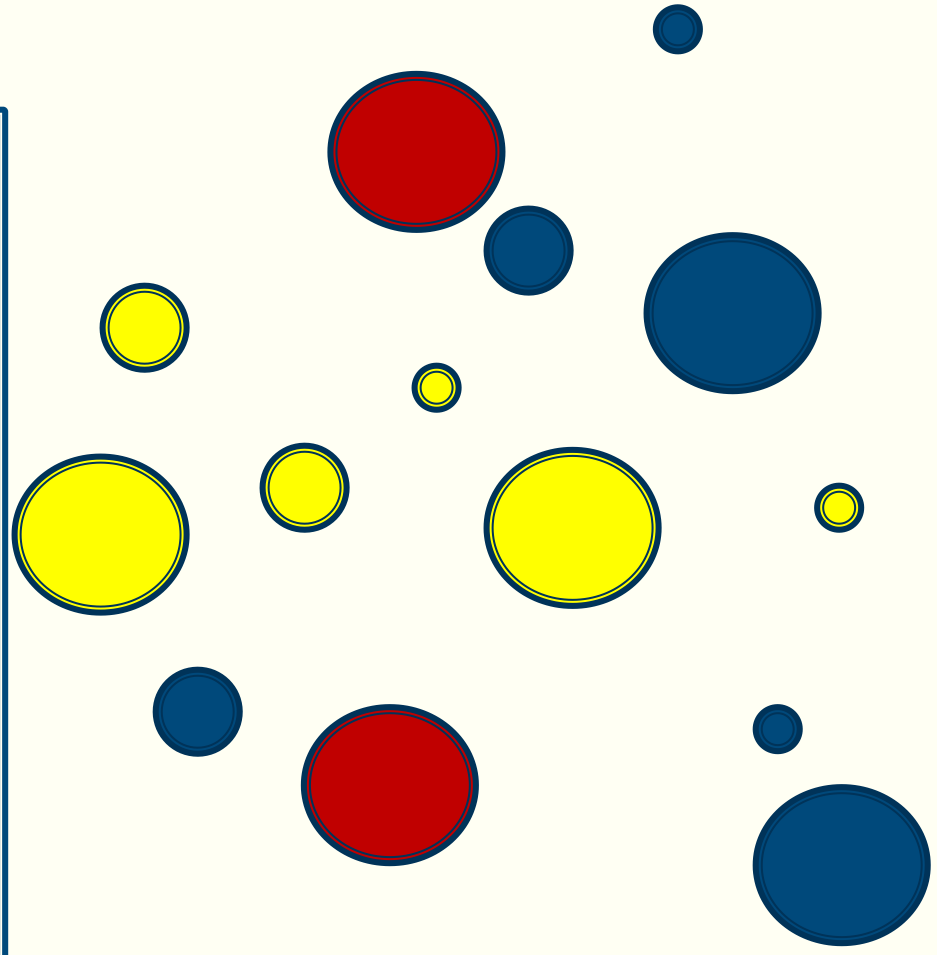
NATURAL EXPERIMENTS AND OBSERVATIONAL STUDIES

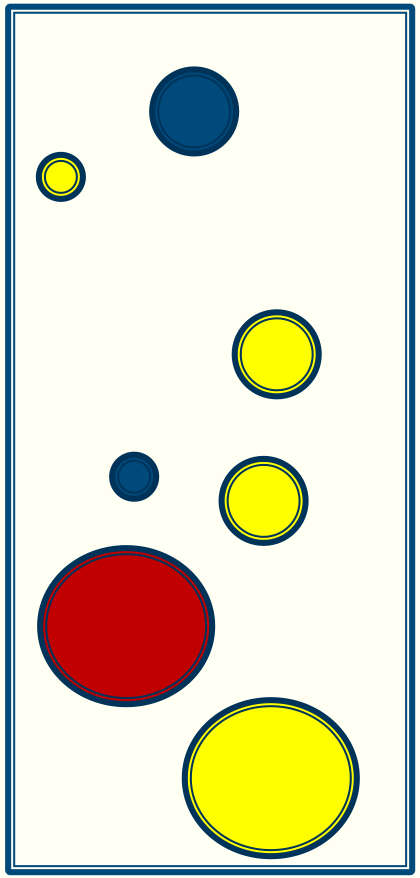
Professor Michael Baiocchi
Department of Epidemiology
Stanford University

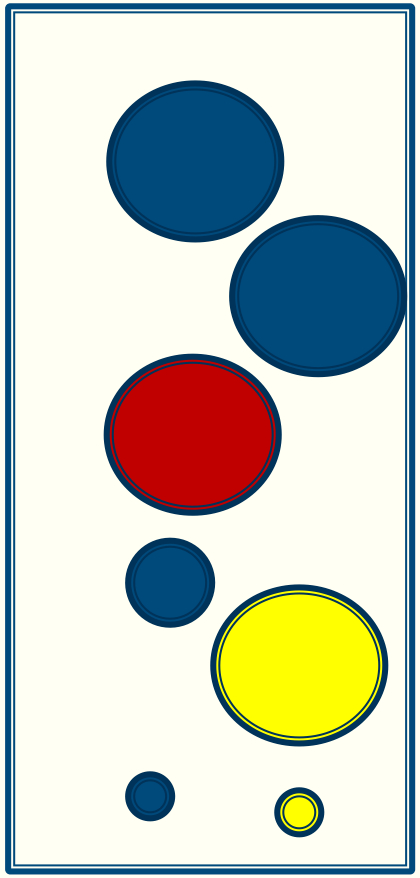
what makes randomized controlled trials special?

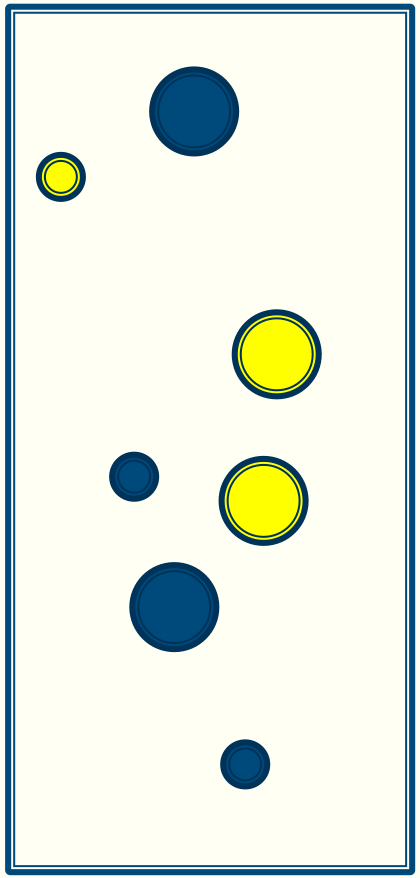
- Randomized controlled trials (RCTs) are commonly viewed as the best way for generating data useful for estimating causal effects.
- We are going to focus on two aspects:
  - Randomization
  - Control

Rosenbaum, P. R. (2005). **Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies**. The American Statistician, 59(2), 147-152.



