Advanced Statistical Methods for Observational Studies

LECTURE 06

class management

- Problem set 1 is posted
- Questions?

analysis plan

the structure of the argument: null effect outcomes

- Basic idea: Suppose that a treatment is known to not change a particular outcome. Then if we see differences between the treatment and control groups on this particular outcome, this must mean that there are differences between the treatment and control group on unmeasured covariates and thus there is hidden bias.

example: methylmercury fish

- Example: [Skerfving (1974)] studied whether eating fish contaminated with methylmercury causes chromosome damage. The outcomes of interest was the percentage of cells exhibiting chromosome damage. Pairs were matched for age and sex.

example: methylmercury fish

- Example: Skerfving (1974) studied whether eating fish contaminated with methylmercury causes chromosome damage. The outcomes of interest was the percentage of cells exhibiting chromosome damage. Pairs were matched for age and sex.

```
control.cu.cells <- c(2.7,.5,0,0,5,0,0,1.3,0,1.8,0,0,1,1.8,0,3.1)
exposed.cu.cells <- c(.7,1.7,0,4.6,0,9.5,5,2,2,2,1,3,2,3.5,0,4);
library(exactRankTests)

wilcox.exact(exposed.cu.cells,control.cu.cells,paired=TRUE)

        Exact Wilcoxon signed rank test

data:  exposed.cu.cells and control.cu.cells
V = 84, p-value = 0.04712

alternative hypothesis: true mu is not equal to 0
```

example: methylmercury fish

- In the absence of hidden bias, there's evidence that eating large quantities of fish containing methylmercury causes chromosome damage.
- Going further, *Skerfving* described other health conditions of these subjects including other diseases such as (i) hypertension, (ii) asthma, (iii) drugs taken regularly, (iv) diagnostic X-rays over the previous three years, (v) and viral diseases such as influenza.
- These can be considered outcomes since they describe the period when the exposed subjects were consuming contaminated fish.
- However, it is difficult to imagine that eating fish contaminated with methylmercury causes influenza or asthma, or prompts X-rays of the hip or lumbar spine.

null effect outcomes

- Power of the test of hidden bias: Let $\boldsymbol{y}$ denote the outcome for which there is a known effect of zero. For a particular unobserved covariate $\boldsymbol{u}$, what unaffected outcome $\boldsymbol{y}$ would be useful in detecting hidden bias from $\boldsymbol{u}$?
- Precise statement of results in:
  - "The Role of Known Effects in Observational Studies"
- Basic result: The power of the test of whether $\boldsymbol{y}$ is affected by the treatment increases with the strength of the relationship between $\boldsymbol{y}$ and $\boldsymbol{u}$. If one is concerned about a particular unobserved covariate $\boldsymbol{u}$, one should search for an unaffected outcome $\boldsymbol{y}$ that is strongly related to $\boldsymbol{u}$.

takeaway

- Having a detailed understanding of how your intervention functions, what the causal pathway includes and excludes, will give you more data sources that may validate or refute your hypothesis.
- Coherence is trying to flesh out your hypothesis.
- Known null effects may help to address unobserved confounding

a second control group

TWO PROBLEMS
TWO CONTROLS

Design of Observational Studies: chapter 5.2.2
Rosenbaum, "The Role of a Second Control Group in an Observational Study"

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). Importantly: all participants can receive C or T.
- In an observational study there are possibly many different reasons for people to have not received the treatment.
- In some situations there are discernable subgroups within the non-treatment group, each subgroup being identifiable by the reason for the subgroup not receiving the treatment.
- In some subset of these situations these subgroups will be open to critiques of bias when compared to the treatment group, but at least two of the subgroups will differ in the nature of their bias.
- The contrast of these two control groups with the treatment group may strengthen your analysis.

second control group: army toxicity

- Example: The army is interested in the long term effects of exposure to a list of specific chemical agents that were suspected of being toxic. Relatively few soldiers were exposed to these chemicals.
- At first pass, one might think to compare these exposed ("treated") service members to service members who were not exposed at all.
- Complicating that comparison, though, is that the army sorted people into jobs which exposed them or to jobs which did not.
- The army used medical examinations – which were not well documented – to sort some individuals out of high-exposure jobs. This leaves the comparison between exposed and strictly unexposed potentially biased due to baseline conditions.

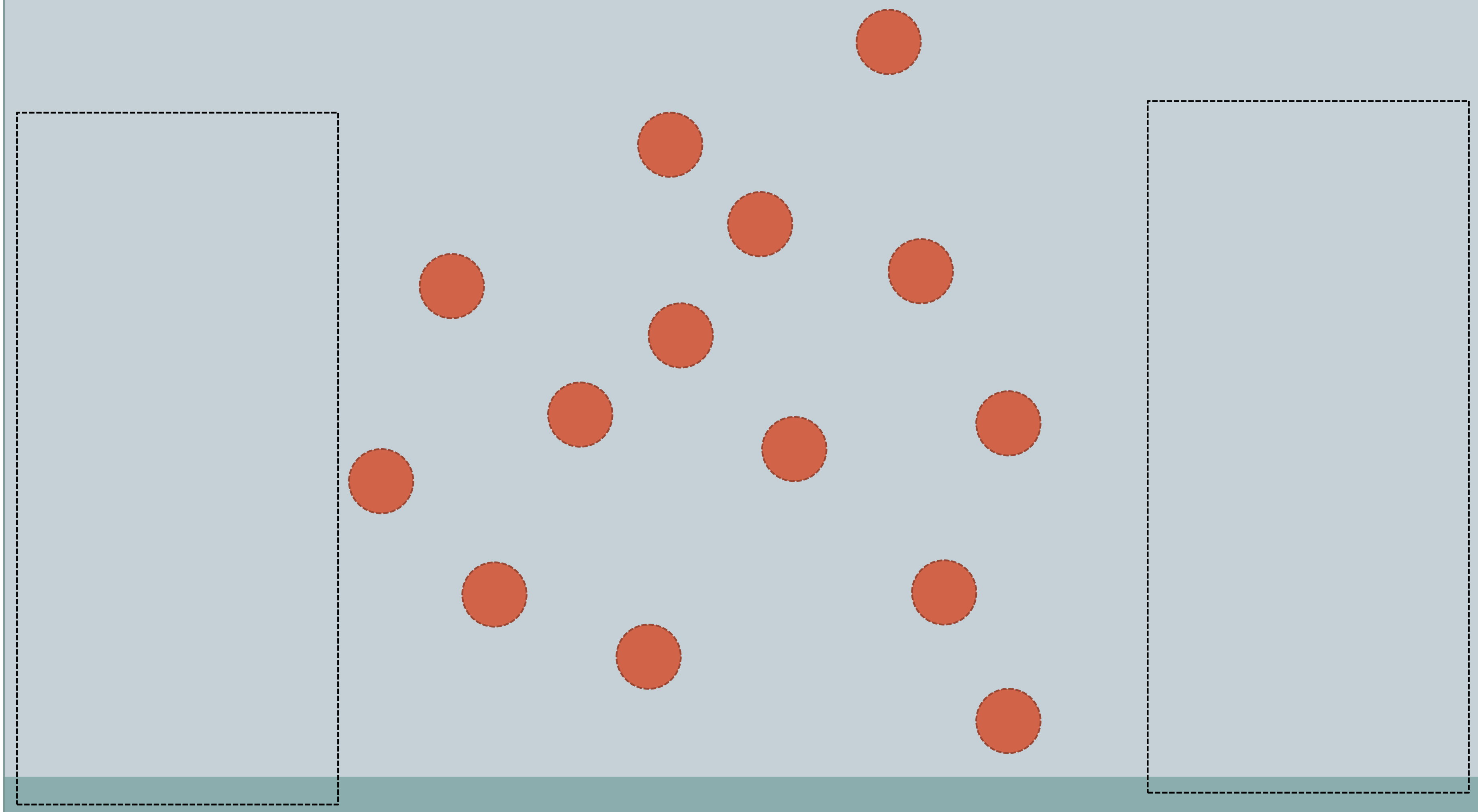
second control group: army toxicity

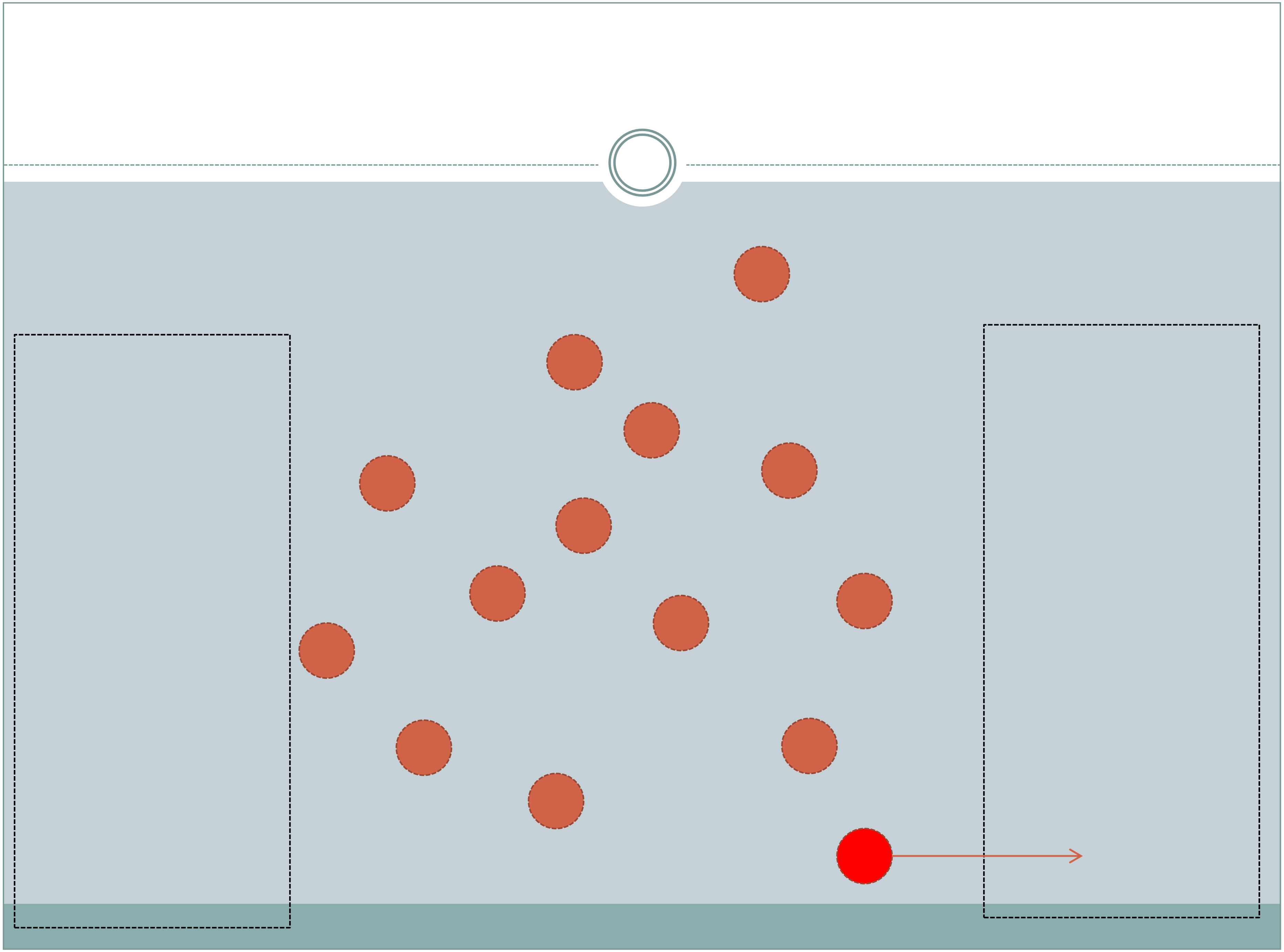
- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects. Thus this group is thought to have received an "ineffective dose" of the exposure.
- Each of these control groups is problematic: the first group is open to critiques of baseline differences in medical conditions; the second group has individuals who were potentially exposed to actively toxic chemical agents.
- But the first control group is unlikely to suffer from the bias encounter in the second control group, and vice versa.

