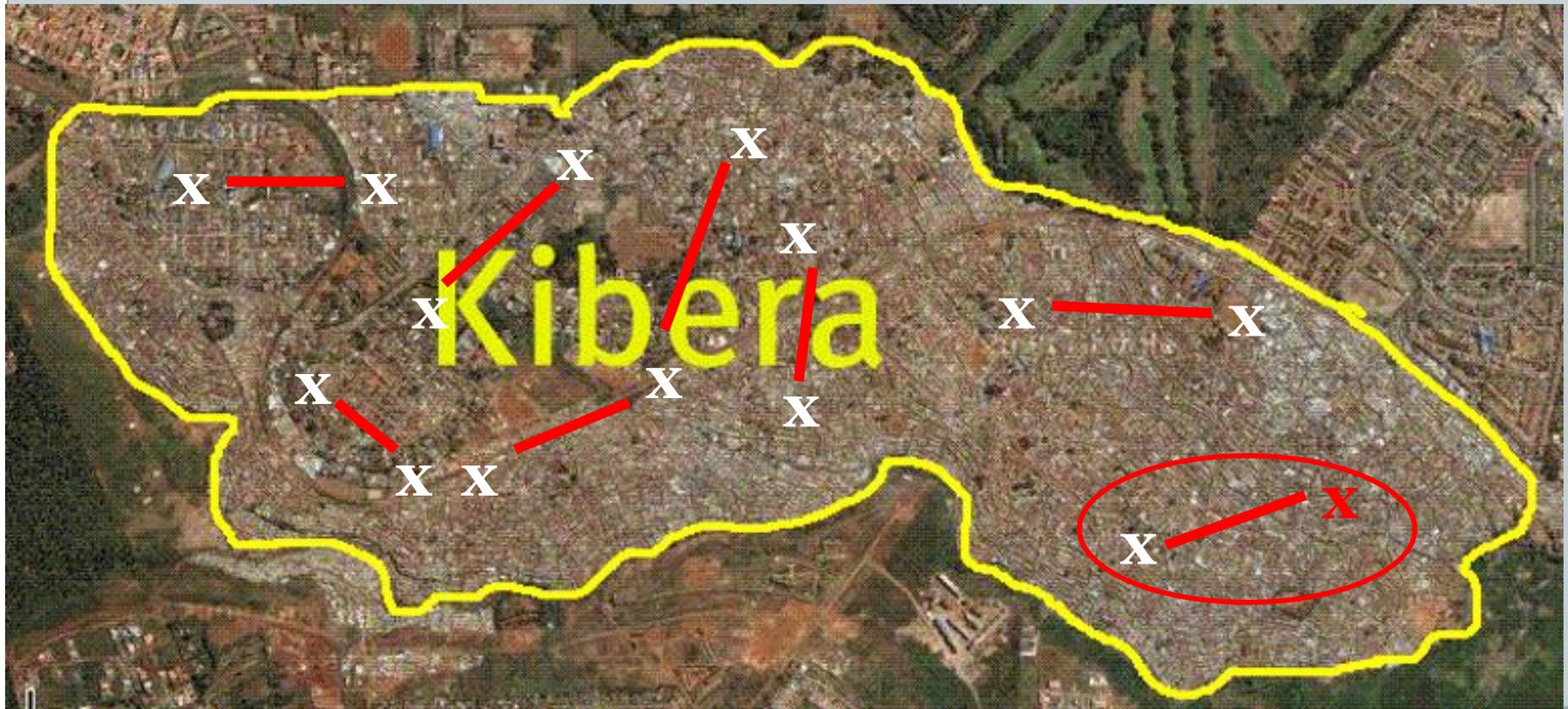


- Done on the cheap, 72 hours from baseline data collection to randomization.
- Constrained baseline covariate information at the cluster level: area, public/private, materials, test scores, number of girls, number of boys.