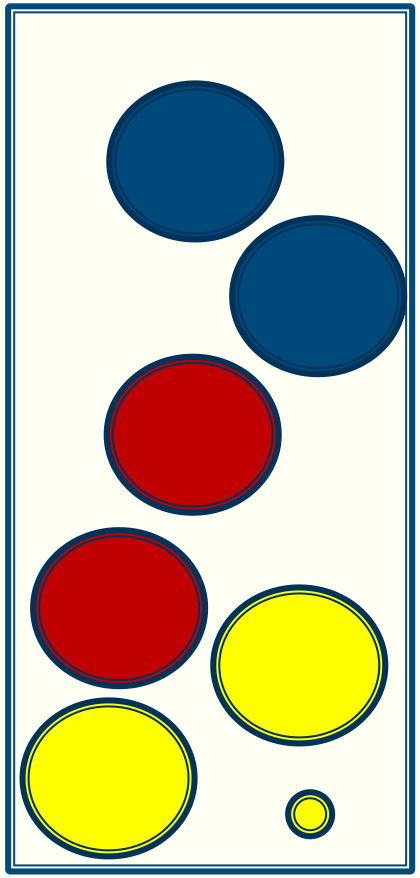


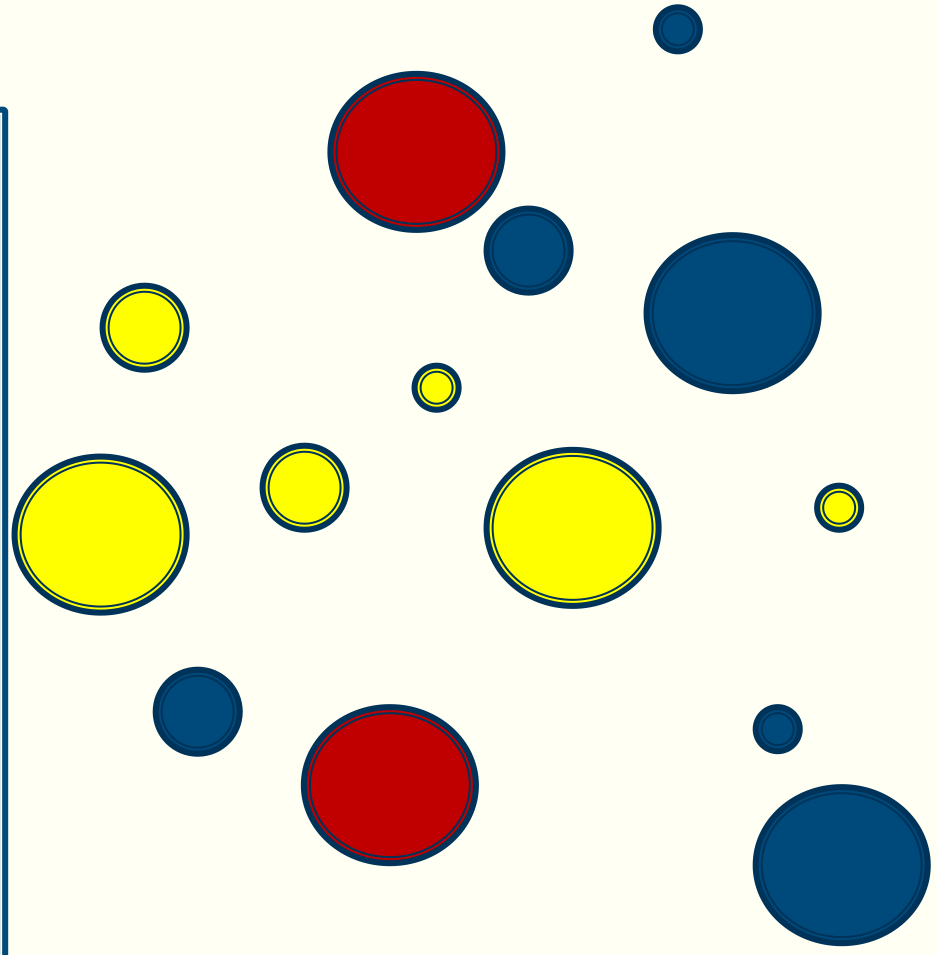


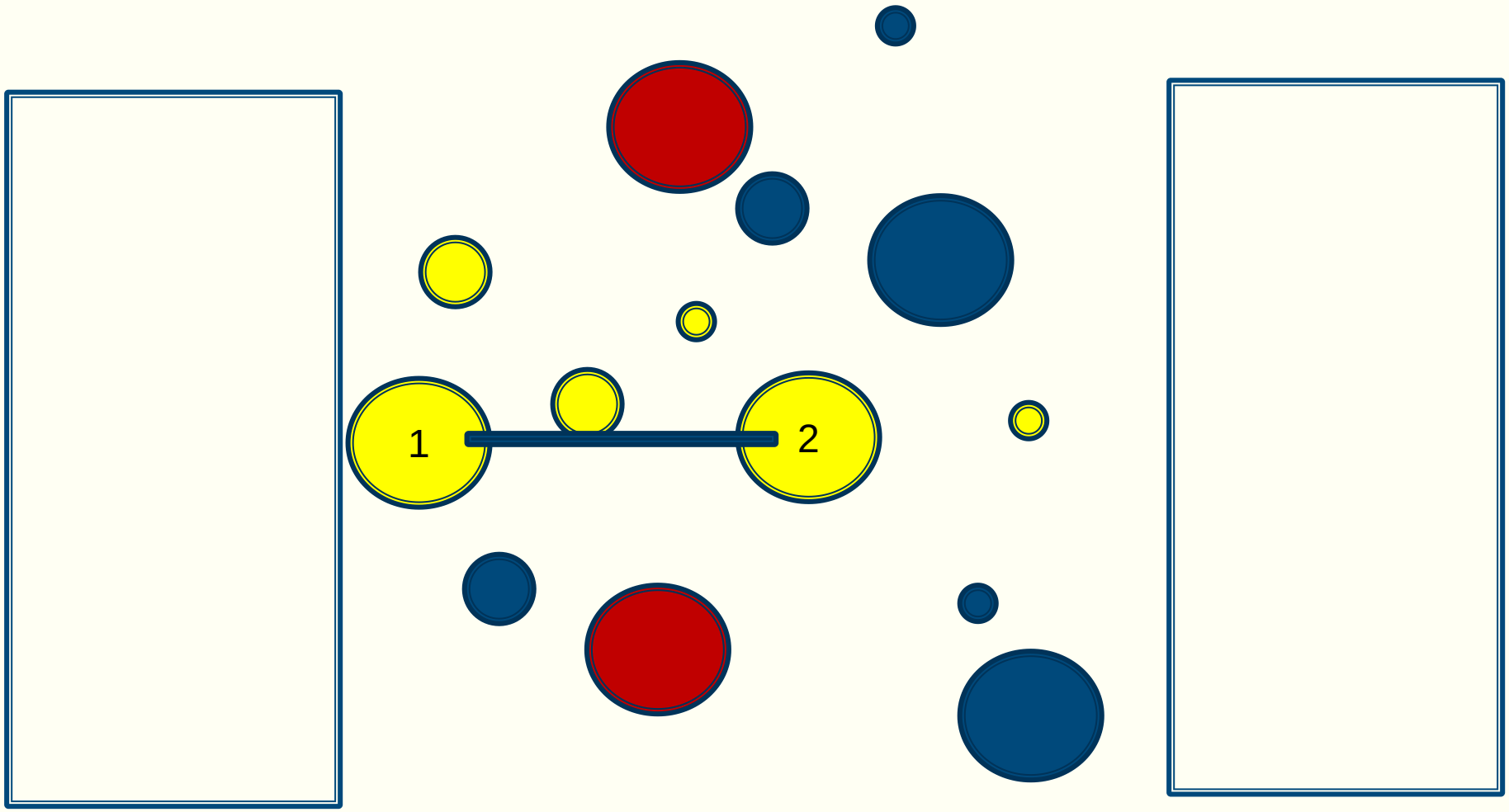


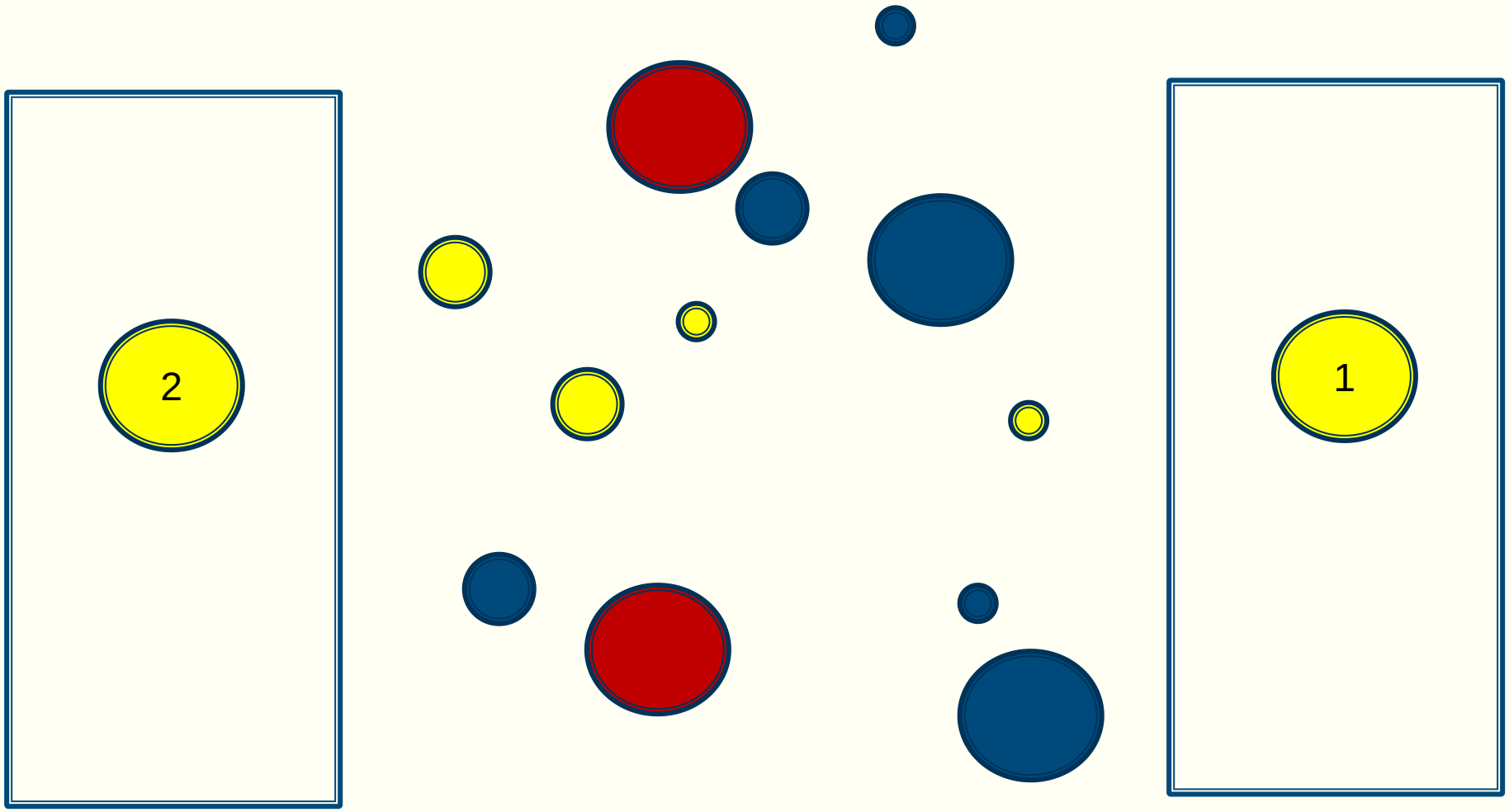
We’re off to a **bad** start.

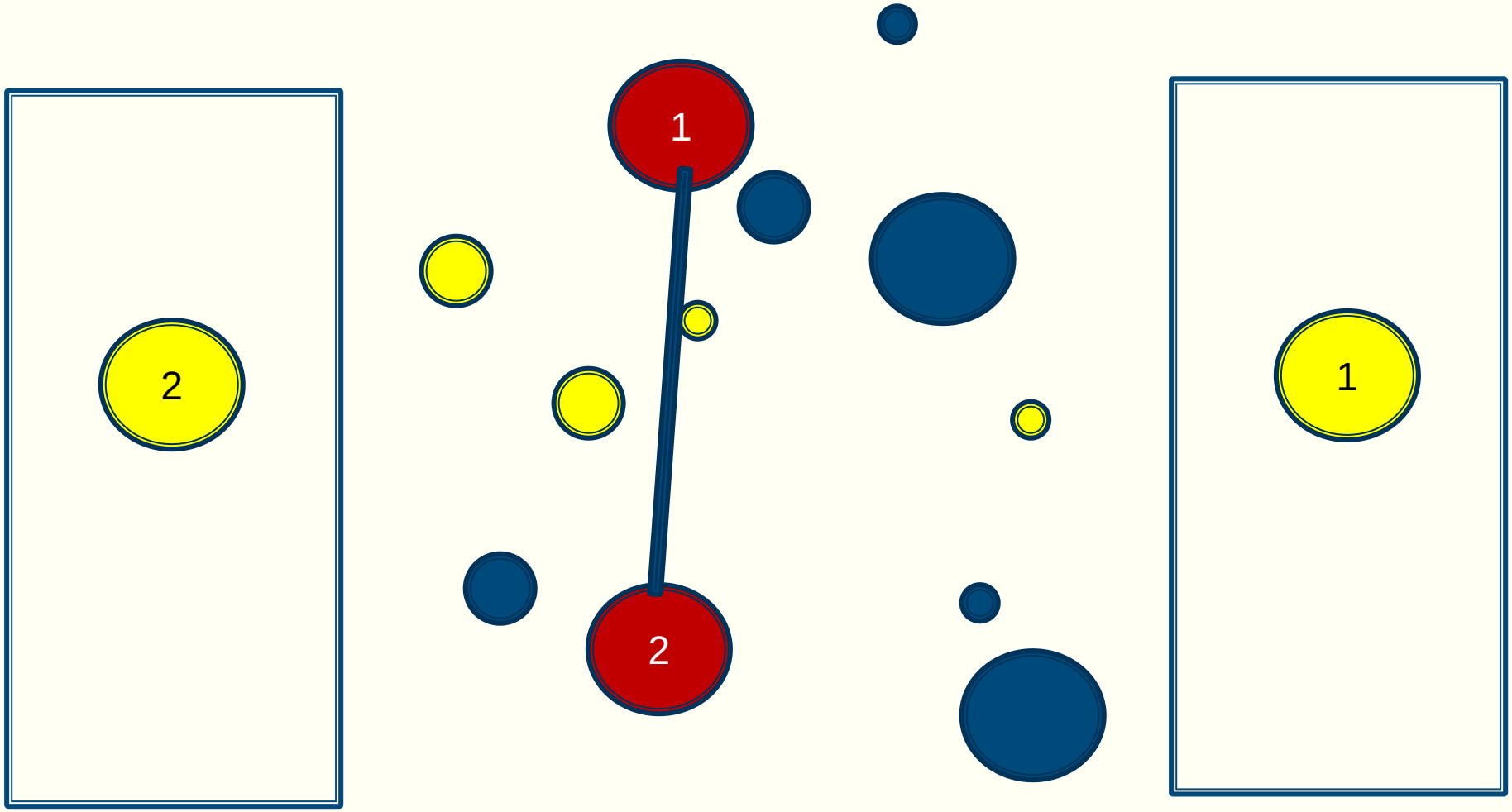
These groups are imbalanced.