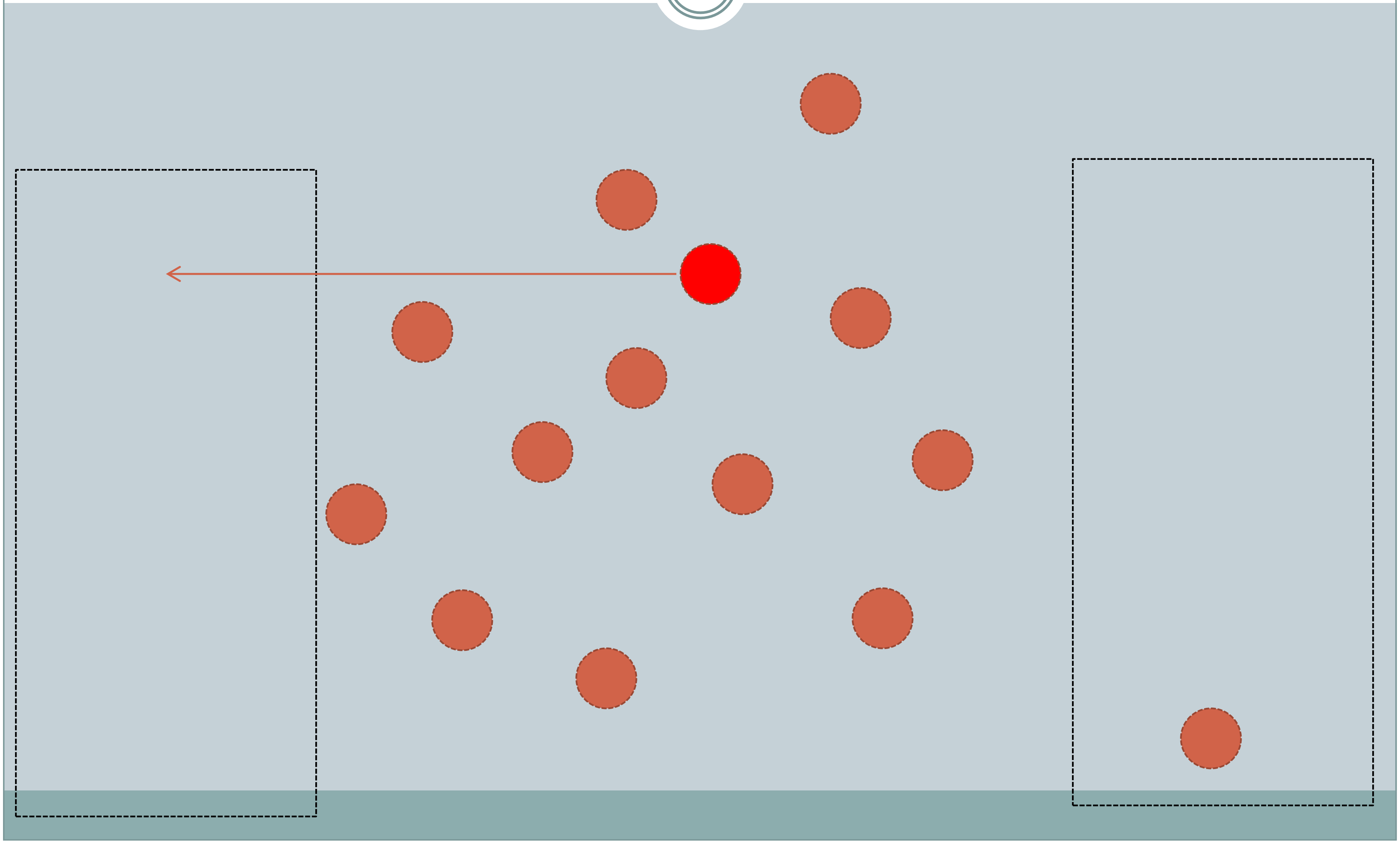
second control group: army toxicity

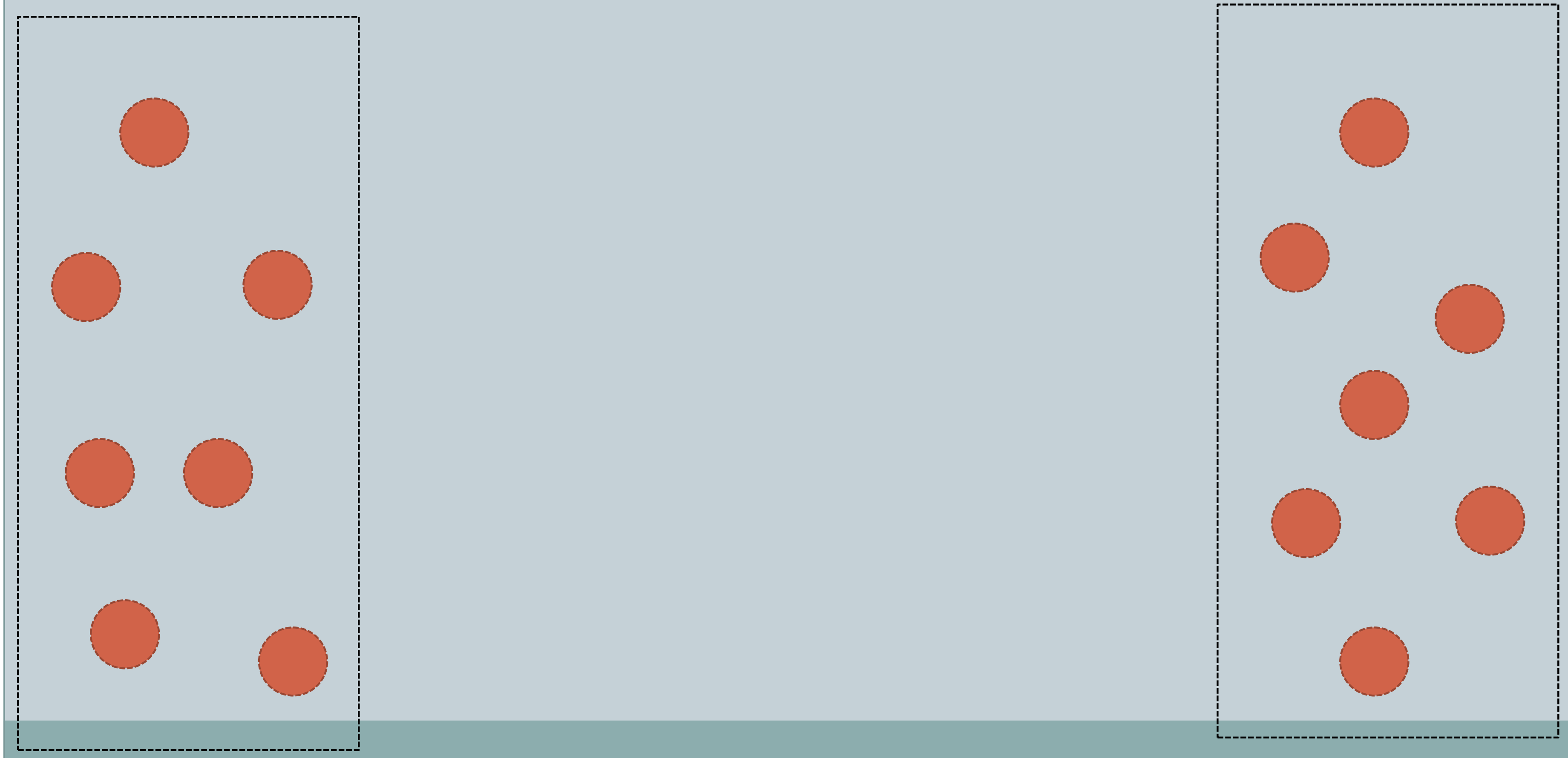
- The hope is that the two control groups will not differ from each other in a meaningful way.
- A rejection of a test of equivalency between the control groups is a strong warning sign of potential bias.
- A non-rejection may arise for several reasons. A false-negative would be problematic.
- The hope is that the control reservoir (i.e., the ratio of controls to treated observations) is large enough that we can reach adequate levels of statistical power for our tests.
- Precise statements of how this argument works statistically, as well as a couple more examples from the literature, can be found in "The Role of a Second Control Group in an Observational Study"