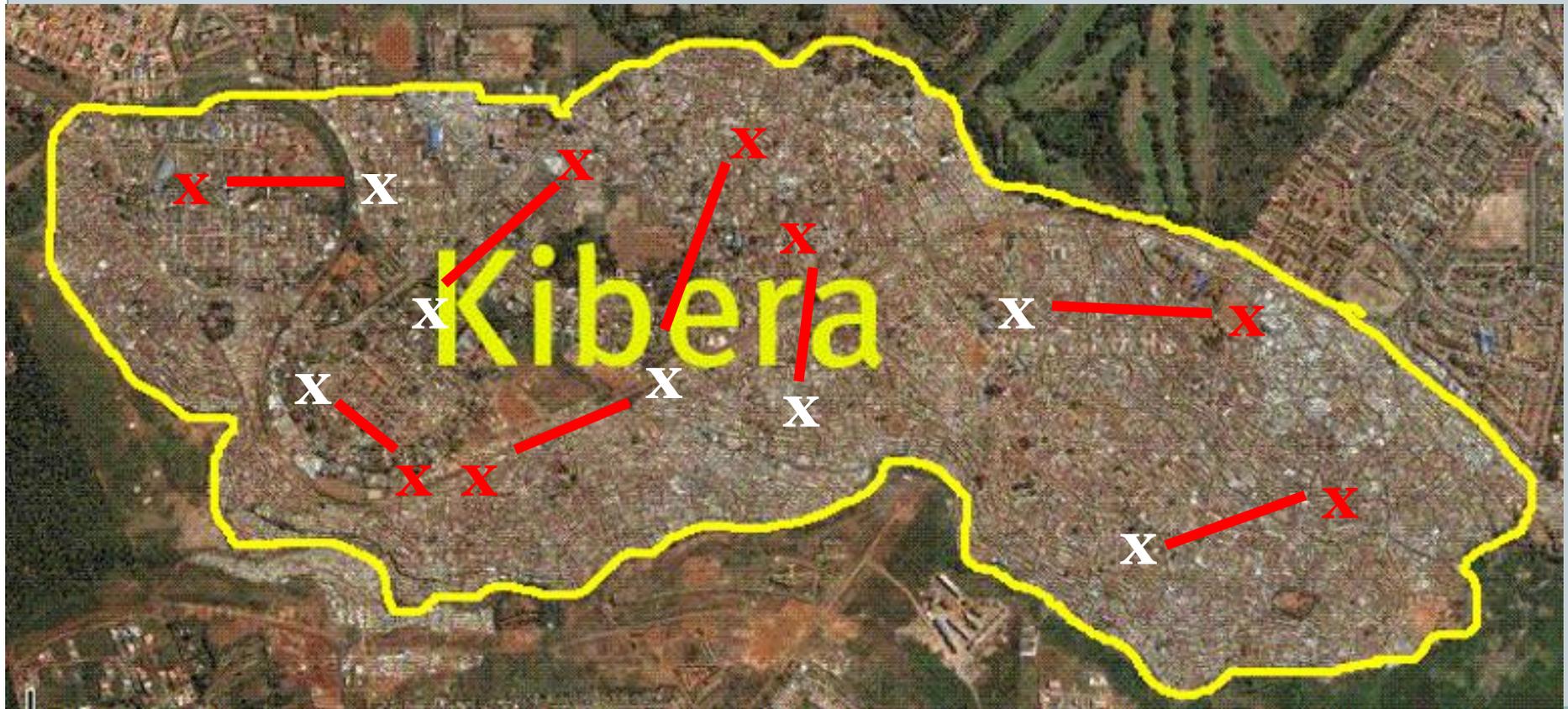
constrained randomization



constrained randomization



constrained randomization



example



- Because of the randomization, we could wait until the end of the 12 month period and compare the rate of rape in the intervention schools vs the rate in the delayed intervention schools.
- There are several issues we're concerned about:
 - The girls are aging into a risky period in their lives. This means you do not want to compare baseline rate in the treated vs endpoint rate in the treated.
 - The rates of rape from school to school vary a lot and it is difficult to "control" this during our randomization. These issues arise because of (i) school homophily, (ii) the delicate nature of our outcome means we can't ask the questions directly before building up a relationship with the girls and (iii) the level of the science isn't at the point that we have great predictors of rates of rape.