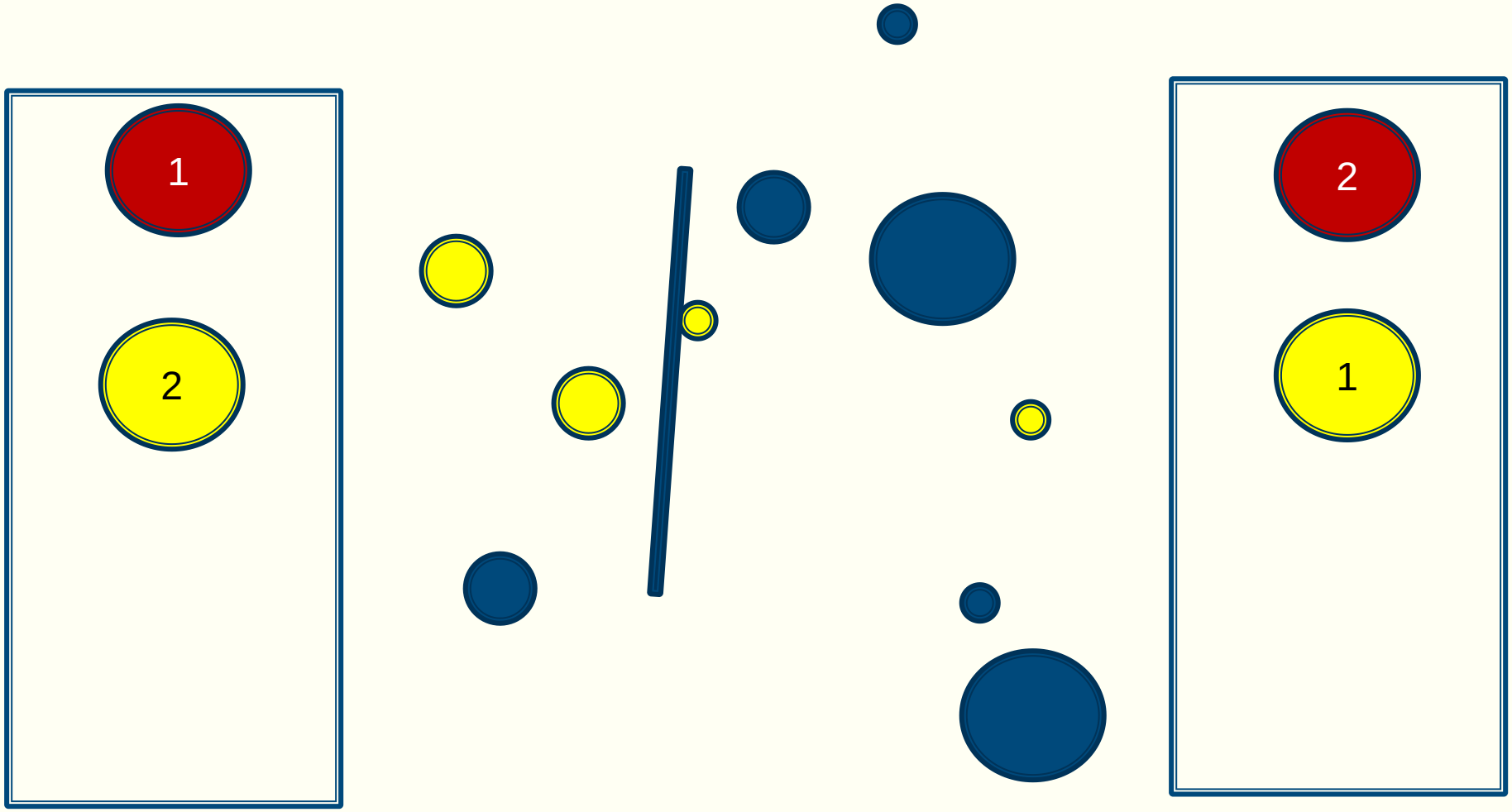


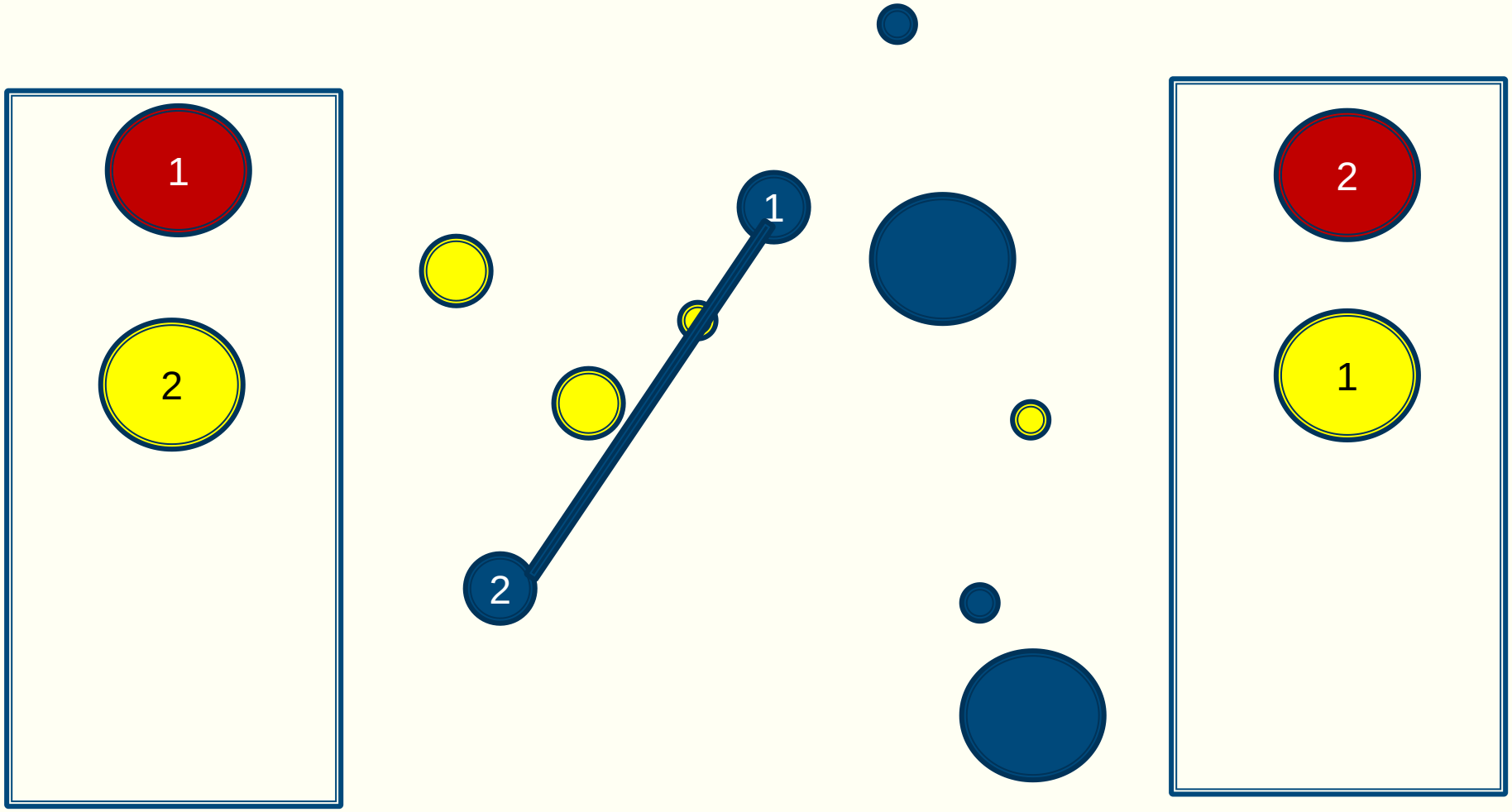


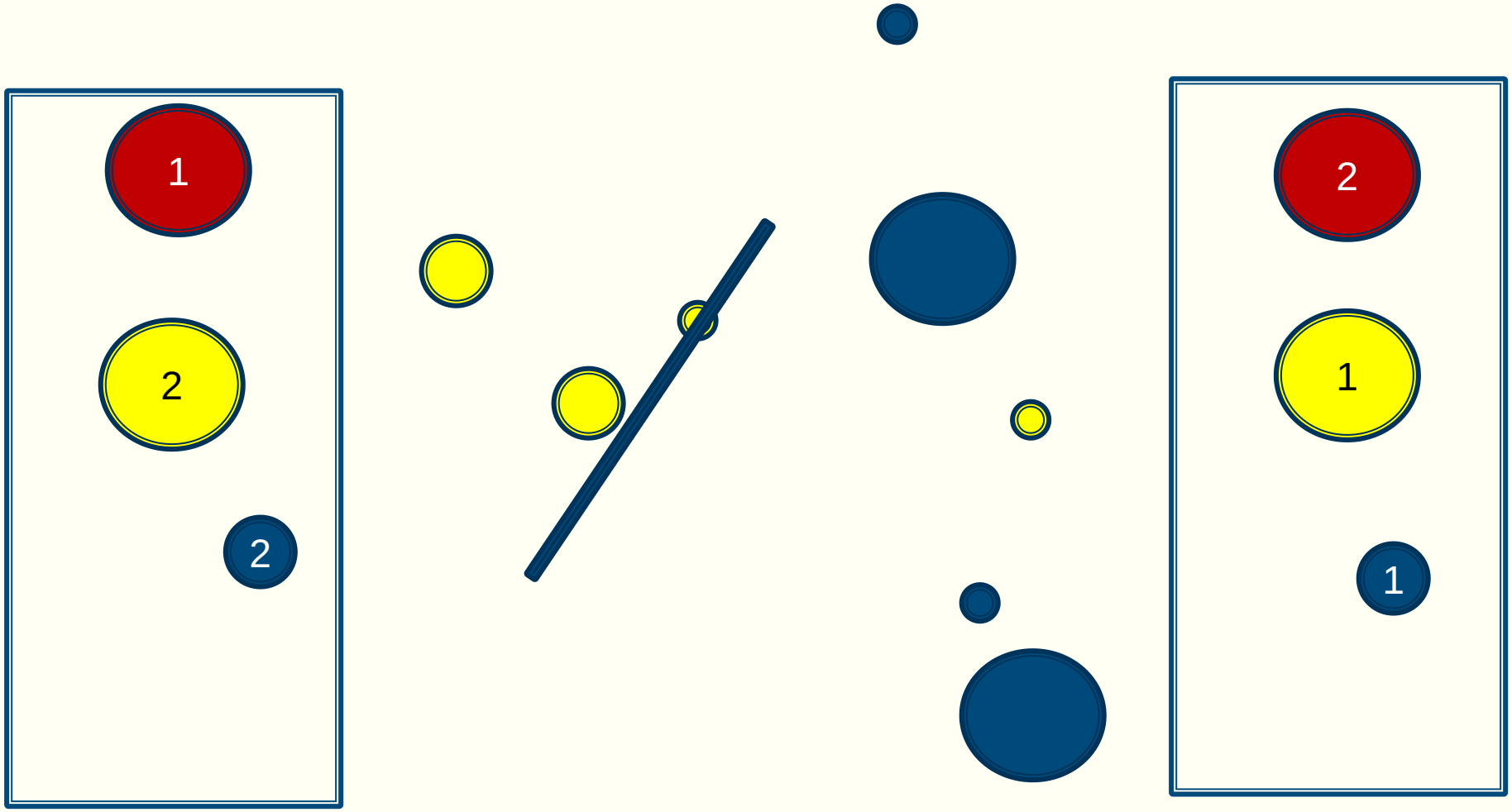

1
2

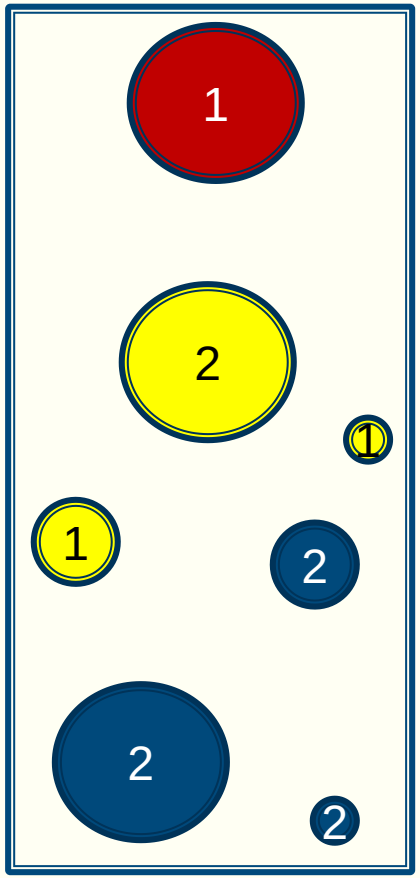

2
1

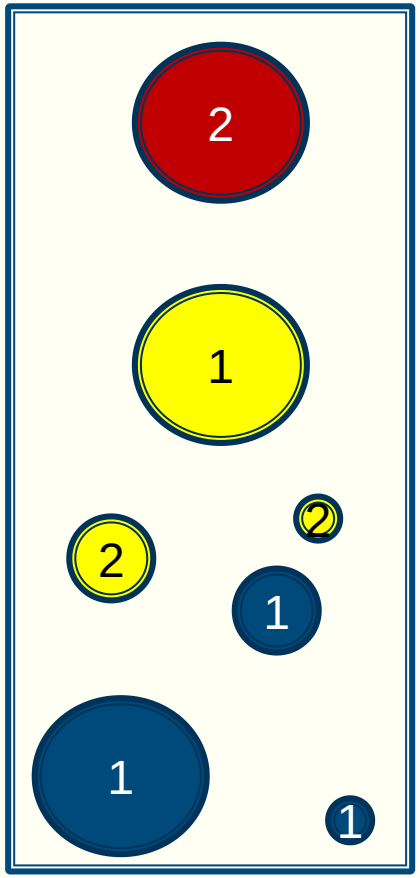

1
2
2
1

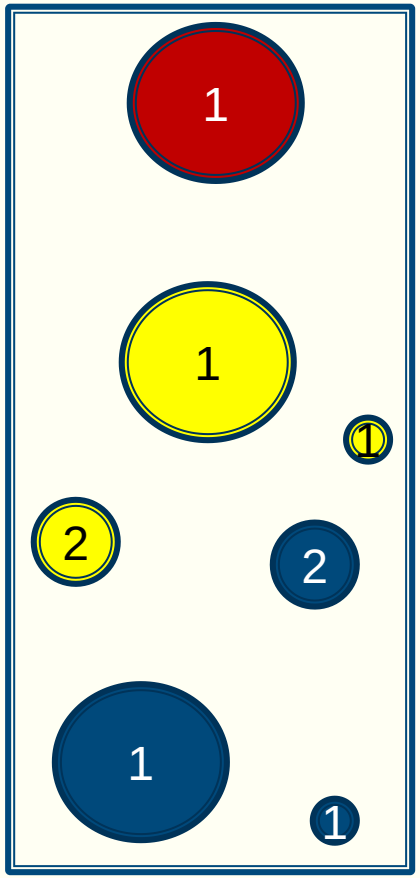

1
2
2
1

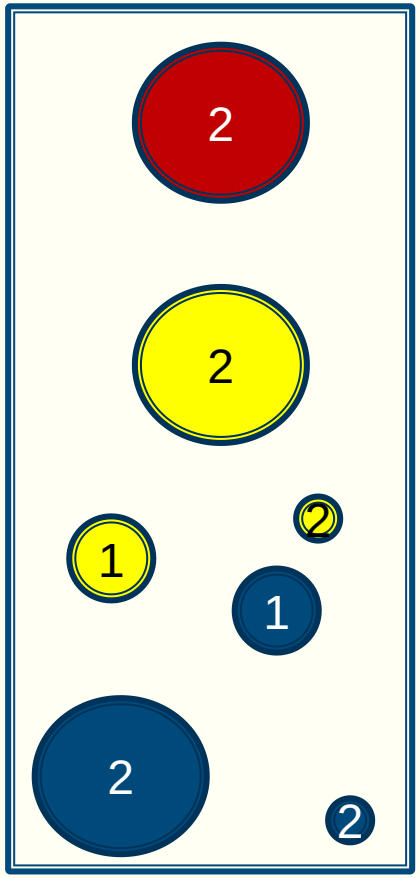

1
2
1
2
2
1

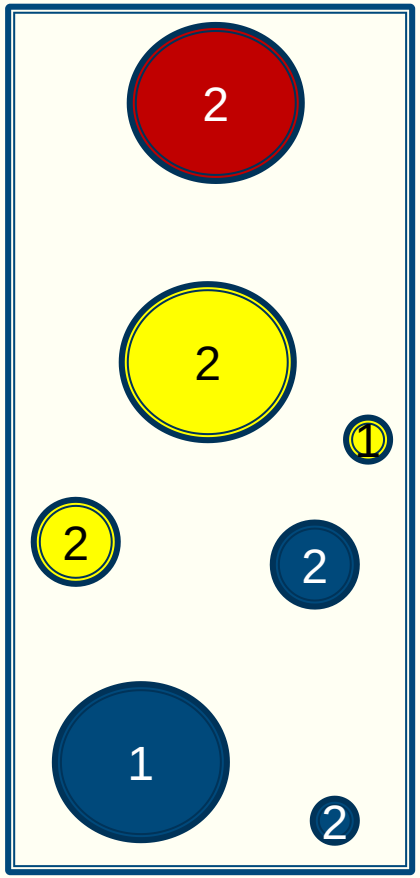

1
2
2
2
1
1

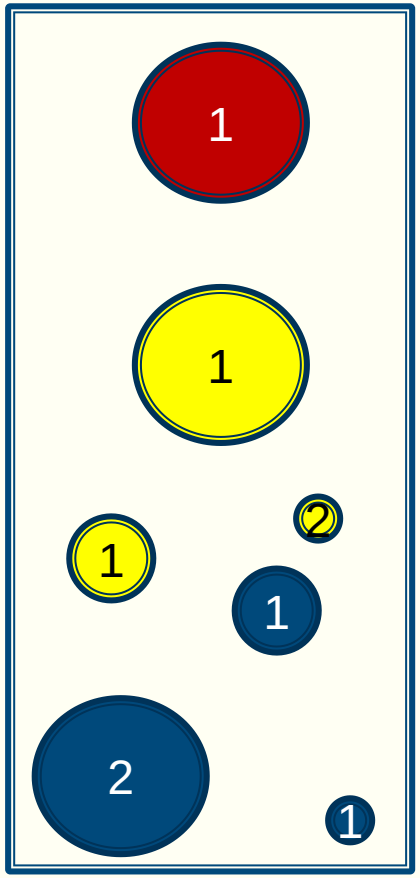

1
2
1
1
2
2
2

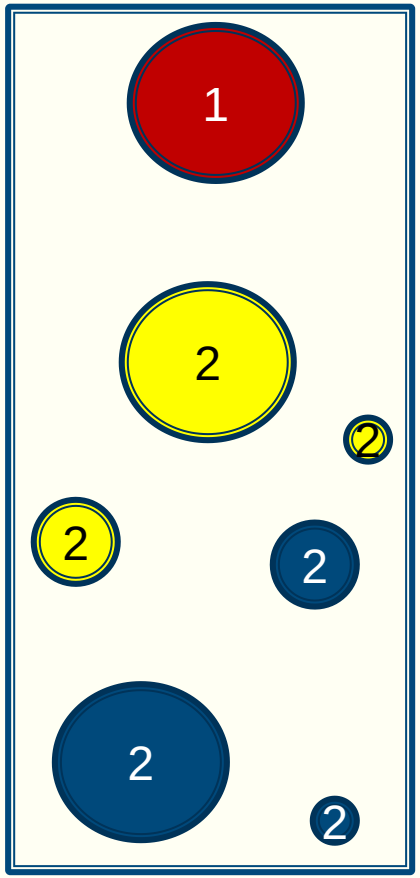

2
1
2
2
1
1
1

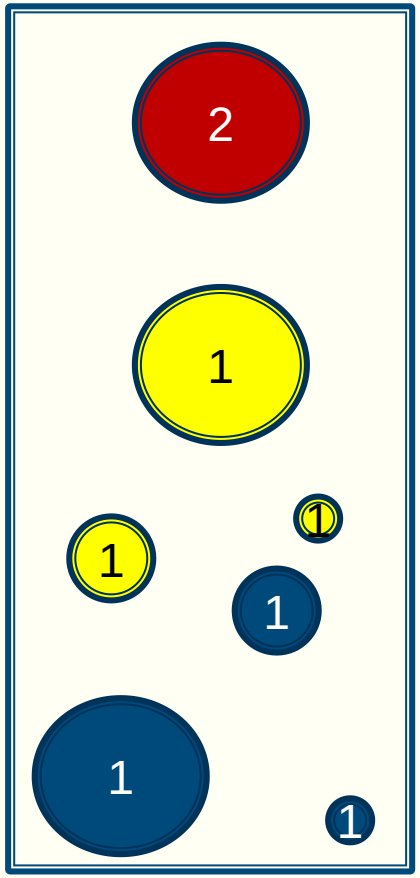

1
1
1
2
2
1
1

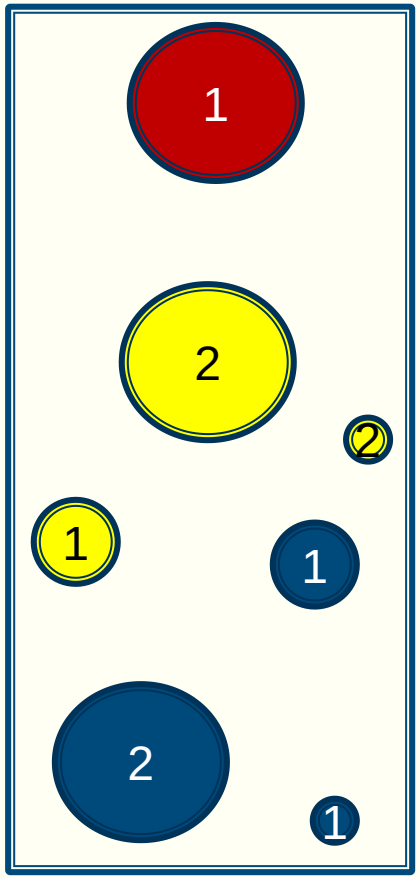

2
2
2
1
1
2
2

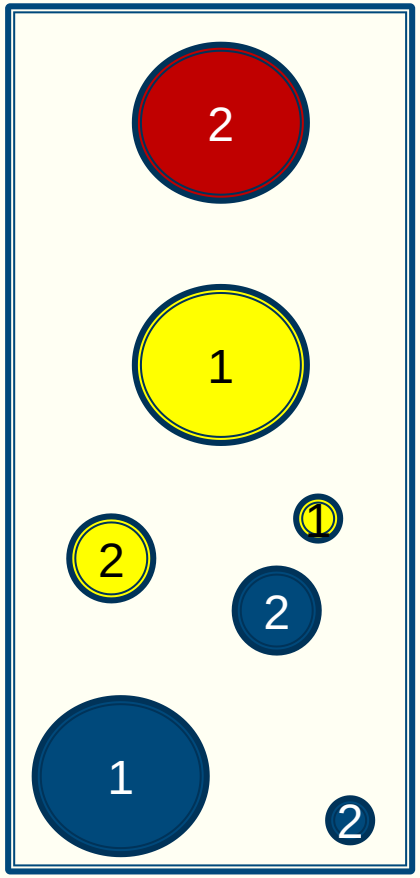

2
2
1
2
2
1
2


1
1
2
1
1
2
1


1
2
2
2
2
2
2


2
1
1
1
1
1
1


1
2
2
1
1
2
1


2
1
1
2
2
1
2

what makes randomized controlled trials special?

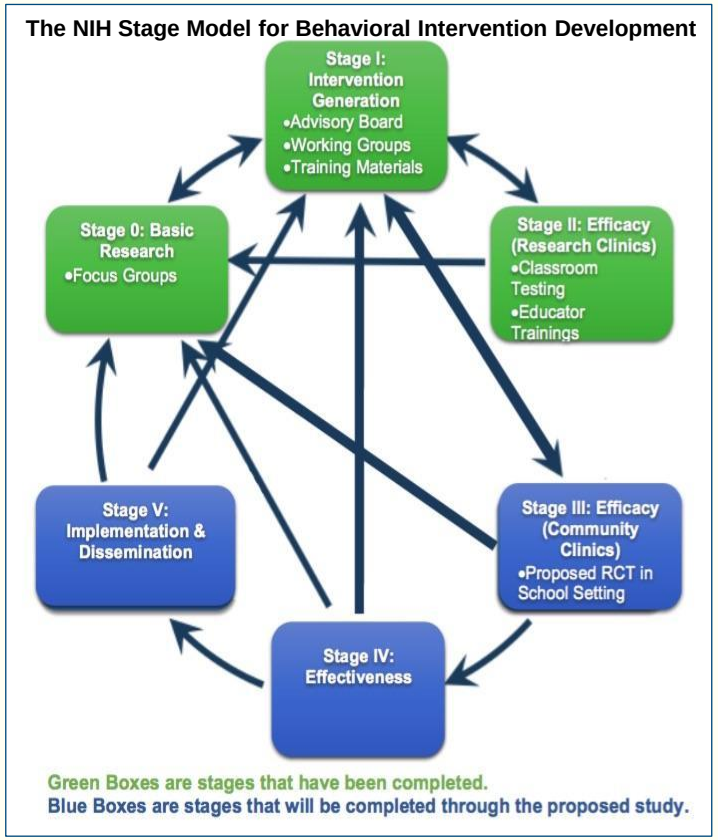
- We can exercise control over the variables we observe, to anticipate and mitigate other explanations of the variation in the outcome.
- The randomization then helps because we have:
  - a non-structured assignment mechanism (i.e., no patterns in *general* but perhaps by chance),
  - a mathematical way of describing what else could have happened (null distributions, standard errors, and p-values).
- It's this synthesis of controlled randomness that gets us powerful results.

Rosenbaum, P. R. (2005). **Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies**. The American Statistician, 59(2), 147-152.

NATURAL EXPERIMENTS

why do natural experiments?