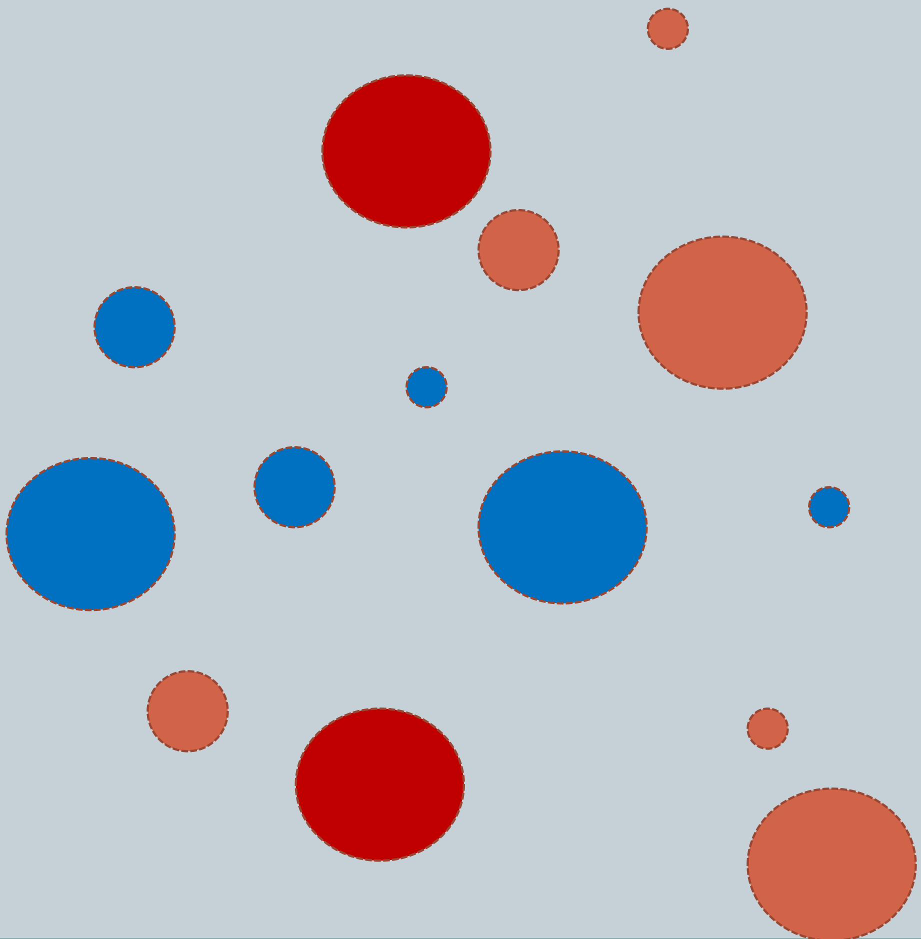
design thus far

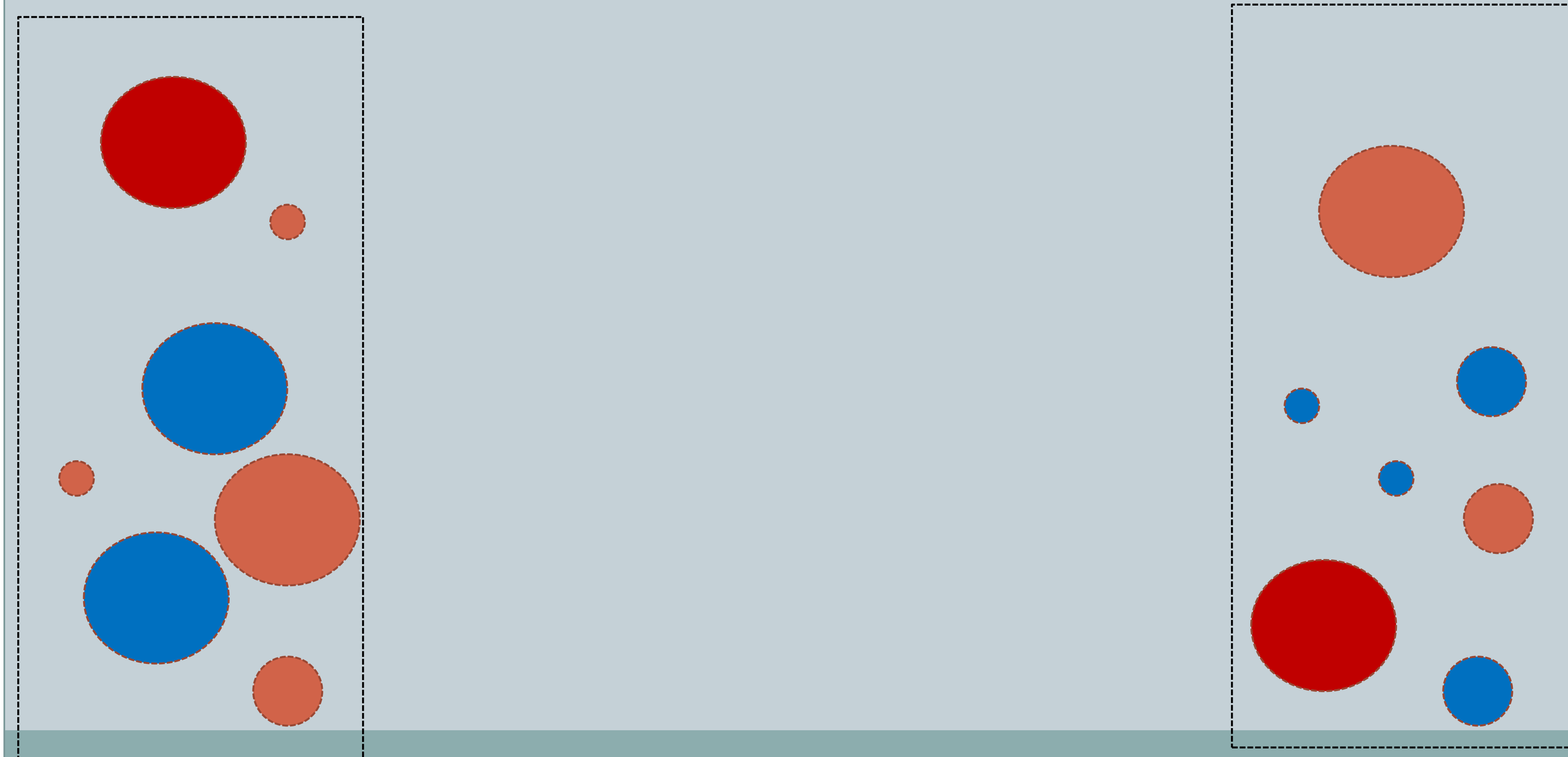


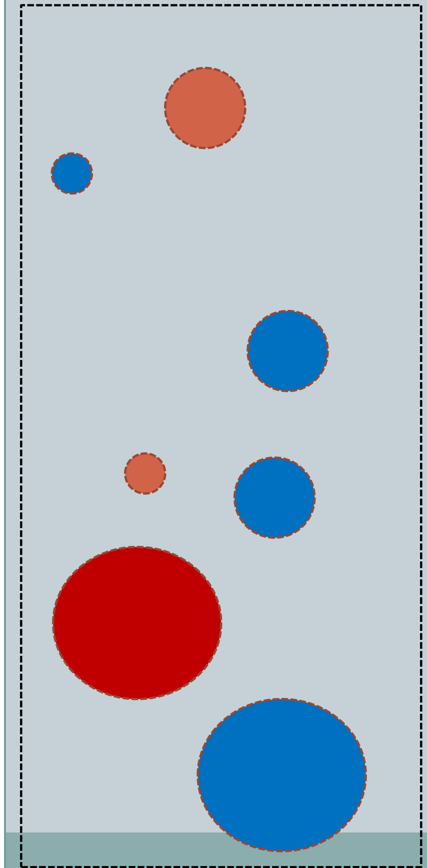


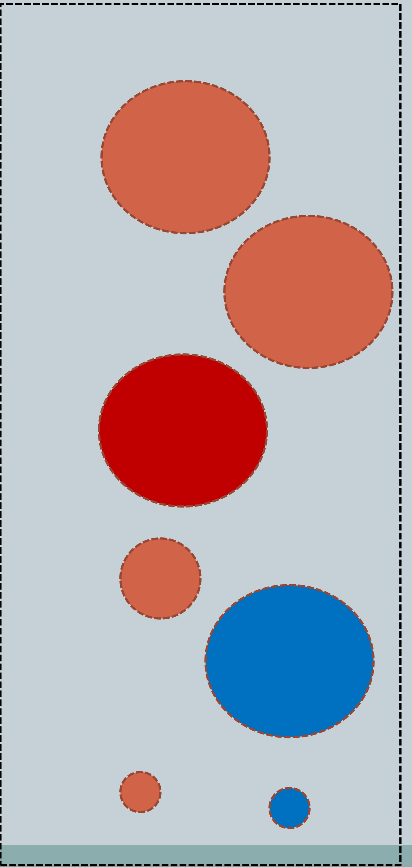


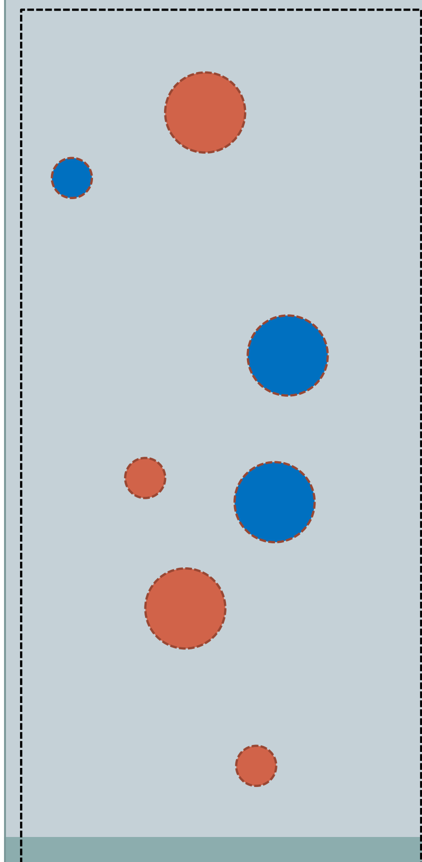