example



- We go into the schools, build up a relationship over the course of a couple visits and then measure the baseline rate of rape in the prior 12 months.
- Our quantity of interest becomes:

Let $p_{i,t}$ be the probability of a randomly sampled girl, from school i , having been raped in the prior 12 months at time $t=\{0,1\}$, where $t=0$ is at baseline and $t=1$ is at the endpoint.

$$(p_{i,1} - p_{i,0}) - (p_{j,1} - p_{j,0})$$

diff-in-diff: inference



- Inference: When someone invoke a “diff-in-diff design” they haven’t necessarily identified their source of randomization. You still need to chase this out.
- In the Kenya example I randomized the schools. So I’ll use the randomization as the basis for a permutation test that is a diff-in-diff estimator.
 - I’m doing this to improve precision.

diff-in-diff: inference



- Briefly consider the example of [Card and Krueger \(1994\)](#):
New Jersey minimum wage increased, but Pennsylvania's did not. Looked at fast food restaurants near the border of the two states. Wanted the impact on employment.
- Belief is that the restaurants were the same before the change, and would have continued but for the minimum wage change.
- Design consideration: try to look at “burn in” period prior to the intervention. The more you can get the two groups similar prior, the easier your case will be.
 - Researchers end up using matching a lot in diff-in-diff designs.

diff-in-diff: inference



- In a structural equation model approach:

$$y_{i,t} = \beta_0 + \beta_t * t_i + \beta_d * d_i + \beta_{t*d} t_i * d_i + \varepsilon_{i,t}$$

where t is time (pre=0, post=1) and d is the intervention (control=0, intervention=1) and $t_i * d_i$ is an interaction term. The quantity of interest is β_{t*d} .

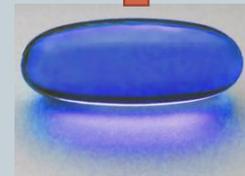
The only difference

$$r_T = f(d = 1, X = \mathbf{x})$$

$$r_C = f(d = 0, X = \mathbf{x})$$



$\mathbf{x} =$

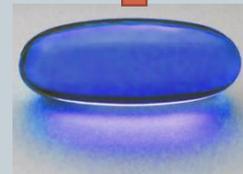


$\mathbf{x} =$

$$r_T = f(d = 1, X = x)$$



$$r_C = f(d = 0, X = x')$$



$x =$

$x' =$

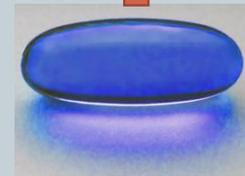
$(\text{contrast 1}) - (\text{contrast 2}) = \text{difference-in-differences}$



contrast 1



contrast 2



diff-in-diff



- Takeaway: difference-in-difference approaches try to reduce the variation in your estimate due to unit-to-unit differences. It does this by focusing on differencing a pre-treatment and post-post measurement.
- Keep in mind: diff-in-diff is an approach that can often be paired with other aspects of design (e.g., pscore matching, IVs, RCTs).
- If you have the data available then doing a diff-in-diff is usually a good idea. (Personally, I've never seen a situation where the diff-in-diff was worse.)

takeaways



takeaways



- RCTs provide high quality information
- Part of that is because of the transparency and clarity of their argumentation.
- We cannot achieve RCT-level in an observational study
- Fortunately, science is not all or nothing. It is an accumulation of facts and insights until we reach a preponderance of evidence and decide to move on.
- Observational studies have a role in this process.
- Observational studies can be better or worse.

takeaways



- Observational studies can be better or worse.
- Worse: no discussion of inclusion/exclusion, just a model
- Better: identification of randomness, detailed discussion of baseline characteristics.
- Excellent: Sensitivity analysis, coherence, multiple controls
- Reporting standards:
 - [The CONSORT Statement](#)
 - [The STROBE Statement](#)

next class



observational studies



- The world of “observational studies” is kind of hard to get into because it grew up in several distinct, but overlapping, disciplines:
 - Epidemiology
 - Demography
 - Economics (econometrics)
 - Political Science
 - Sociology
 - Biostatistics
 - Statistics
 - Psychology (psychometrics)
 - Computer Science

fin.