- If RCTs are our best way of generating data for causal inferences then why not always do RCTs?
  - Ethics (smoking studies)
  - Cost (RAND Health Insurance study)
  - Speed of innovation (Valve Replacement study)
  - We may not be at the right point in scientific discovery (e.g., ORBIT model, NIH Stage Model for Behavioral Intervention Development)
  - RCTs may not have great generalizability



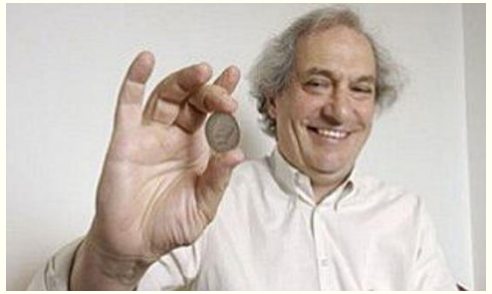
[COST] Brook, R. H., Ware Jr, J. E., Rogers, W. H., Keeler, E. B., Davies, A. R., Donald, C. A., ... & Newhouse, J. P. (1983). Does free care improve adults' health? Results from a randomized controlled trial. New England Journal of Medicine, 309(23), 1426-1434.

[SPEED OF INNOVATION] Goldstone, A. B., Chiu, P., Baiocchi, M., Lingala, B., Patrick, W. L., Fischbein, M. P., & Woo, Y. J. (2017). Mechanical or biologic prostheses for aortic-valve and mitral-valve replacement. New England Journal of Medicine, 377(19), 1847-1857.

[STAGE OF SCIENCE] Czajkowski, S. M., Powell, L. H., Adler, N., Naar-King, S., Reynolds, K. D., Hunter, C. M., ... & Epel, E. (2015). From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychology, 34(10), 971.

how natural experiments?

- When designing an observational study, think of the RCT you would ideally run to answer your question.
  - How will you exercise control over the data?
  - Consider looking at the CONSORT guidelines.
- It turns out that we do not actually need something truly random.
  - Percy Diaconis



- Now think about how to isolate the natural experiment – i.e., find some “randomness.”