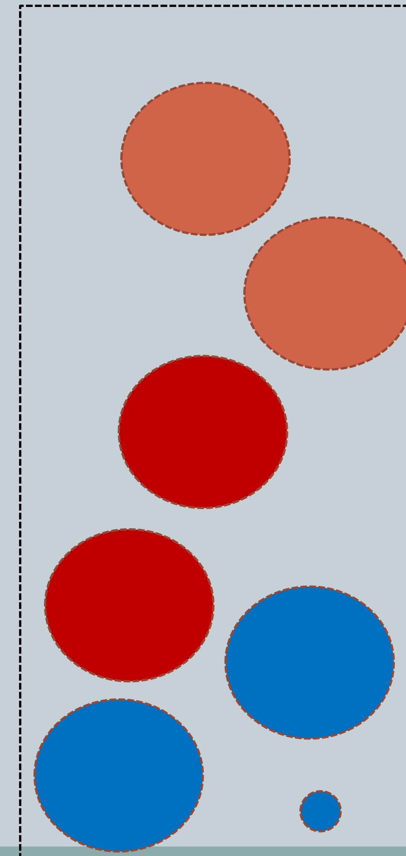


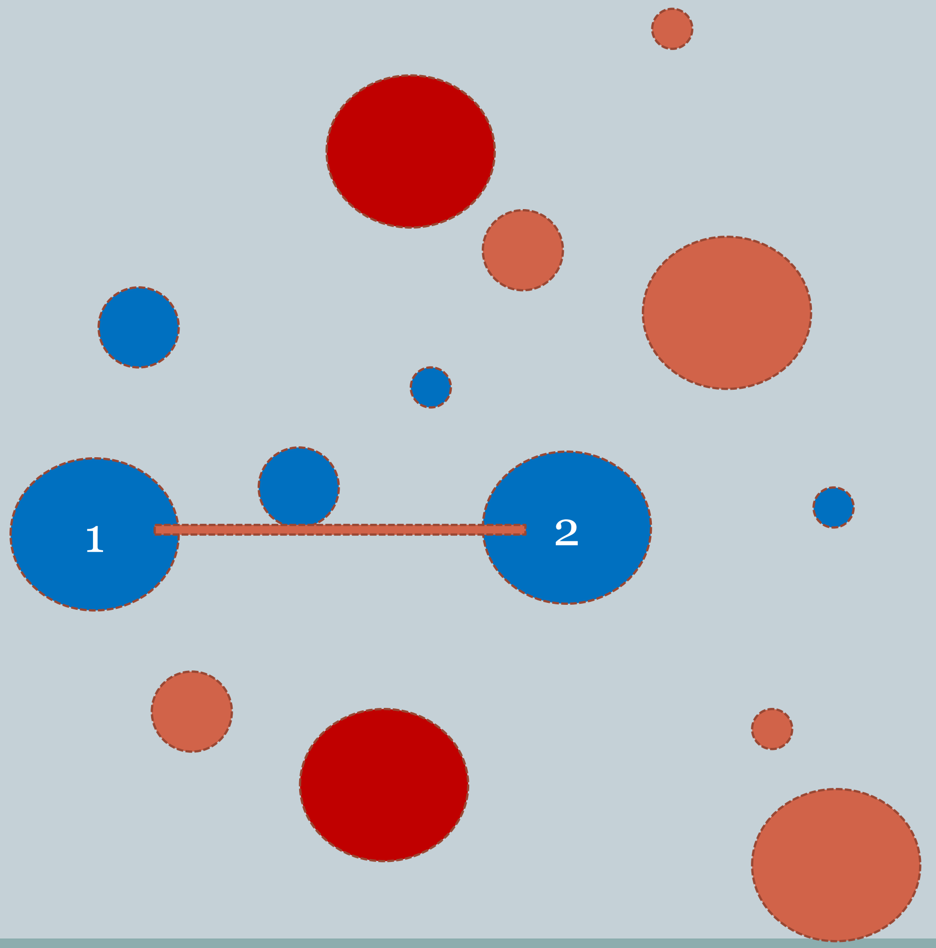


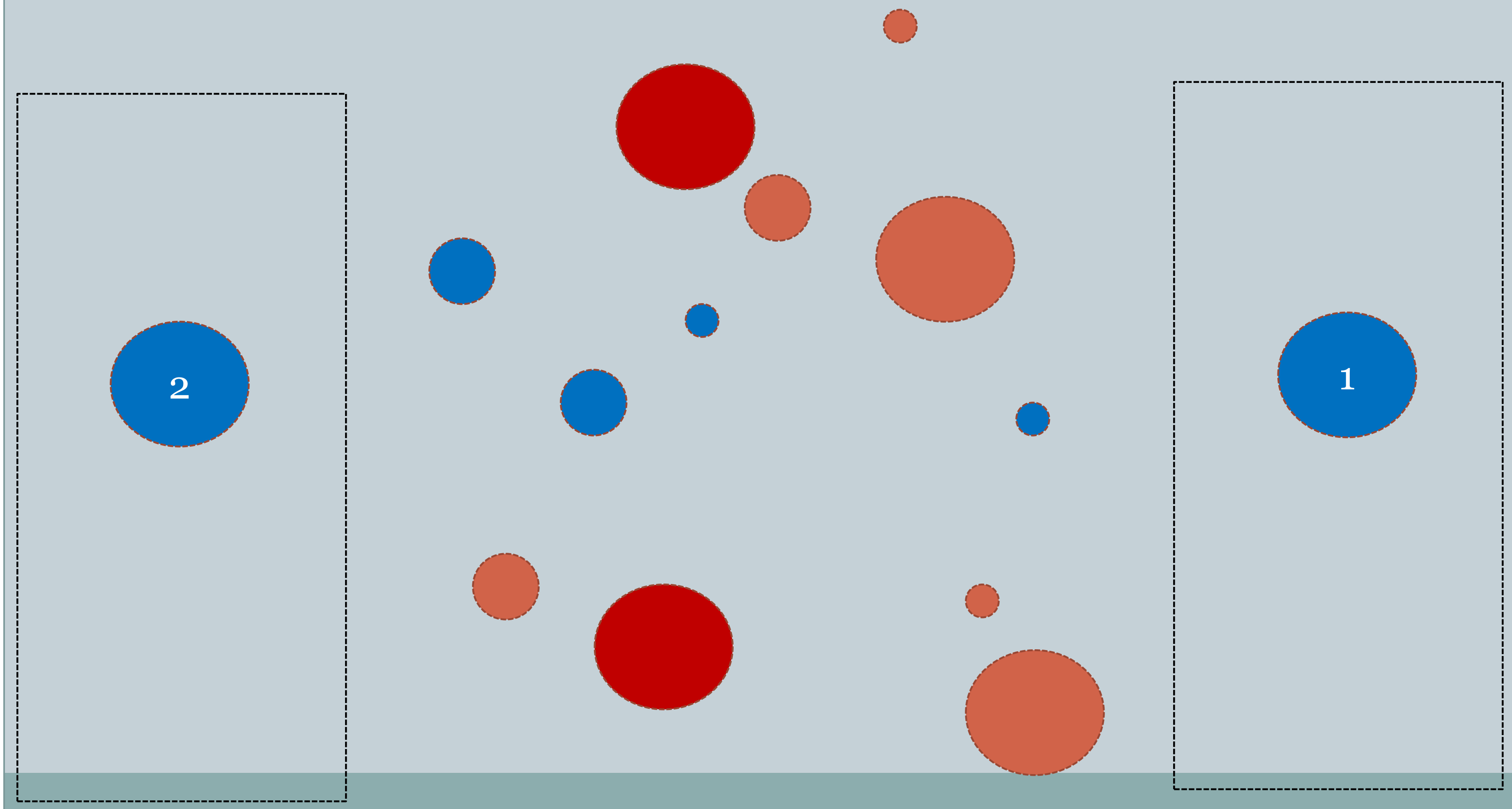


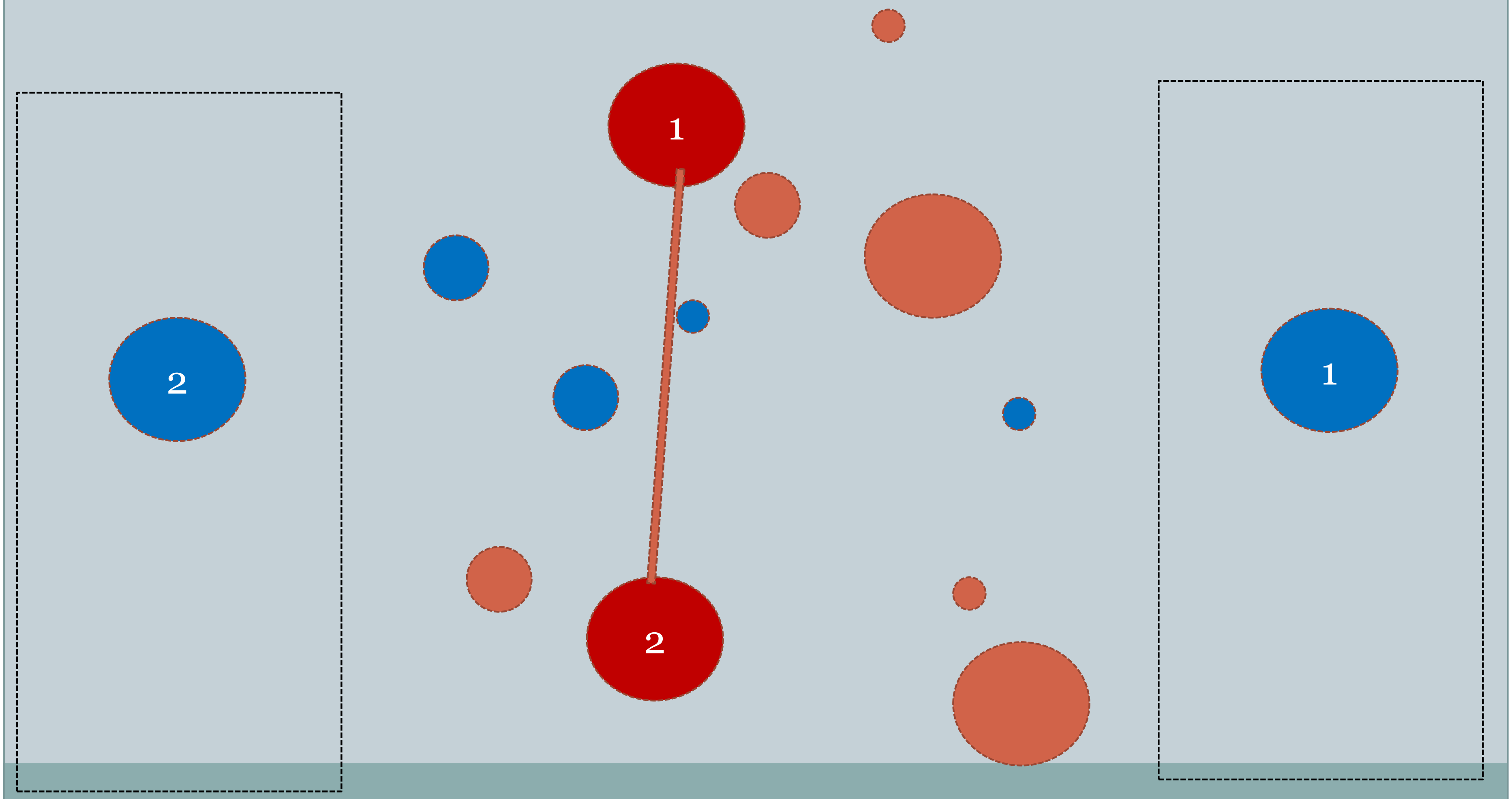


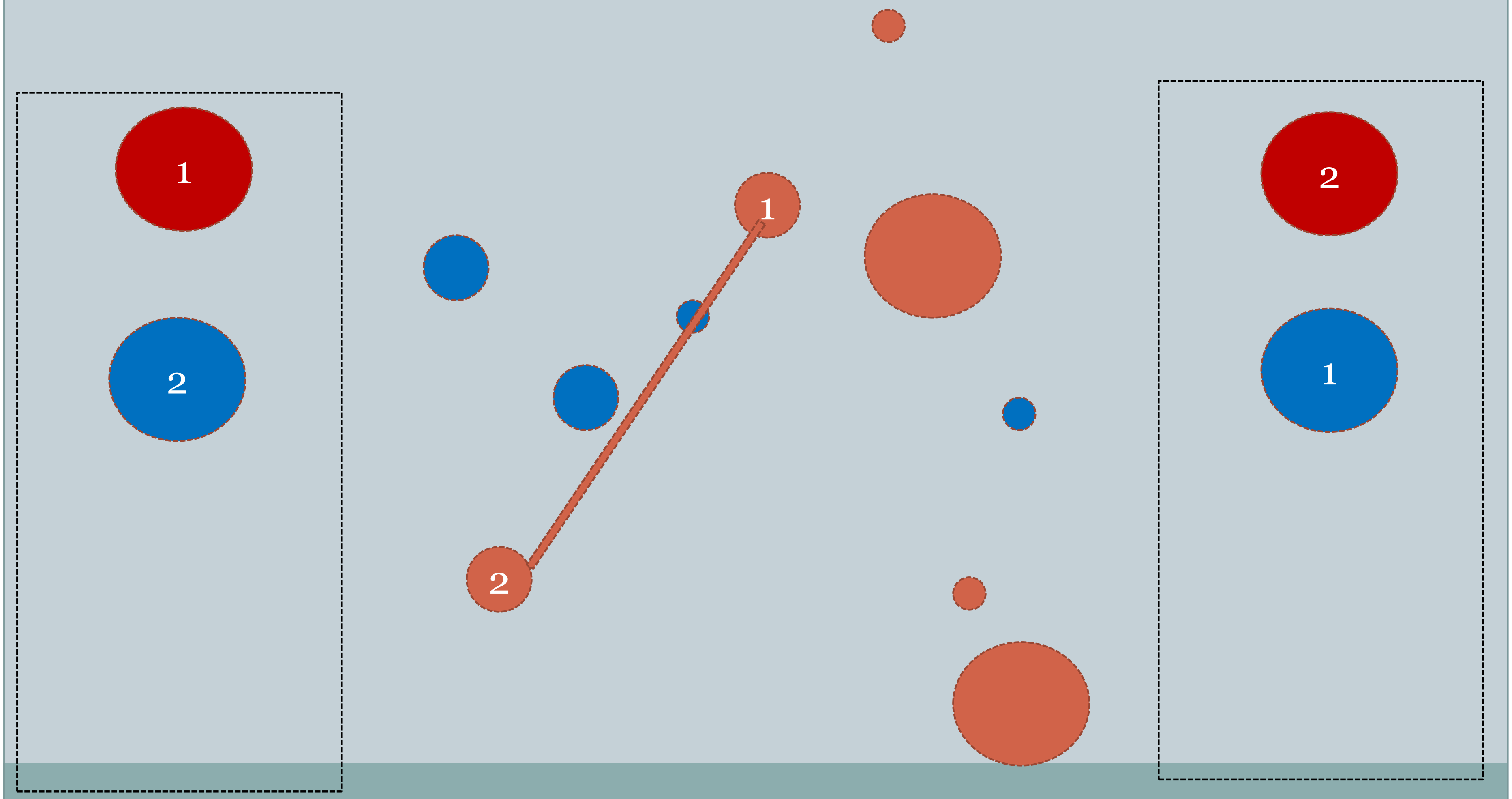
We’re off to a **bad** start.

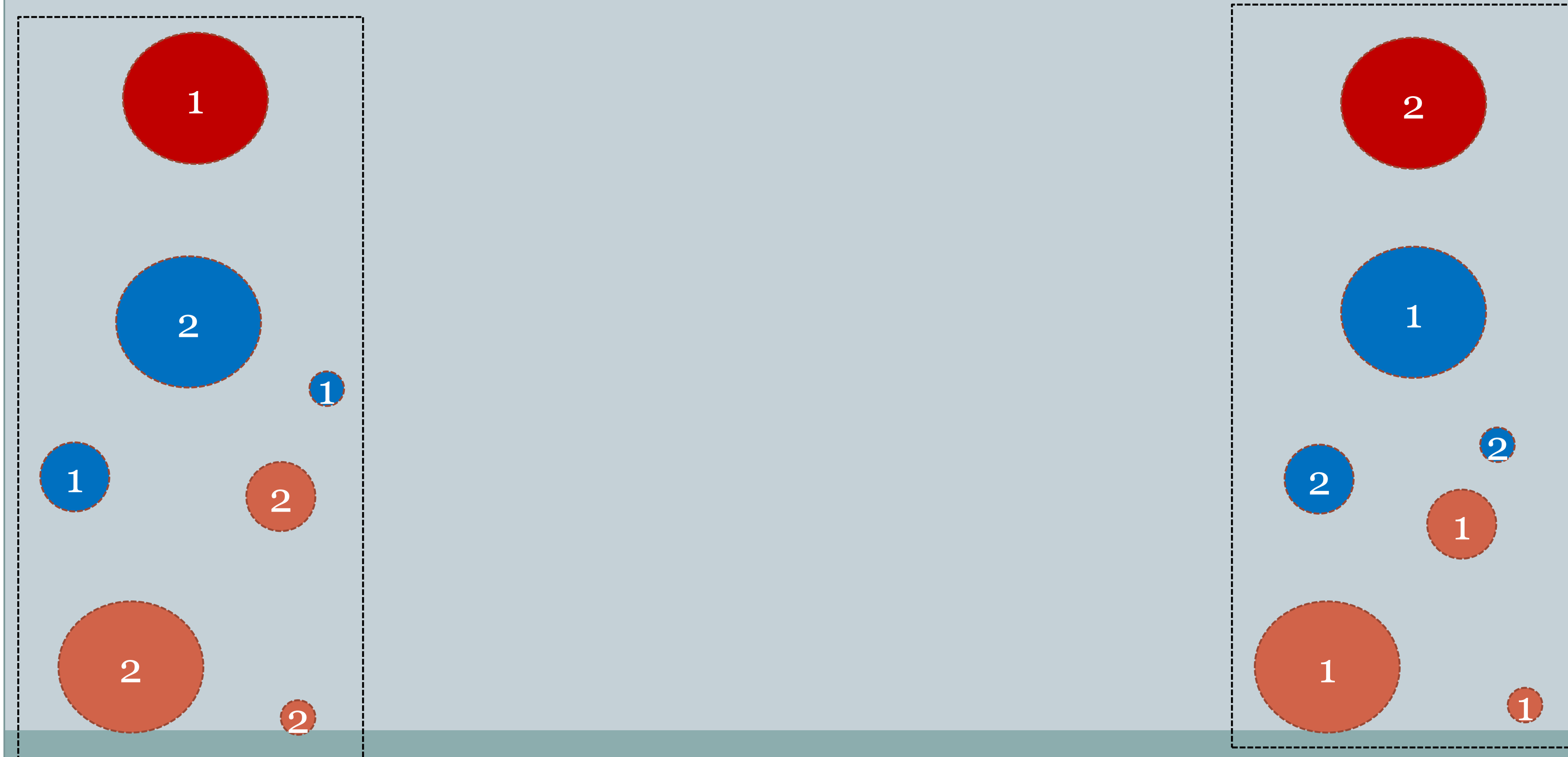


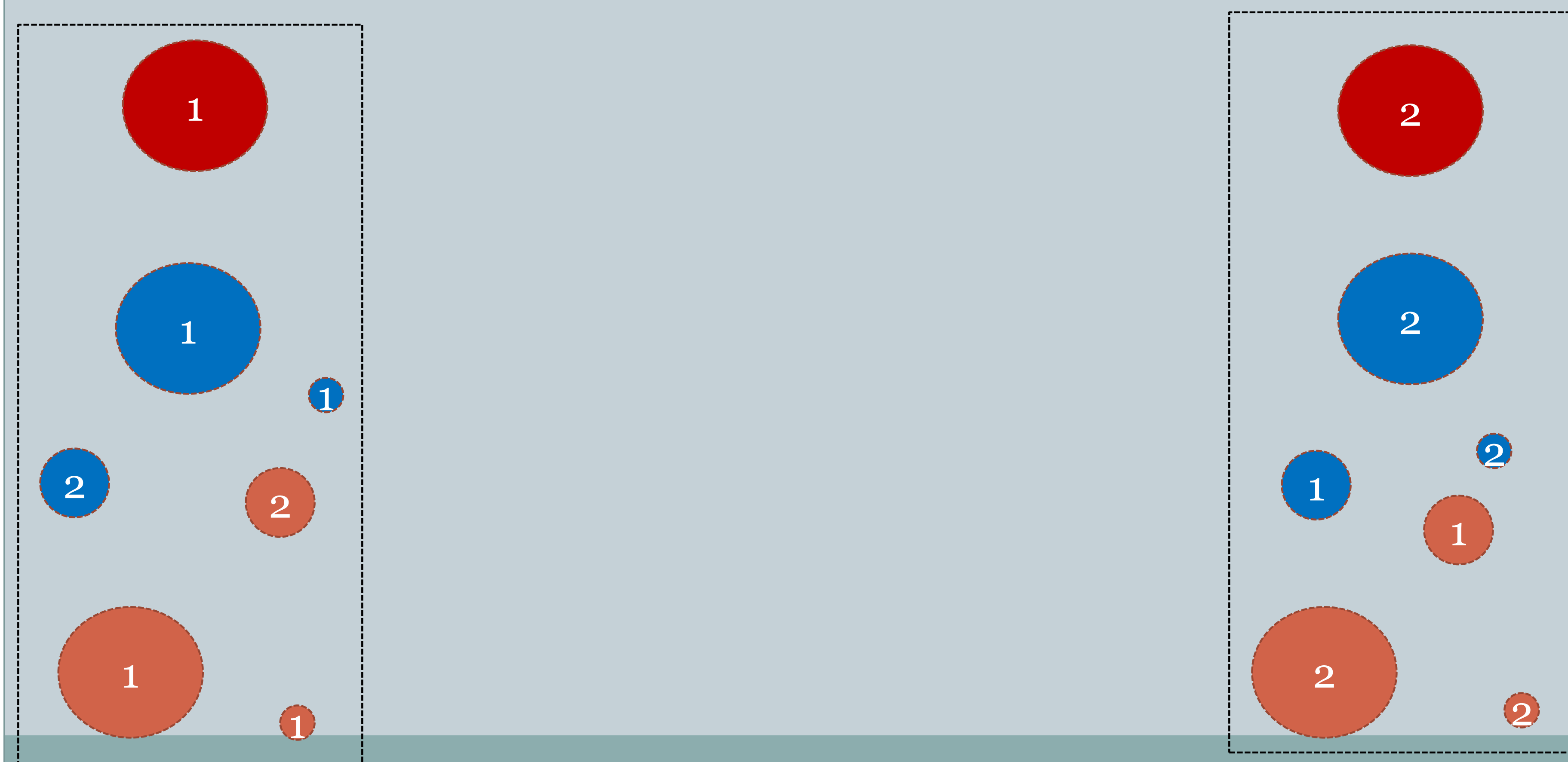

1
2

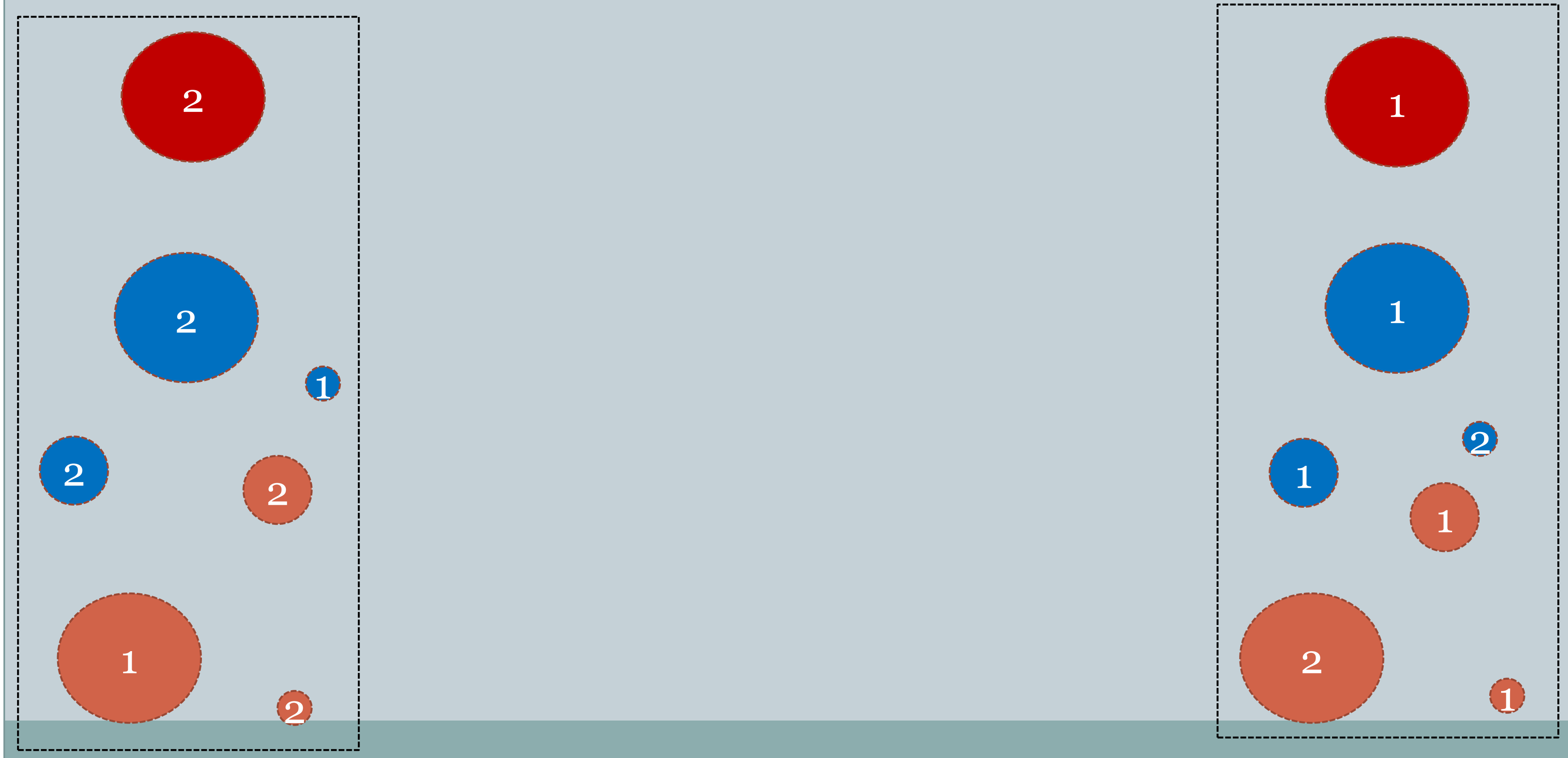

2
1


1
2
2
1


1
2
1
2
2
1


1
2
1
1
2
2
2
2
1
2
2
1
1
1


1
2
1
1
2
2
2
1
2
2
1
1
1
2


2
2
1
2
2
1
2
1
1
2
1
1
2
1

the design limits what could have happened

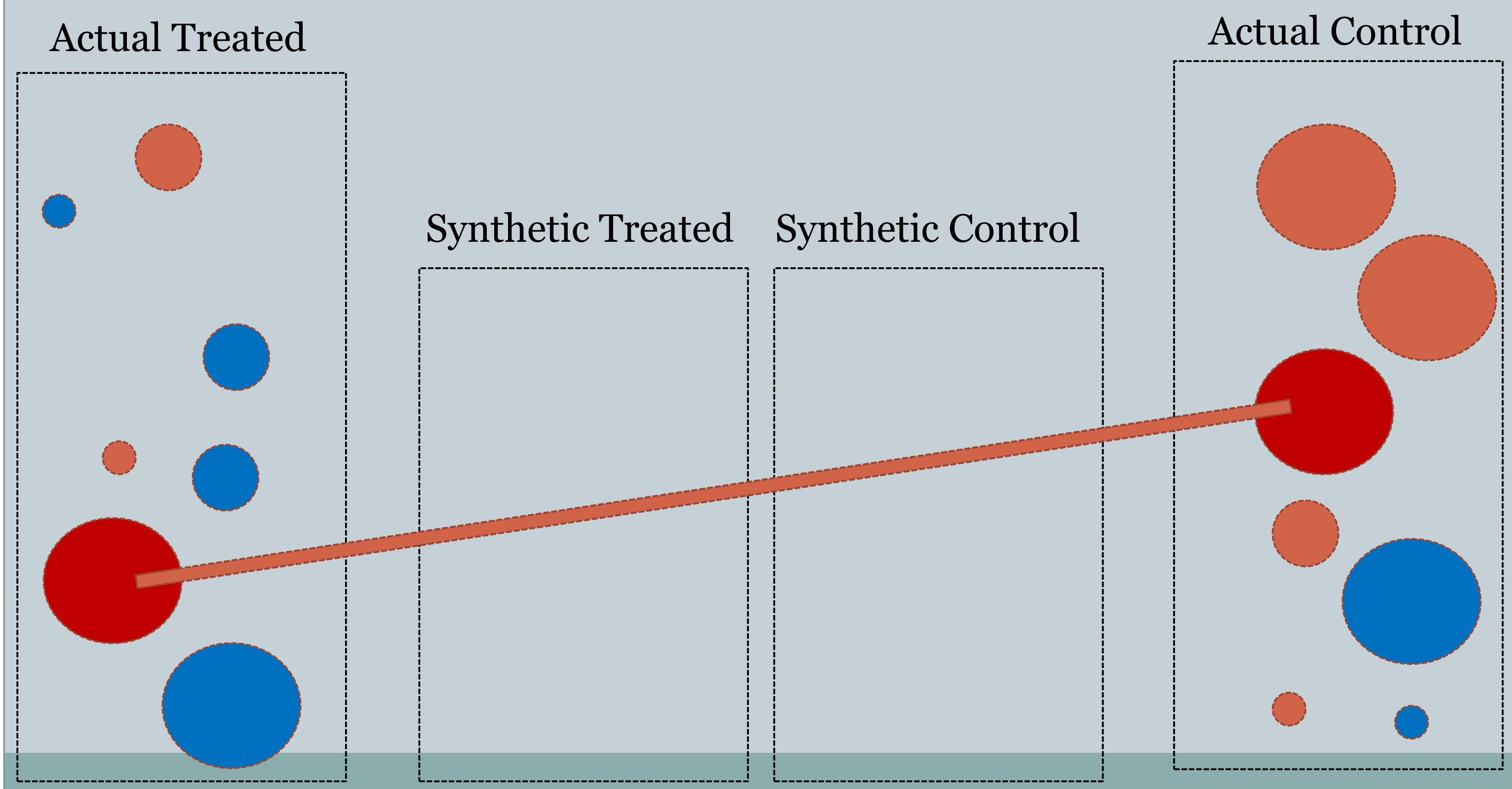
- Uniform randomization allows for quite different possibilities. (${}_nC_r$)
- Matched-pairs randomization limits the size, and range, of possible assignments. ($2^{(\frac{n}{2})}$)
- In some sense, we're losing something when we go to matched-pairs…
- … but what are we losing? The "crazy" options that we know are going to lead us astray.