Zubizarreta, J. R., Small, D. S., & Rosenbaum, P. R. (2014). **Isolation in the construction of natural experiments.** The Annals of Applied Statistics, 2096-2121.

Hernán, M. A., & Robins, J. M. (2016). **Using big data to emulate a target trial when a randomized trial is not available.** American journal of epidemiology, 183(8), 758-764.

TWO FLAVORS

two flavors of “natural” experiments

- Discontinuity design
  - There is some variable (usually called a “running variable”) that assigns people to T or C, often in quite a stark way.
  - We can use the rapid change in treatment probabilities as a type of randomizer.
  - The randomness is such slightly off-stage. We have a strong story for *why* (the running variable) but we don’t get to see exactly how someone got assigned.

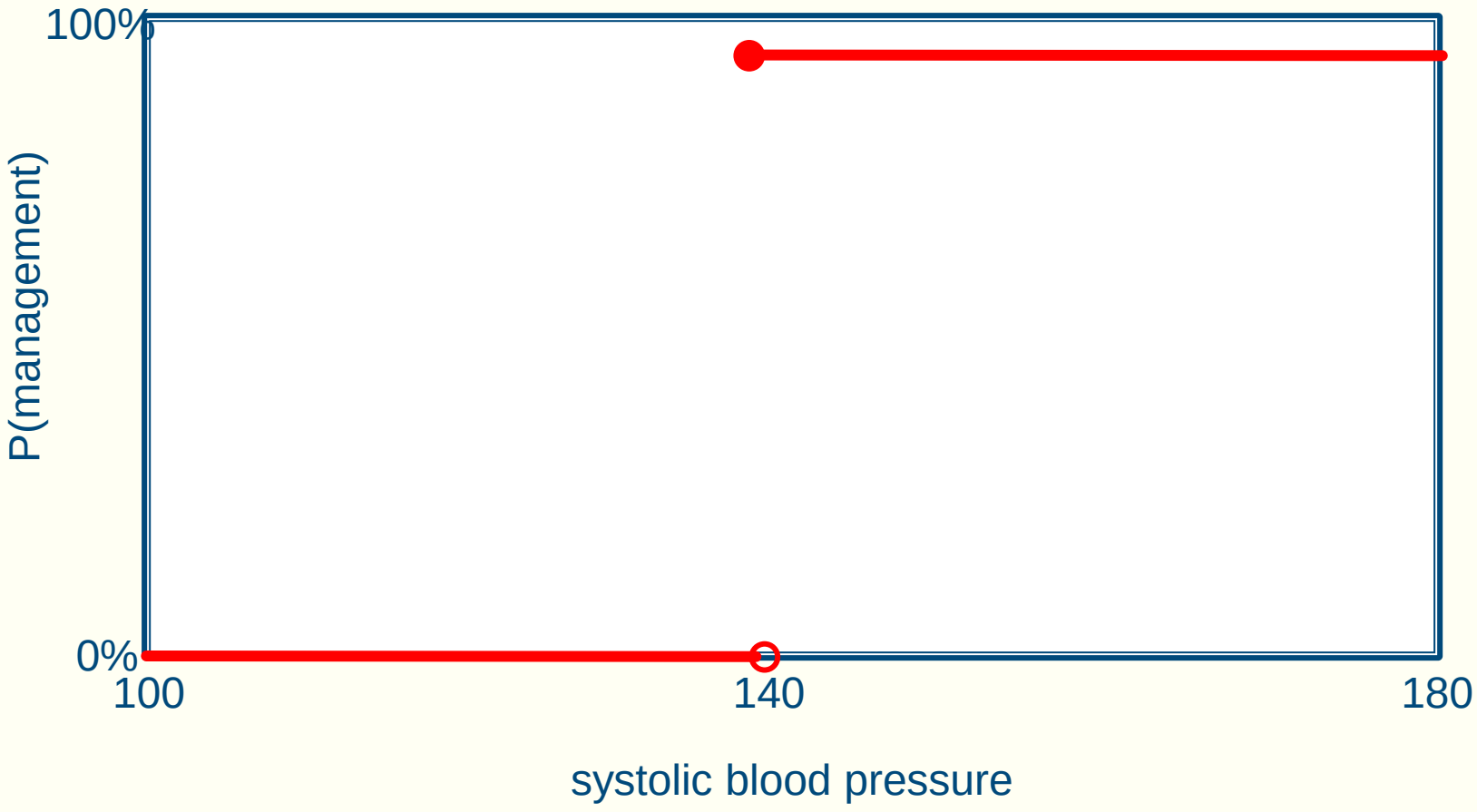
- Instrumental Variables
  - This is as close to a pseudo-randomizer as we have. It is a variable that assigns T *vs* C but is not connected to the outcome, except through the treatment.
  - In this design, we have a variable – the IV itself – that quantifies the randomness.

two flavors of “natural” experiments: discontinuity design

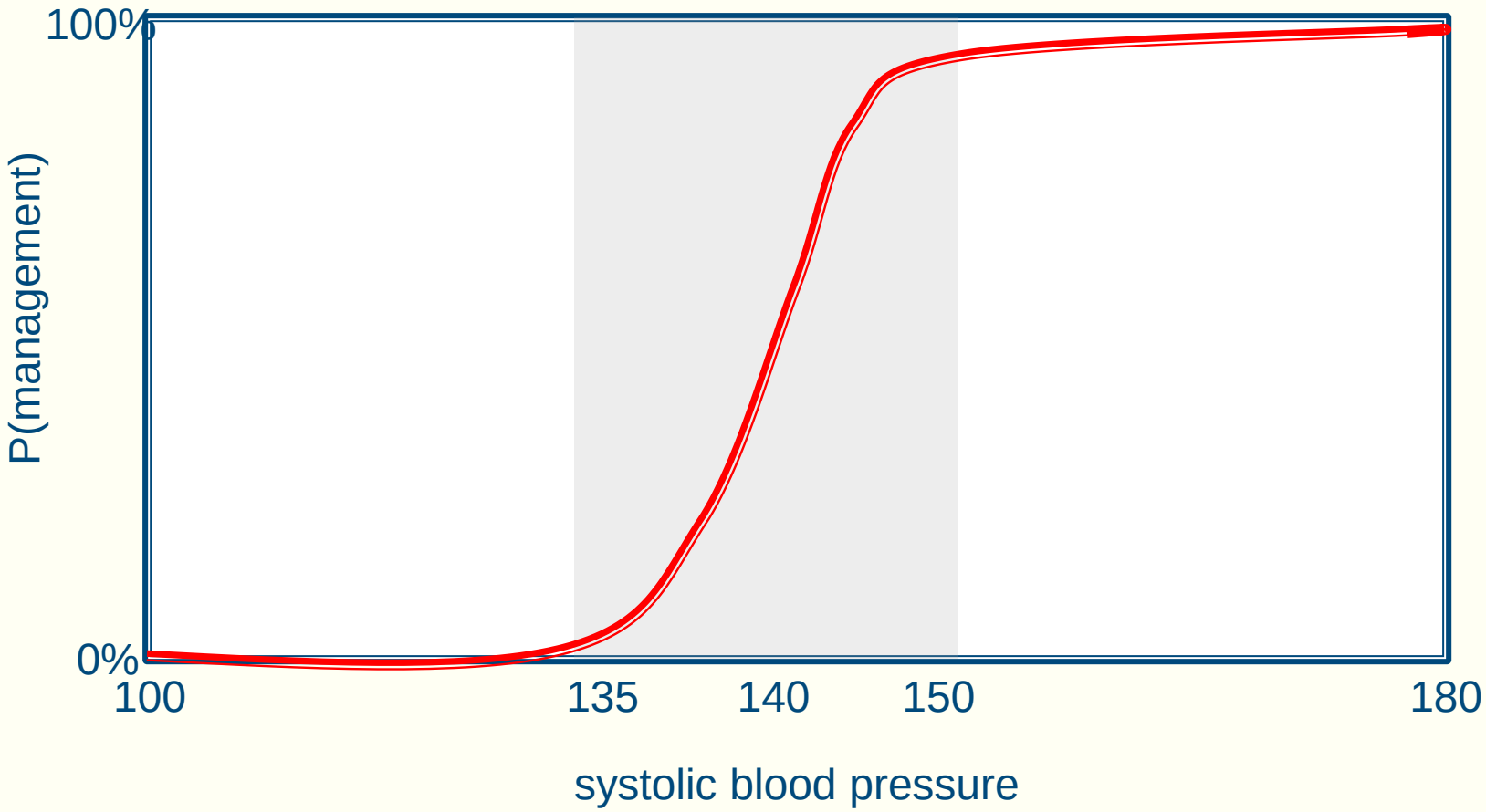
- Discontinuity design
  - Example study: what benefit is there to aggressively managing blood pressure?

Imbens, G. W., & Lemieux, T. (2008). **Regression discontinuity designs: A guide to practice**. Journal of econometrics, 142(2), 615-635.

discontinuity design: blood pressure



discontinuity design: blood pressure



two flavors of “natural” experiments: discontinuity design

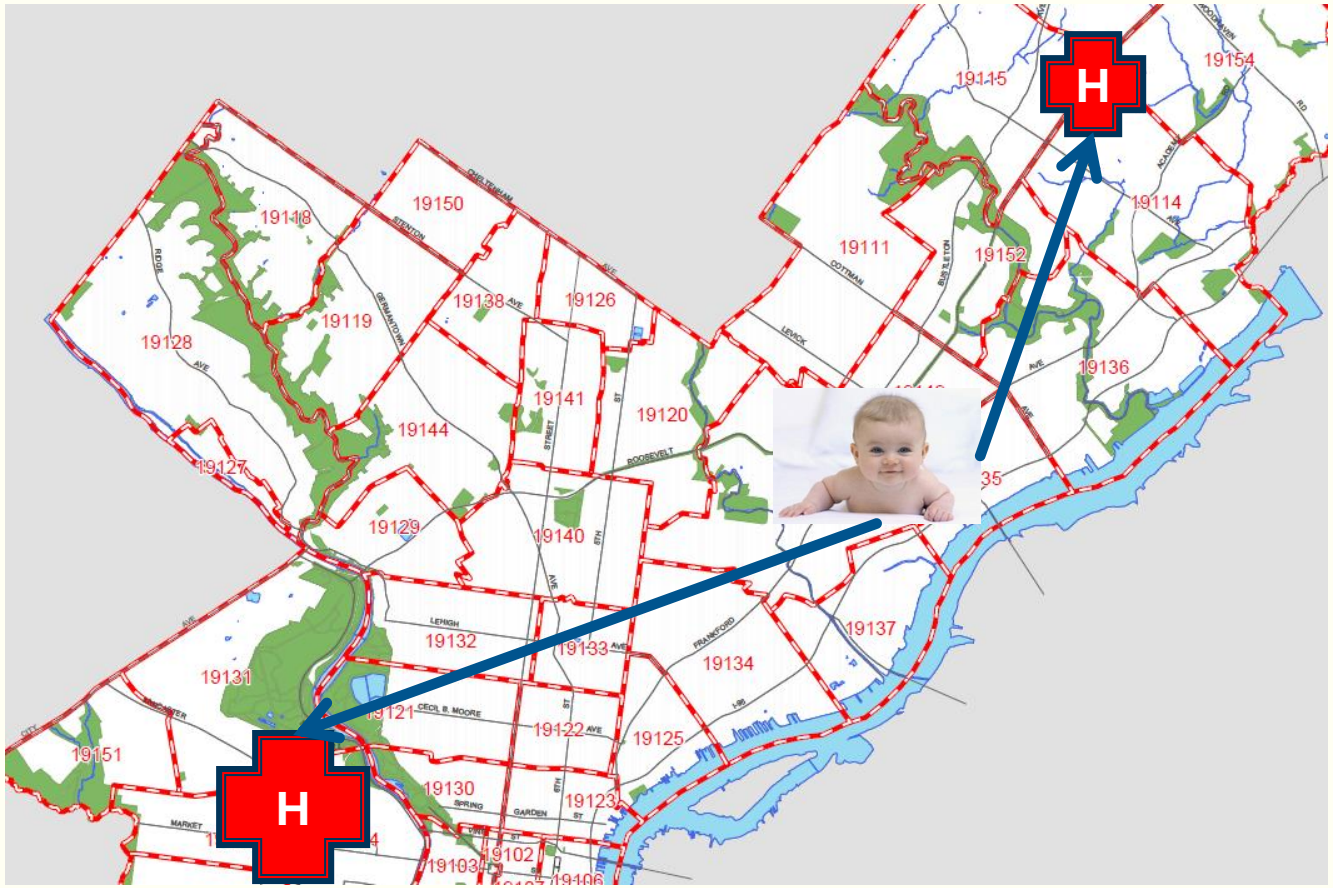
- Discontinuity design
  - Example study: what benefit is there to aggressively managing blood pressure?
- We now have a detailed explanation of assignment into T vs C, which is that it was because someone was above/below the cutoff point. But we should still be a little bit bothered by not knowing why someone ended up above/below.
- Examples of discontinuities: time (e.g., competitor enters the market, black box warning from the FDA), political boundary (e.g., insurance access in two neighboring cities), and expert guidelines (e.g., initiate treatment at a given level).