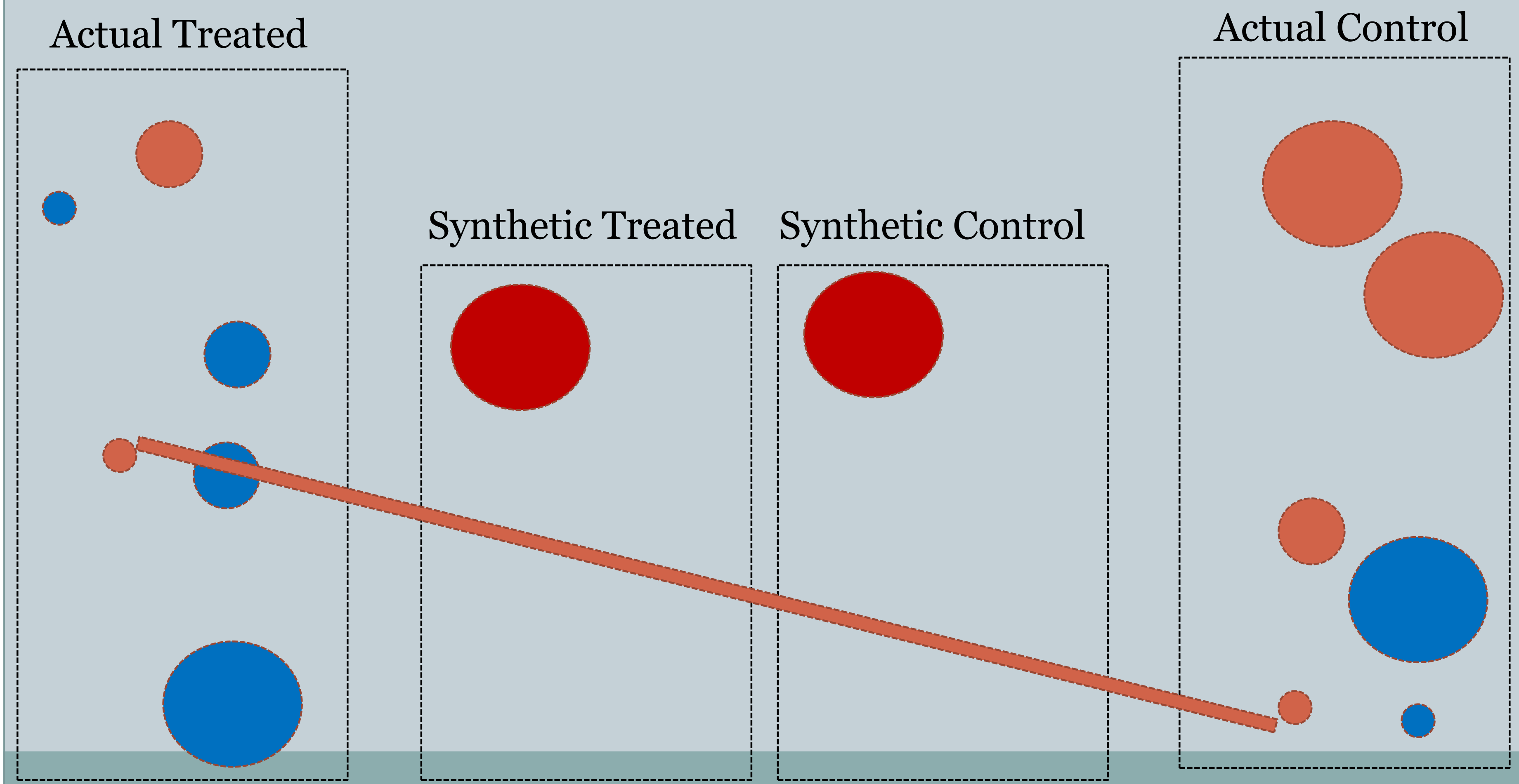
takeaway

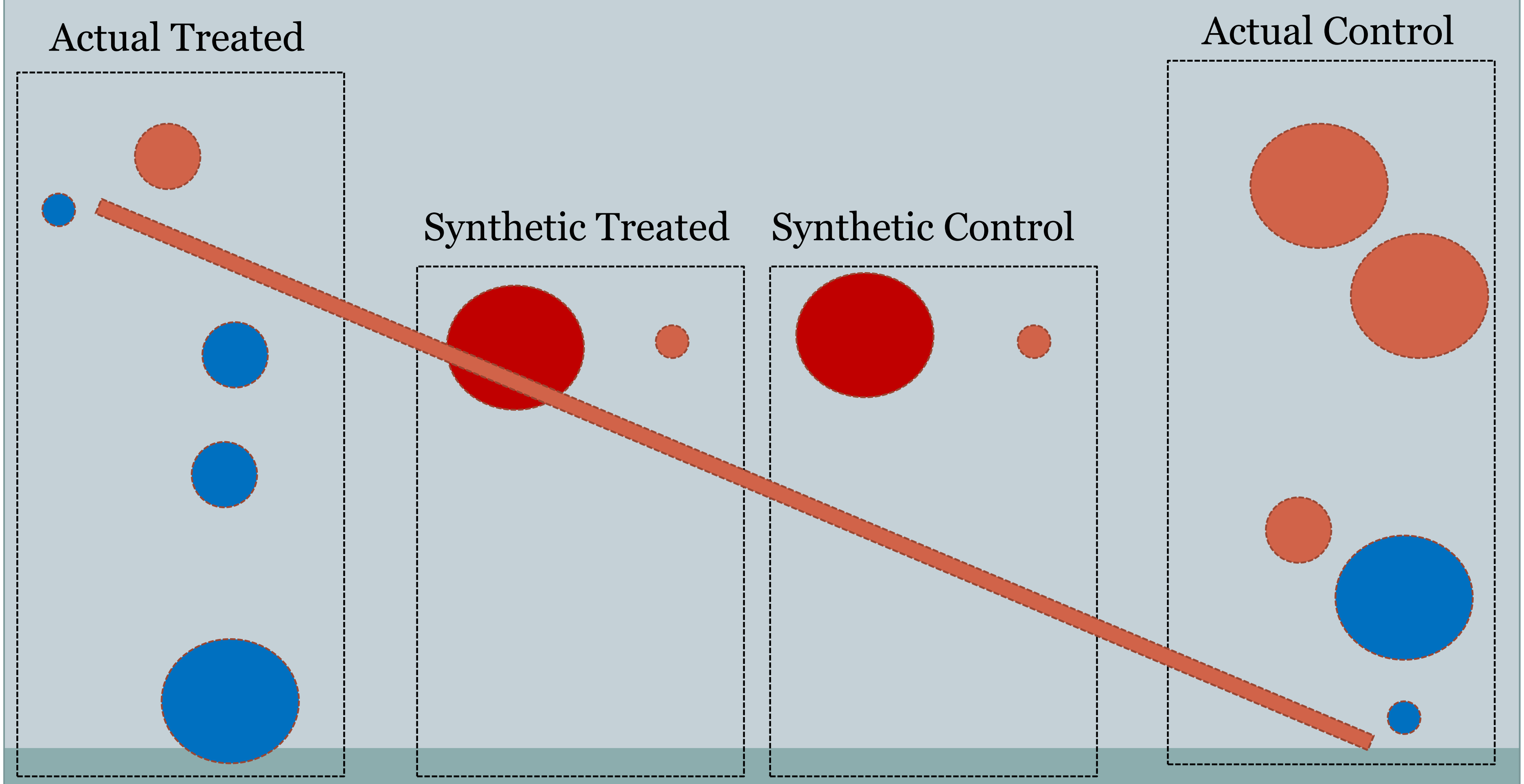
- Data come from somewhere.
- Often, we can preferentially select aspects of the data to be analyzed (i.e., design our study).
- We will design our study to improve the quality of our resulting analysis – strengthen our inference.

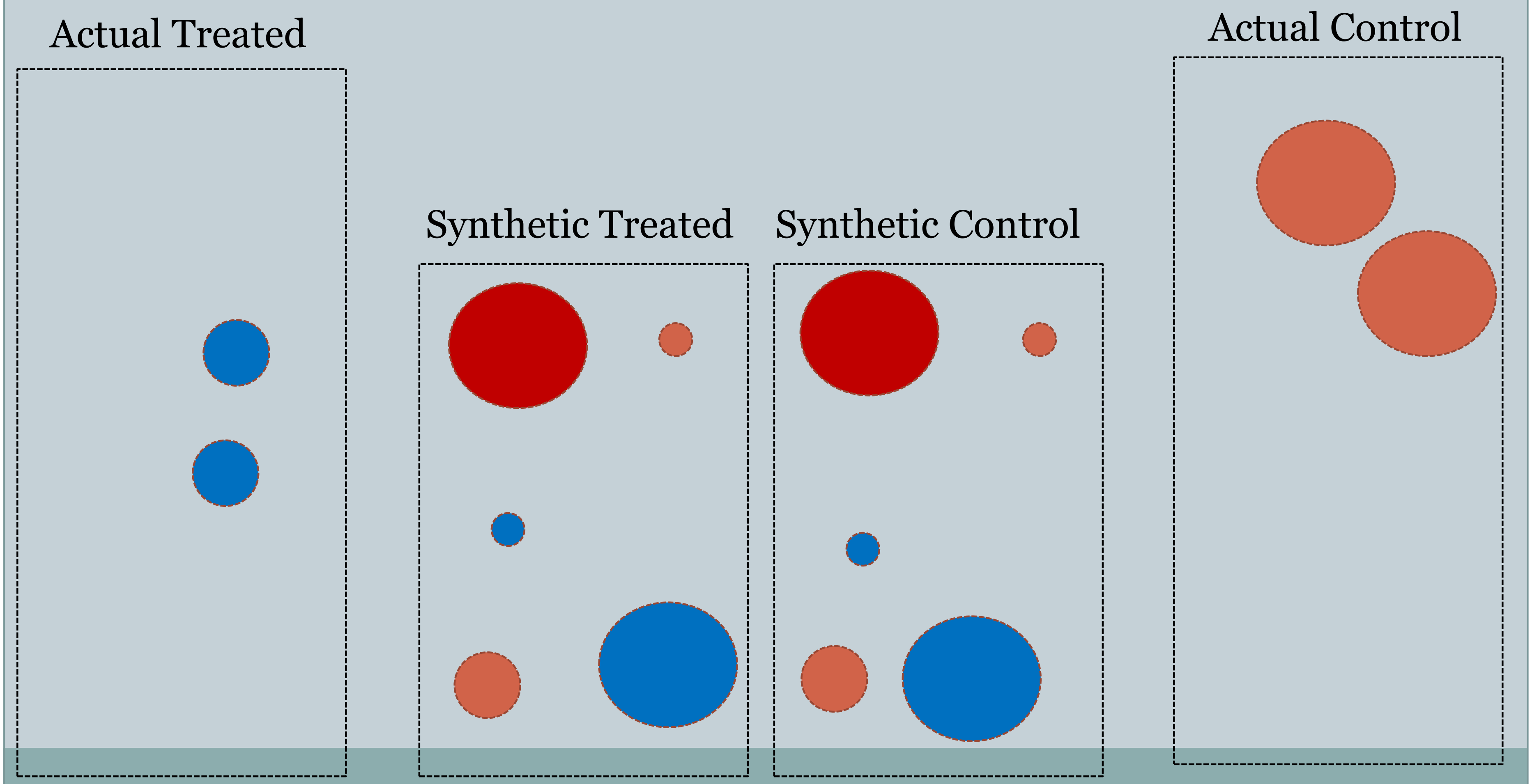
matching

RECREATING AN RCT


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control

inverse probability weighting

RECREATING SURVEYS

inverse-probability weights

- A gentle introduction to best practices here.
- Let Z denote treatment assignment.
- Let $e = \Pr(Z = 1|\boldsymbol{X})$; that is e is the propensity score.
- The inverse probability treatment weight is defined as:

$$w = \frac{Z}{e} + \frac{1 - Z}{1 - e}$$

- Each subject's weight is equal to the inverse of the probability of receiving the treatment that the subject received.

inverse-probability weights

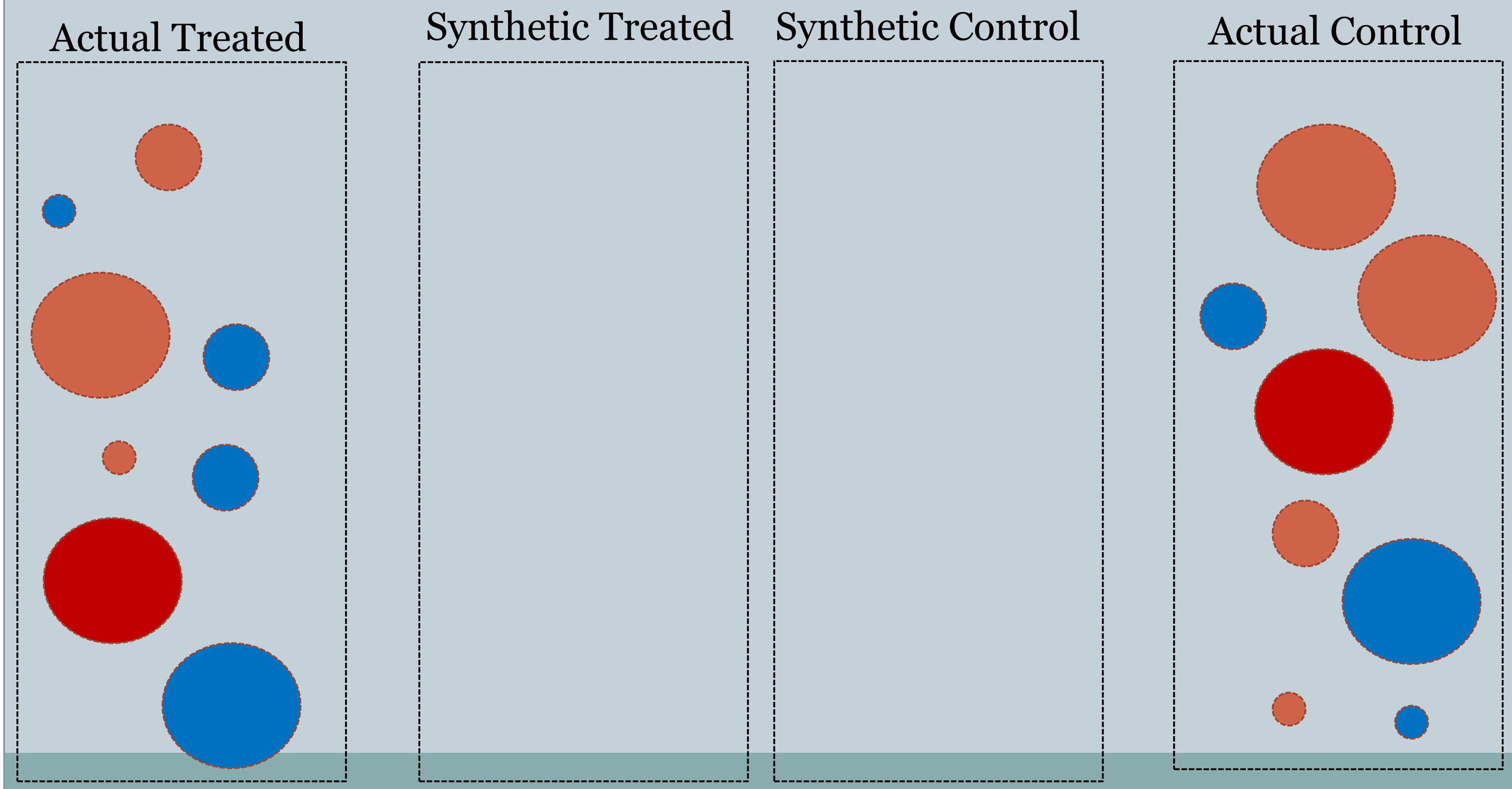
- To estimate the average treatment effect (ATE):