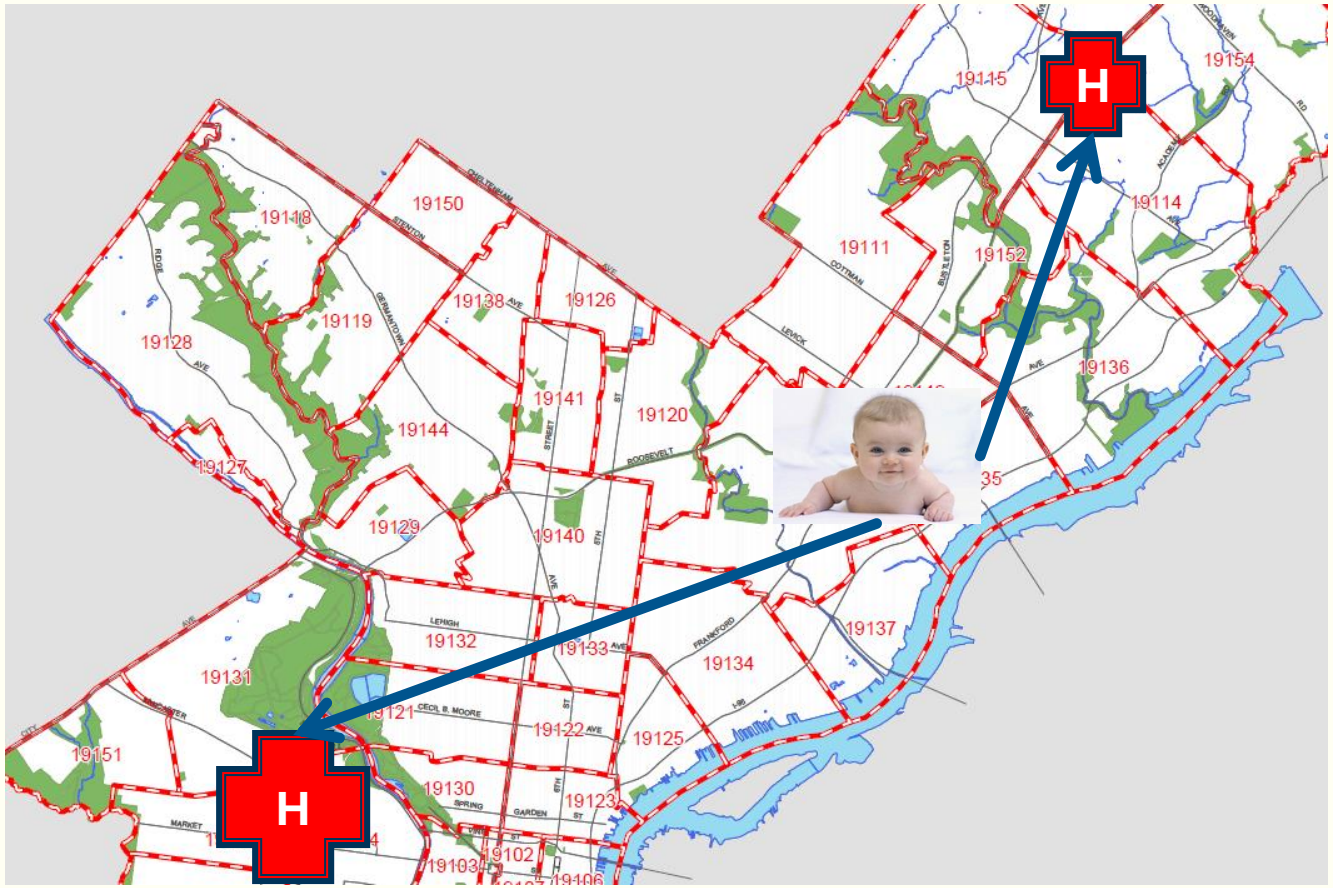
Imbens, G. W., & Lemieux, T. (2008). **Regression discontinuity designs: A guide to practice**. Journal of econometrics, 142(2), 615-635.

two flavors of “natural” experiments: instrumental variable

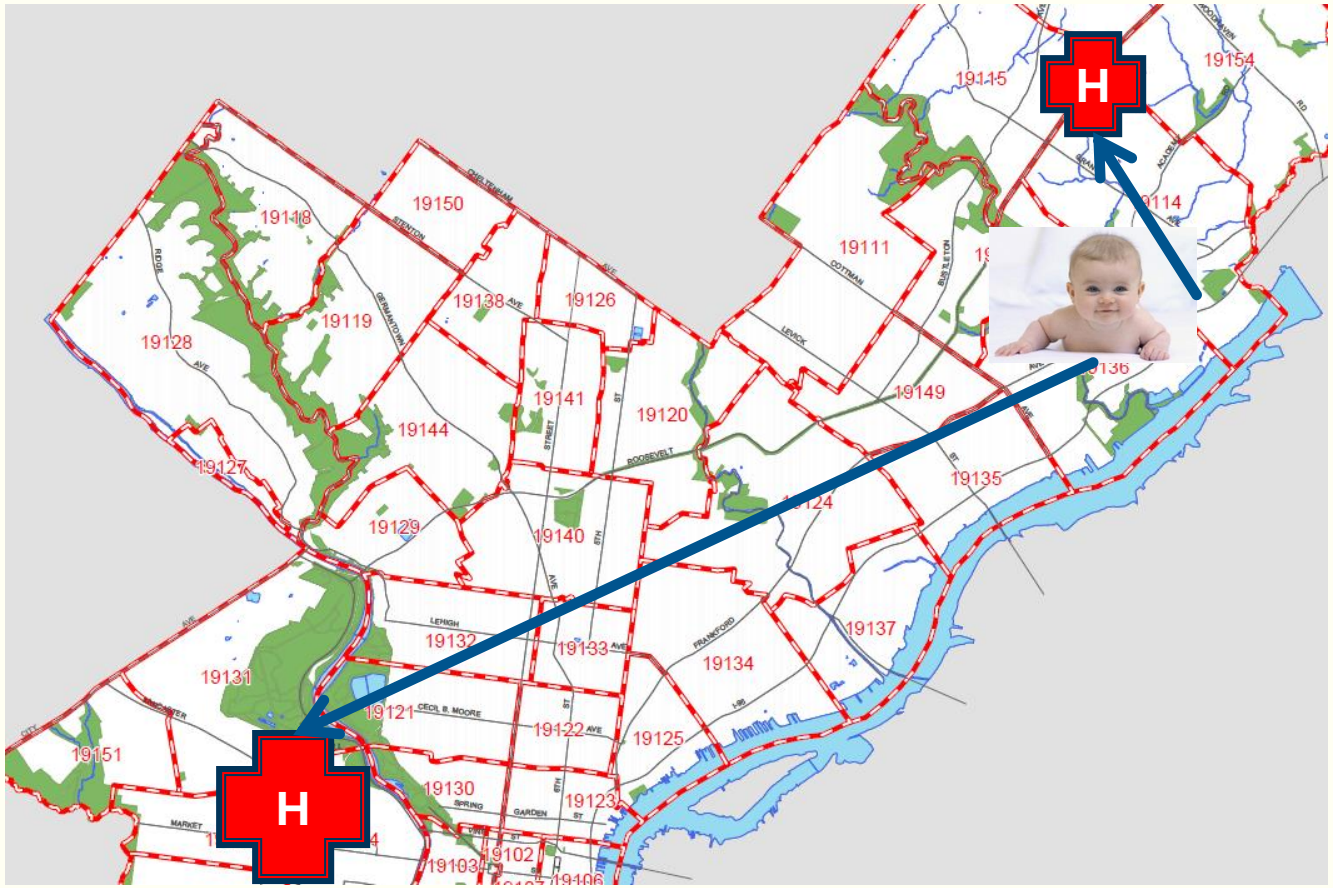
- Instrumental variable design
  - Example study: would regionalizing premature babies to higher-volume facilities result in better outcomes?

Baiocchi, M., Cheng, J., & Small, D. S. (2014). **Instrumental variable methods for causal inference.** Statistics in medicine, 33(13), 2297-2340.

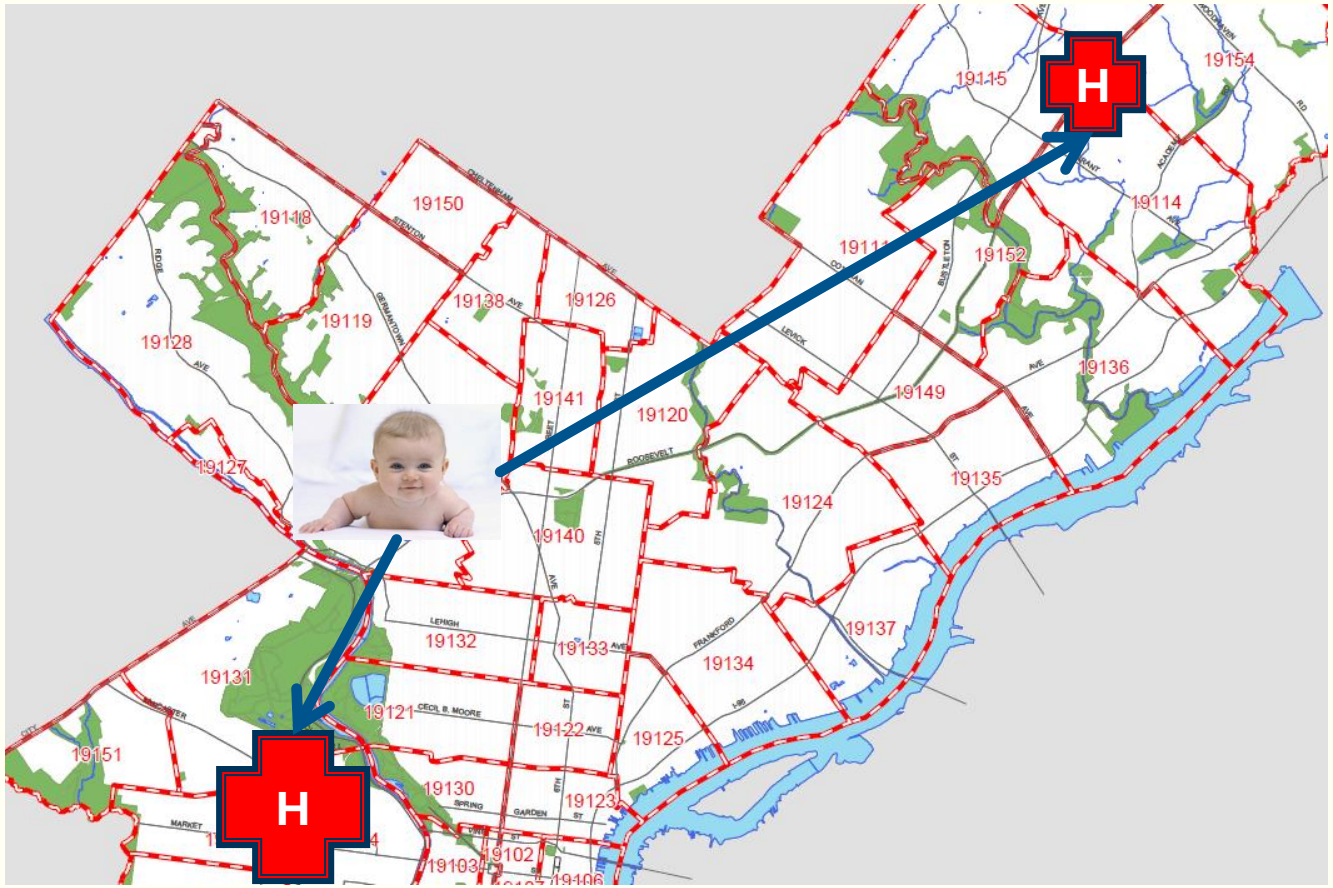

H
H
19154
19115
19114
19150
19118
19111
19152
19138
19126
19119
19128
19136
19141
19120
19144
19127
19129
19140
19132
19133
19134
19137
19131
19121
19122
19125
19151
19130
19123
19102
19103



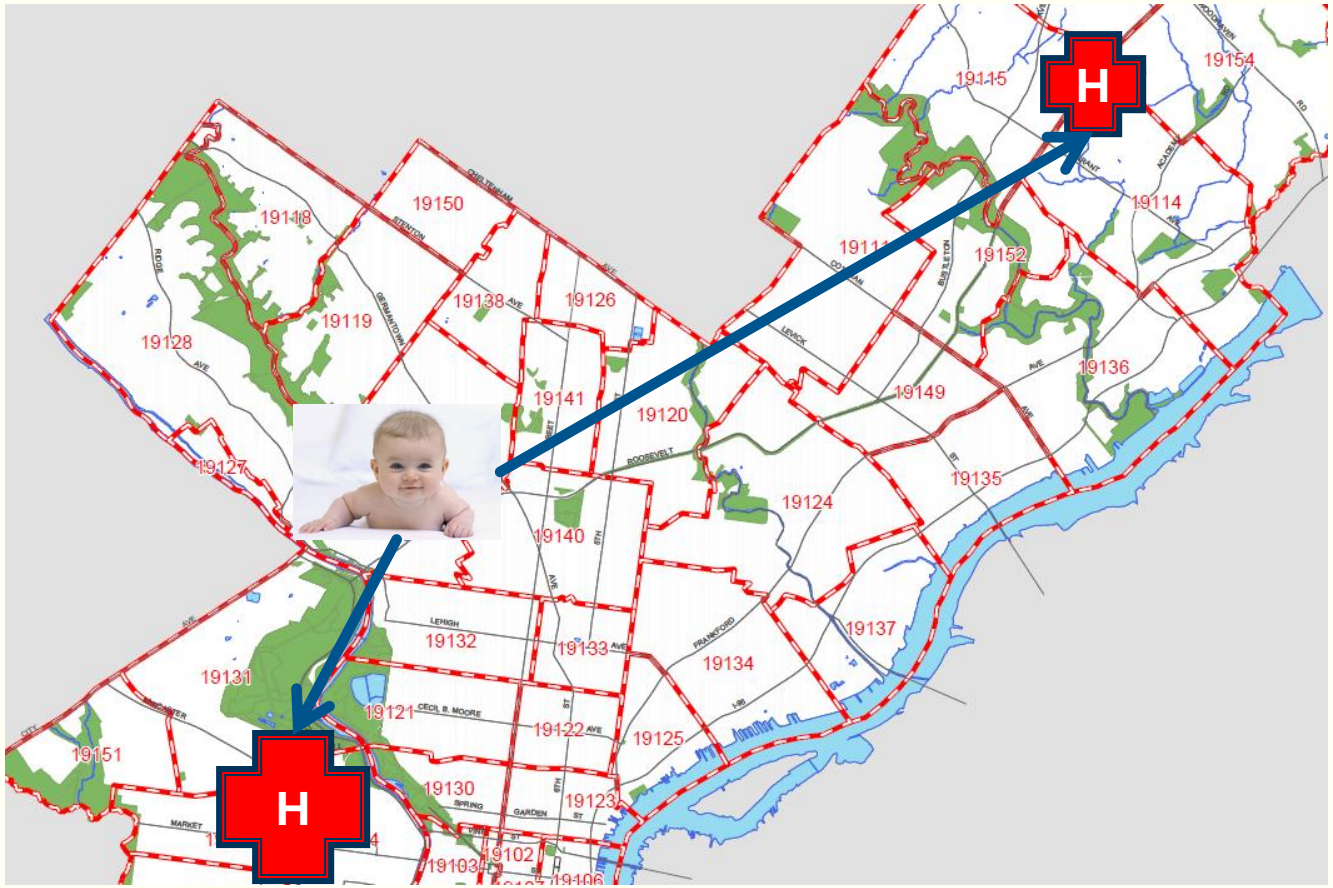
Excess Travel Time



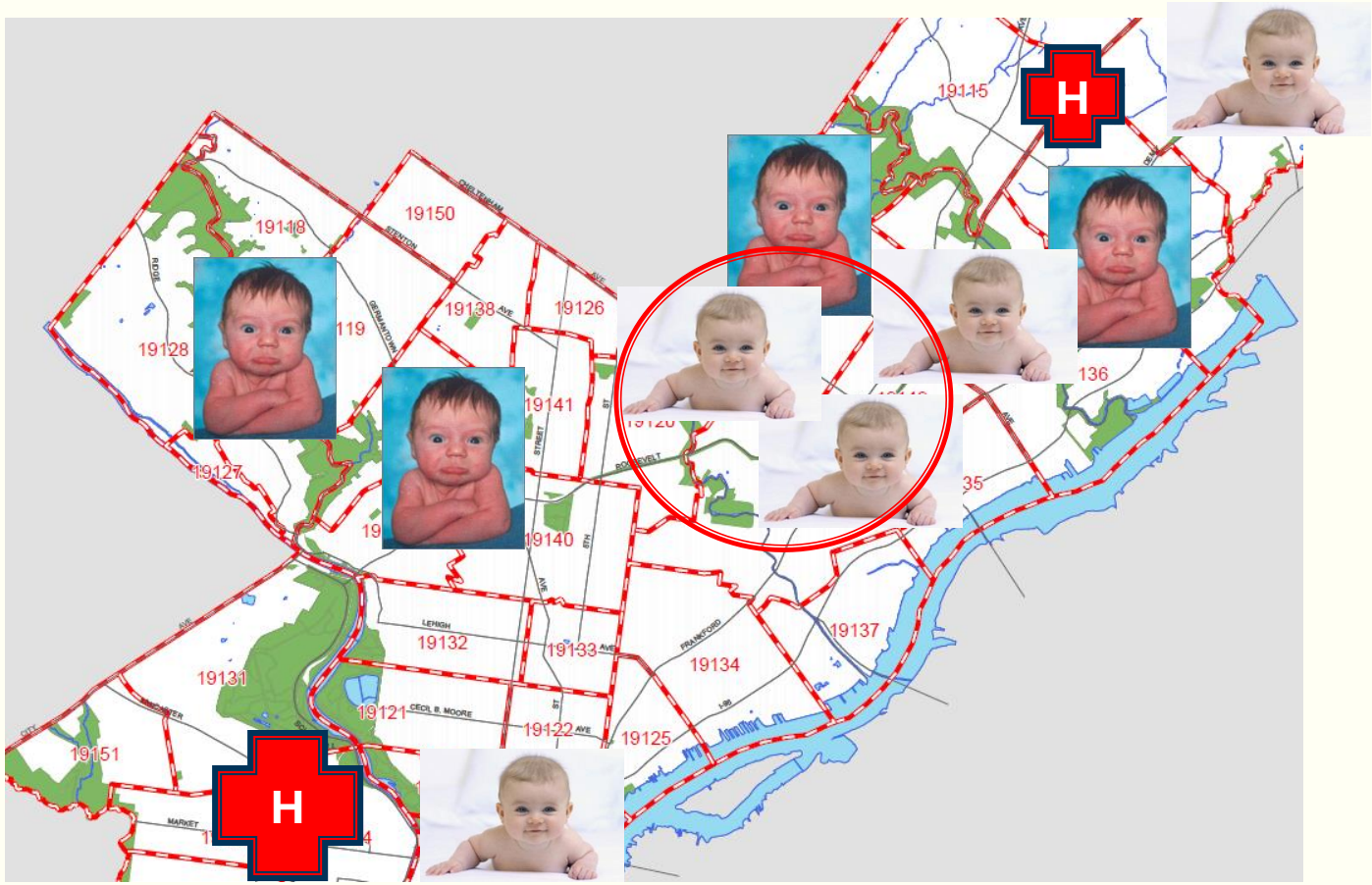
Excess Travel Time

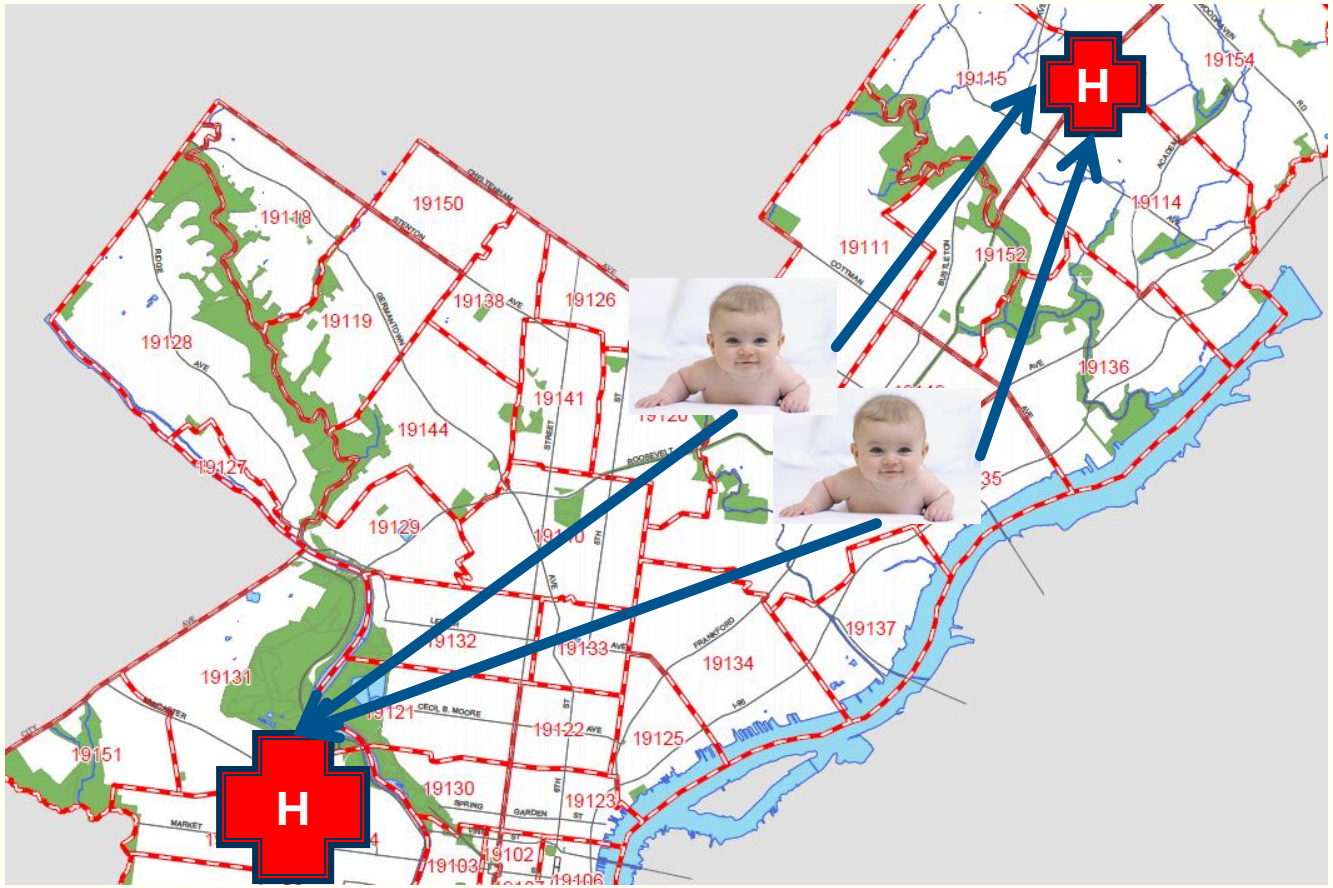


Excess Travel Time

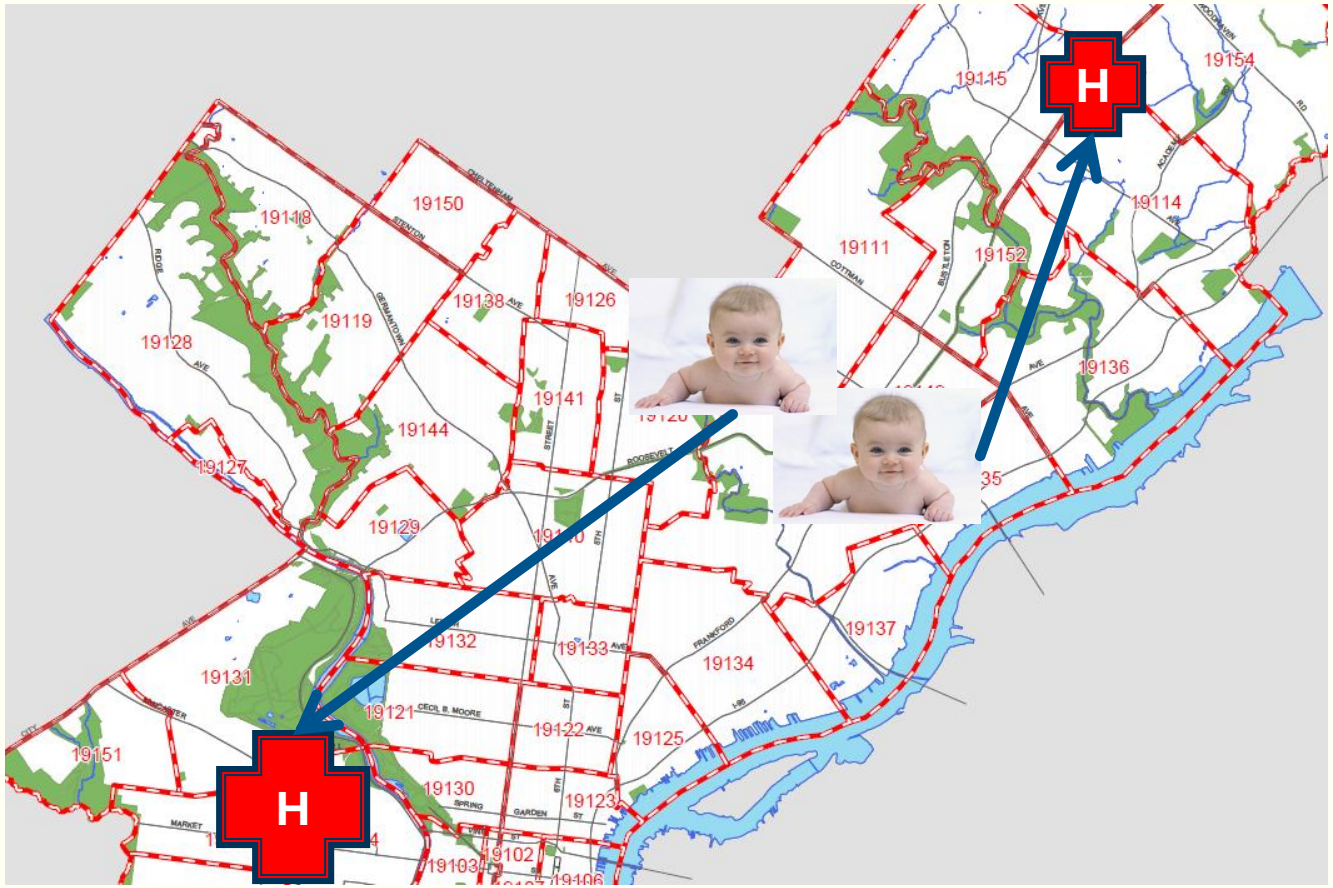


McClellan, McNeil & Newhouse; "**Does more intensive treatment of acute myocardial infarction reduce mortality?**" *JAMA.* 272(11): 859-66, September 1994

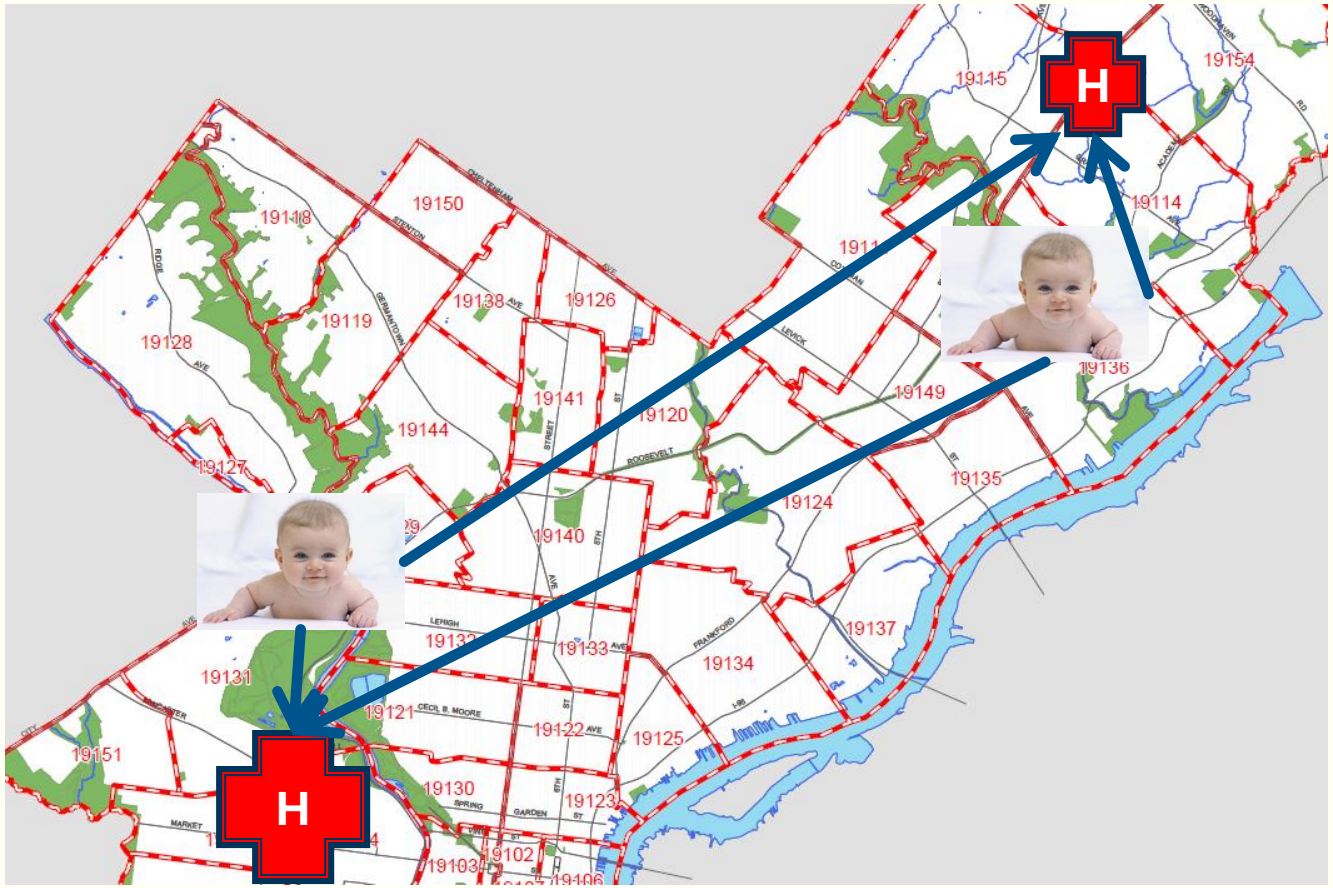

19115
H
19150
19118
19138
19126
19128
119
19141
136
19127
19140
35
19132
19133
19137
19134
19131
19121
19122
19125
19151
H


H
H
19154
19115
19114
19111
19152
19150
19148
19138
19126
19119
19128
19136
19141
19144
19127
19129
19137
19132
19133
19134
19131
19121
19125
19151
19130
19123
19102
19103

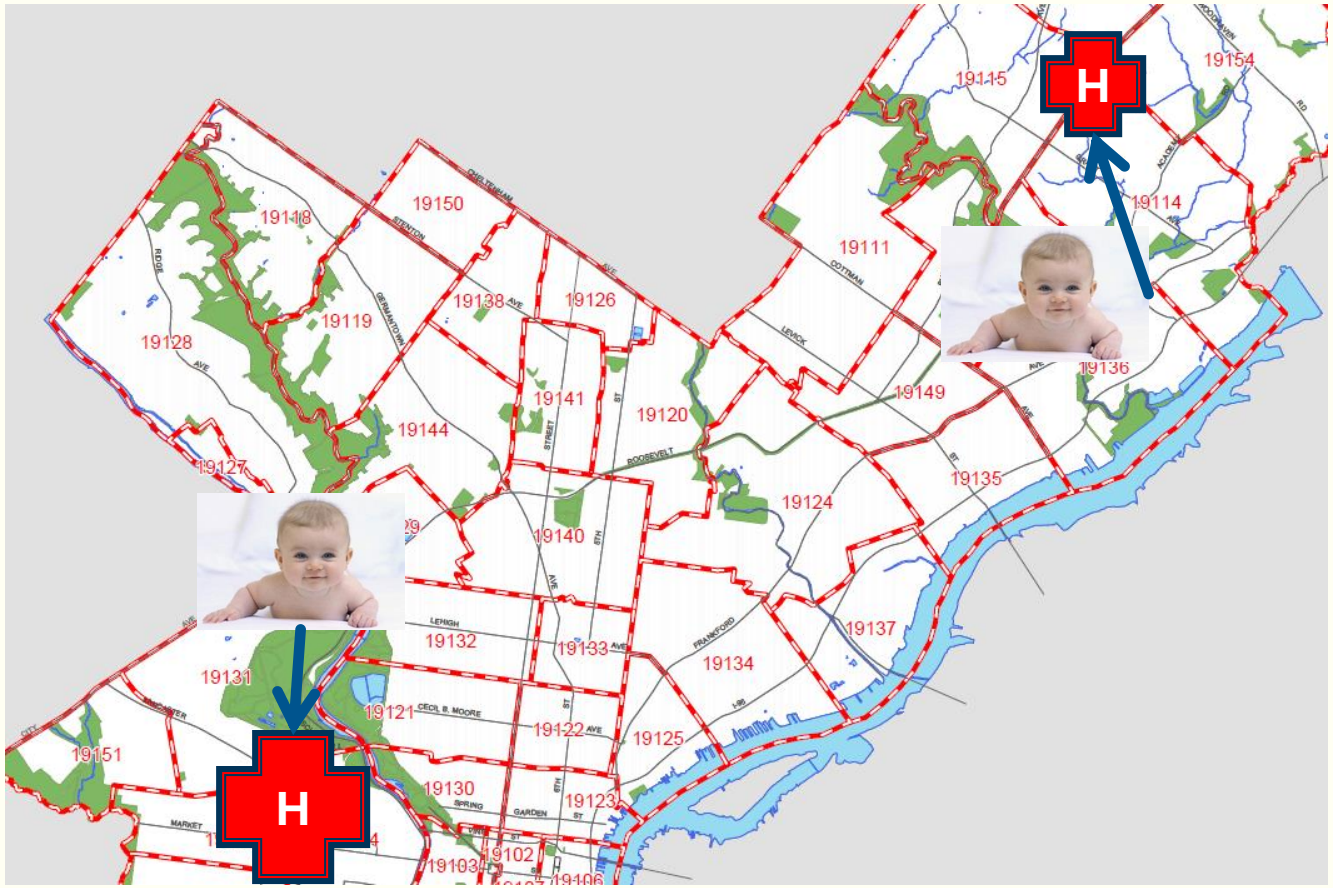
Bhattacharya and Vogt (2007) – Do Instrumental Variables Belong in Propensity Scores?



Sorting is potentially biased!


H
H
19154
19115
19114
19150
19118
19119
19138
19126
19128
19136
19149
19141
19120
19144
19127
19135
19124
19140
19137
19134
19133
19131
19121
19125
19151
19122
19130
19123
19102
19103

Baiocchi, Small, Lorch and Rosenbaum (2010) – Building a Stronger Instrument in an Observational Study



Sorting largely due to the randomness!

two flavors of “natural” experiments: instrumental variable

- Instrumental variable design
  - Example study: would regionalizing premature babies to higher-volume facilities result in better outcomes?
- We now have a detailed explanation of assignment into T vs C. It is quantified. We can debate whether it is a good pseudo-randomizer, but at least we’re being explicit about a source of potential randomness.