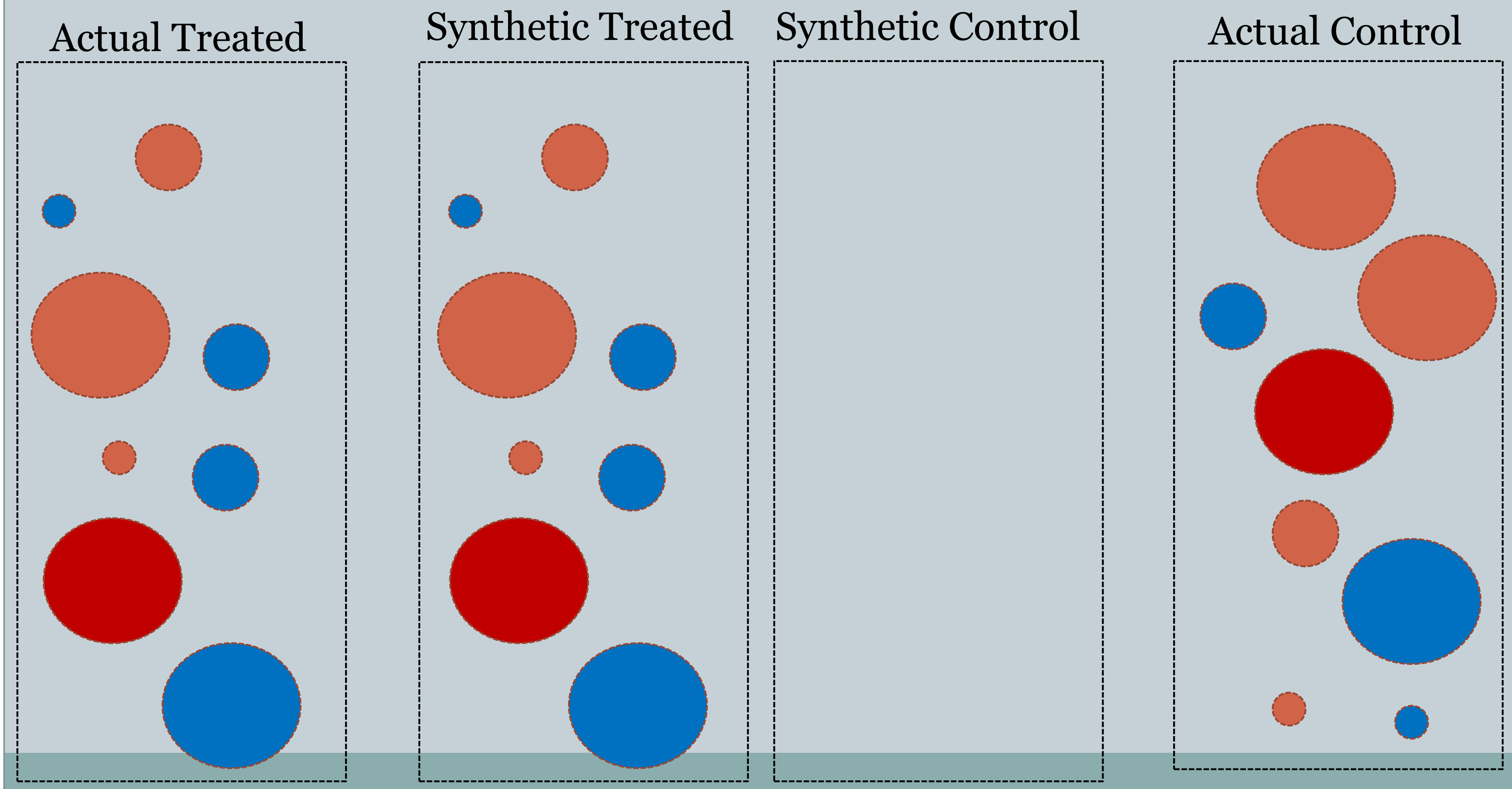
$$\frac{1}{n}\sum_{i=1}^{n}\frac{Z_iY_i}{e_i}-\frac{1}{n}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i}{(1-e_i)}$$

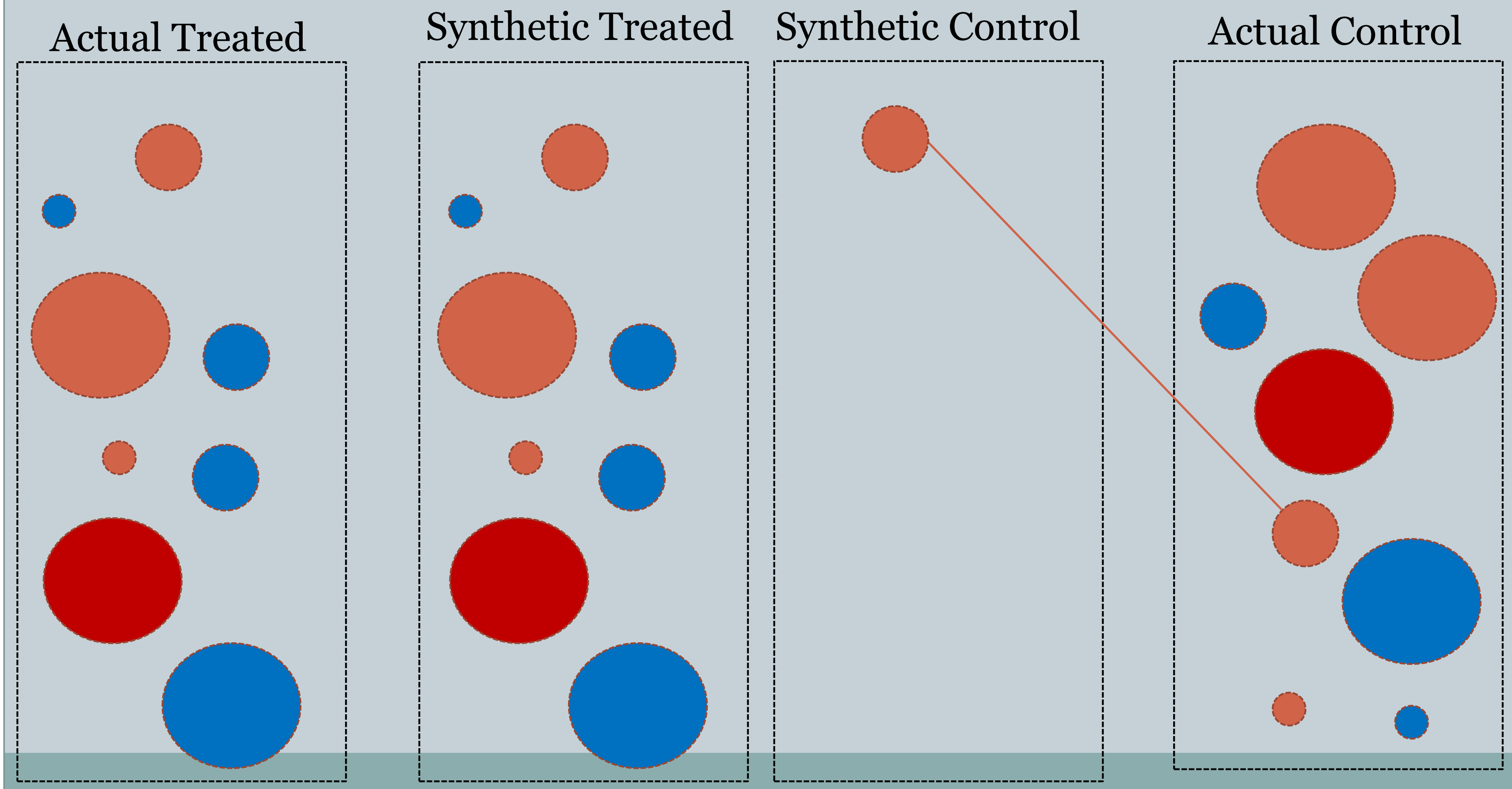
- An improved estimator is:

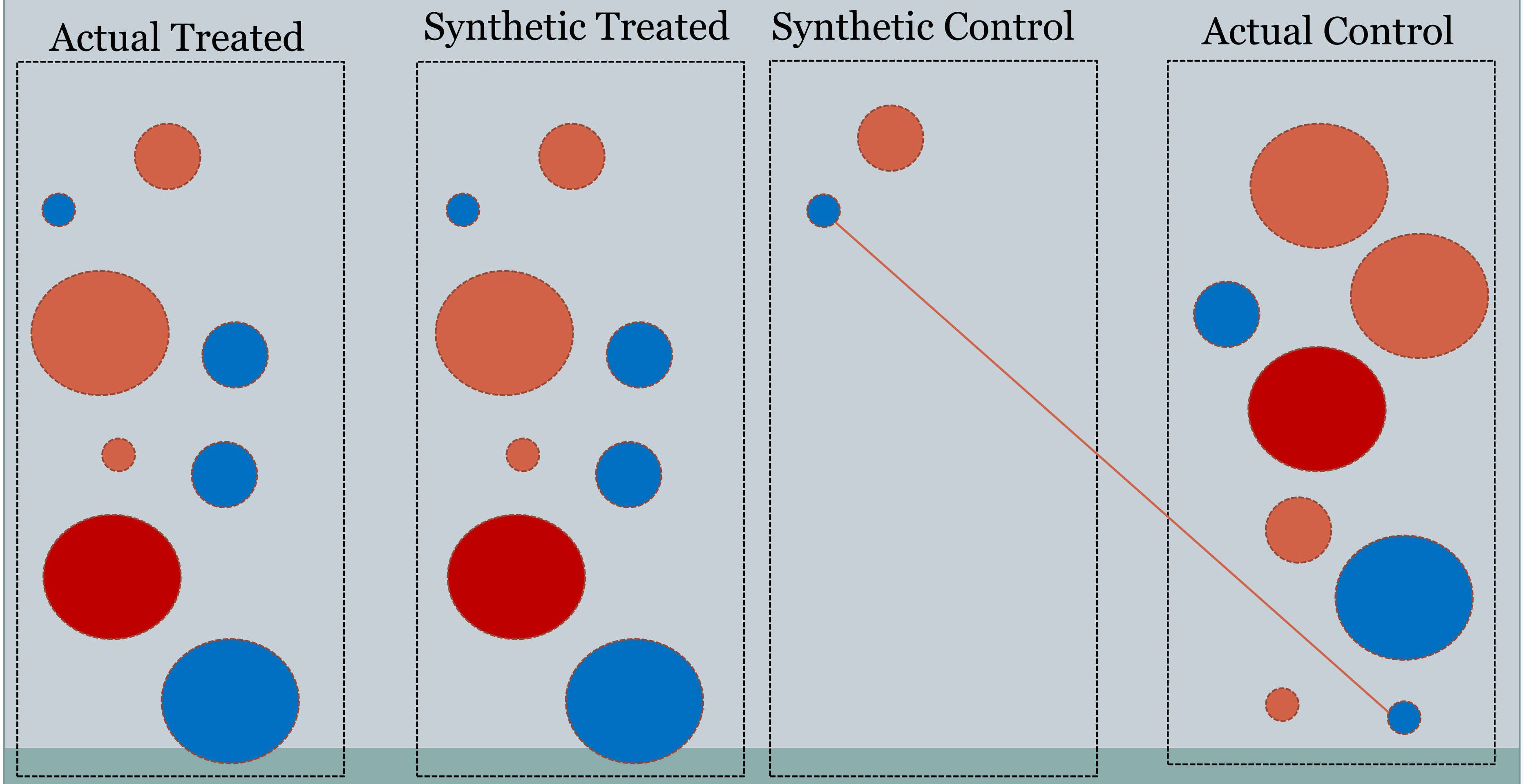
$$\left(\sum_{i=1}^{n}\frac{Z_i}{e_i}\right)^{-1}\sum_{i=1}^{n}\frac{Z_iY_i}{e_i}-\left(\sum_{i=1}^{n}\frac{(1-Z_i)}{(1-e_i)}\right)^{-1}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i}{(1-e_i)}$$

which uses an estimated number of observations, based on the pscore.