- Examples of IVs: distance to treatment, physician preference (e.g., painkiller medication, surgical procedure), and Mendelian randomization (e.g., genes that cause higher cholesterol levels).

Goldstone, A. B., Chiu, P., Baiocchi, M., Wang, H., Lingala, B., Boyd, J. H., & Woo, Y. J. (2018).
Second arterial versus venous conduits for multivessel coronary artery bypass surgery in California.
Circulation, 137(16), 1698-1707.

Baiocchi, M., Cheng, J., & Small, D. S. (2014). **Instrumental variable methods for causal inference.**
Statistics in medicine, 33(13), 2297-2340.

TAKEAWAYS

takeaways

- Natural experiments have a role in evidence building. They are certainly not as definitive as RCTs. Rather, these kinds of studies can cheaply accumulate and build a larger literature that points in a direction, perhaps warranting a more definitive RCT.
- The chief concern is usually the imbalances in baseline-covariates. The hope is that we can find some part of the T vs C assignment process that is haphazard and use that to emulate a target trial.
- While RCTs have higher internal validity, natural experiments tend to have higher external validity (a.k.a., “generalizability”) because we see more types of people.
- Loosely speaking, there is a spectrum of “natural experiments” as defined by how well the researcher can describe the source and behavior of the pseudo-randomizer.

less — pseudo-randomizer described — more

descriptive **propensity score** **discontinuity design** **instrumental variable** **RCT**

FIN.

baiocchi@stanford.edu