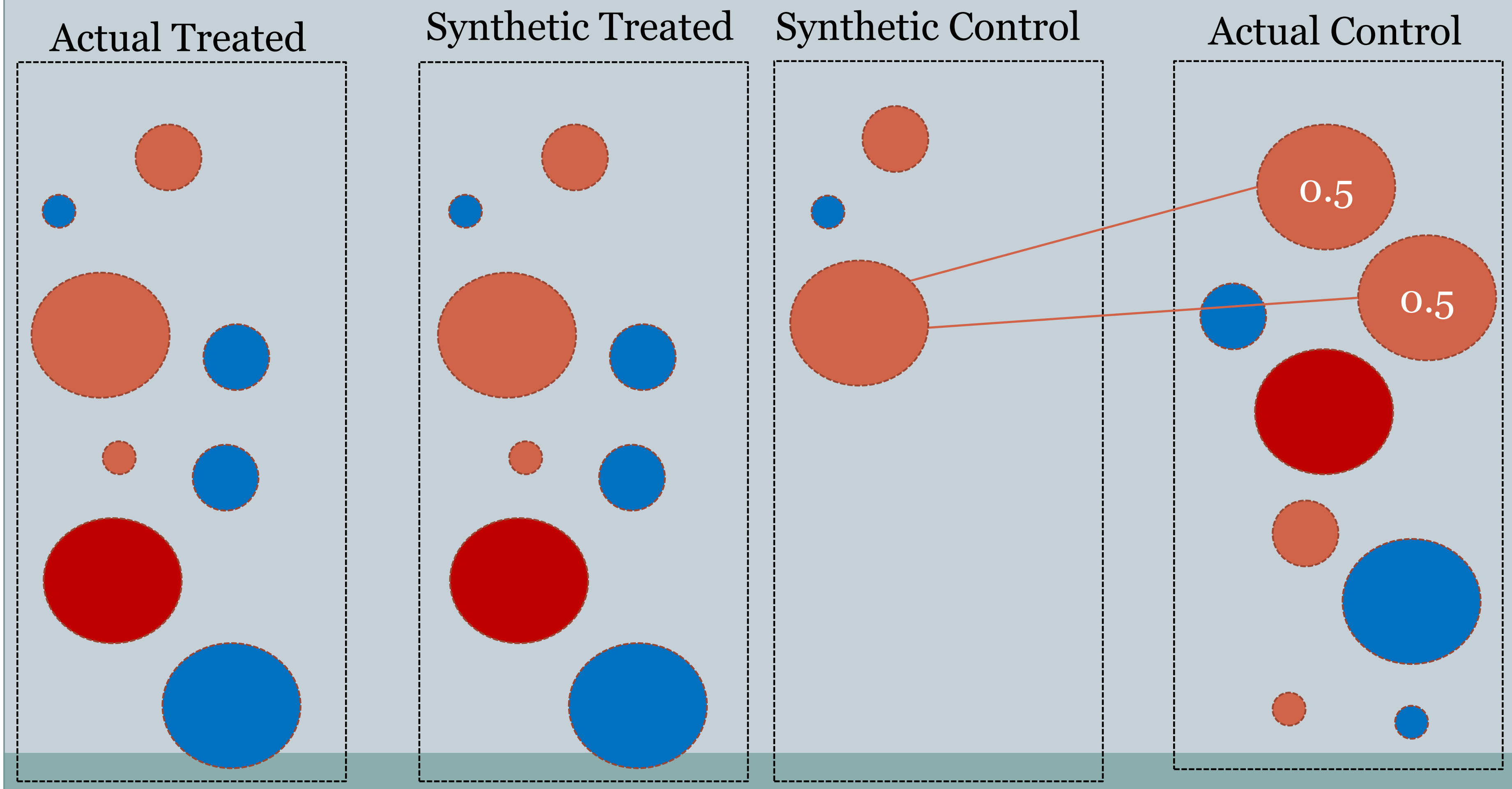

Actual Treated
Synthetic Treated
Synthetic Control
Actual Control


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control

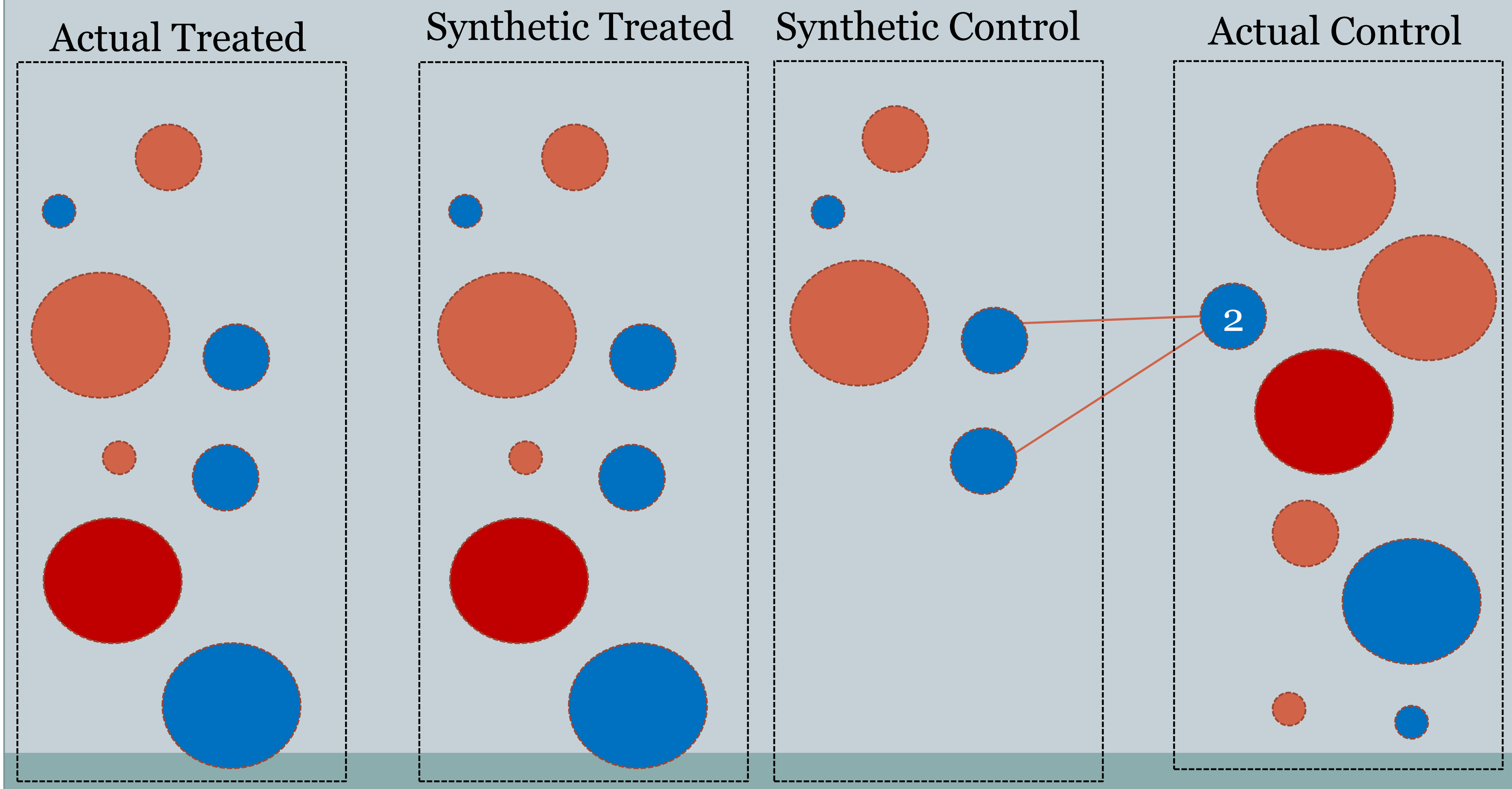

Actual Treated
Synthetic Treated
Synthetic Control
Actual Control

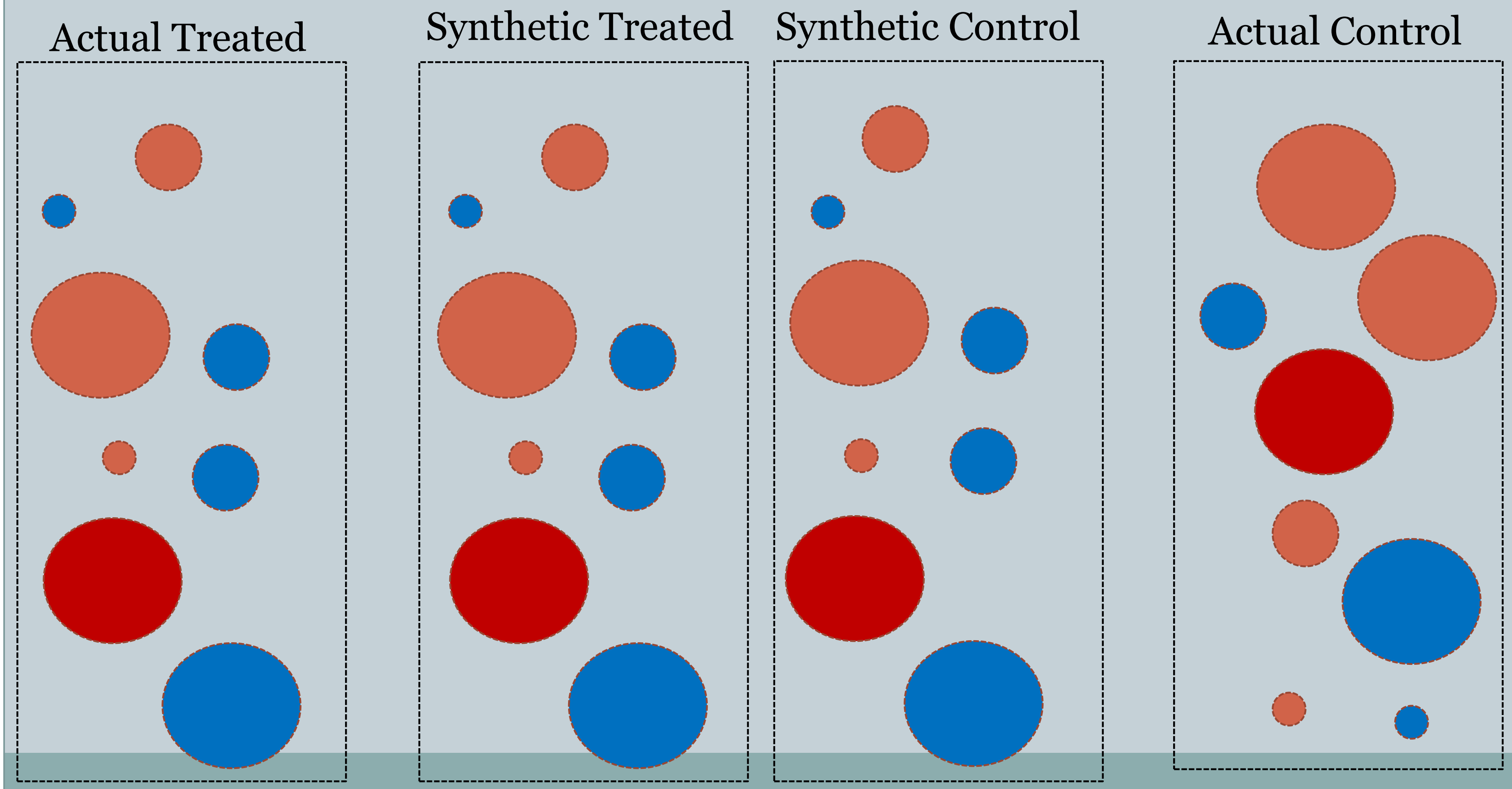


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control



Actual Treated
Synthetic Treated
Synthetic Control
Actual Control
0.5
0.5


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control
2


Actual Treated
Synthetic Treated
Synthetic Control
Actual Control

inverse-probability weights

- A gentle introduction to best practices here.
- Let Z denote treatment assignment.
- Let $e = \Pr(Z = 1|\boldsymbol{X})$; that is e is the propensity score.
- The inverse probability treatment weight is defined as:

$$w = \frac{Z}{e} + \frac{1 - Z}{1 - e}$$

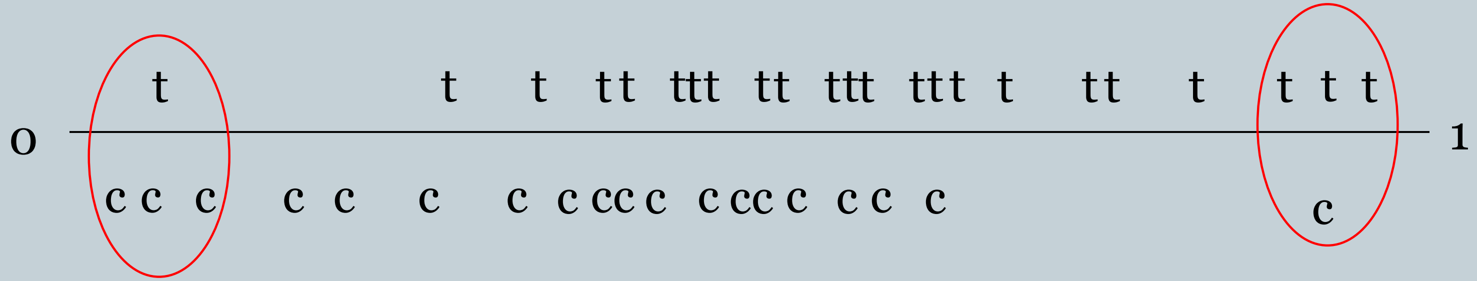
Small note: you can get the ATT or ATC with simple changes to the weights above.

IPW

THE ACHILLES HEEL

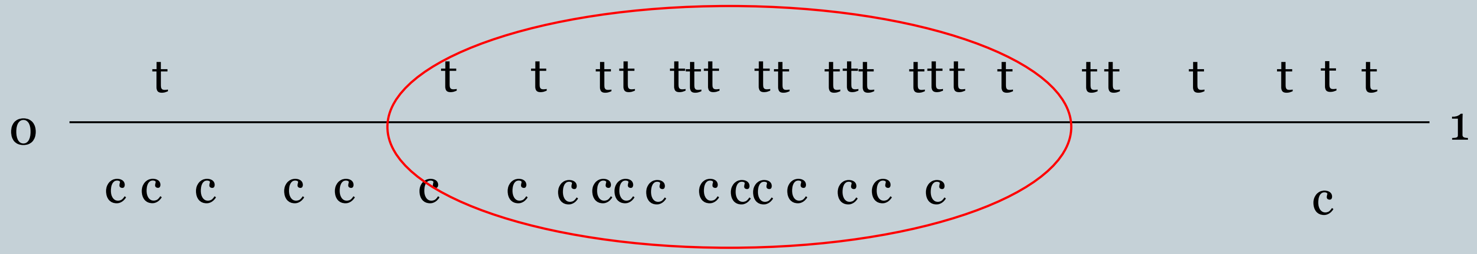
assumptions

- No unobserved confounders
- Possibly even exacerbates the problem in the tails



assumptions

- No unobserved confounders
- Possibly even exacerbates the problem in the tails



inverse-probability weighting

- The tails are often up-weighted heavily.
- The standard errors greatly impacted by these weights.
- These tails are the most problematic for strongly ignorable treatment assignment.

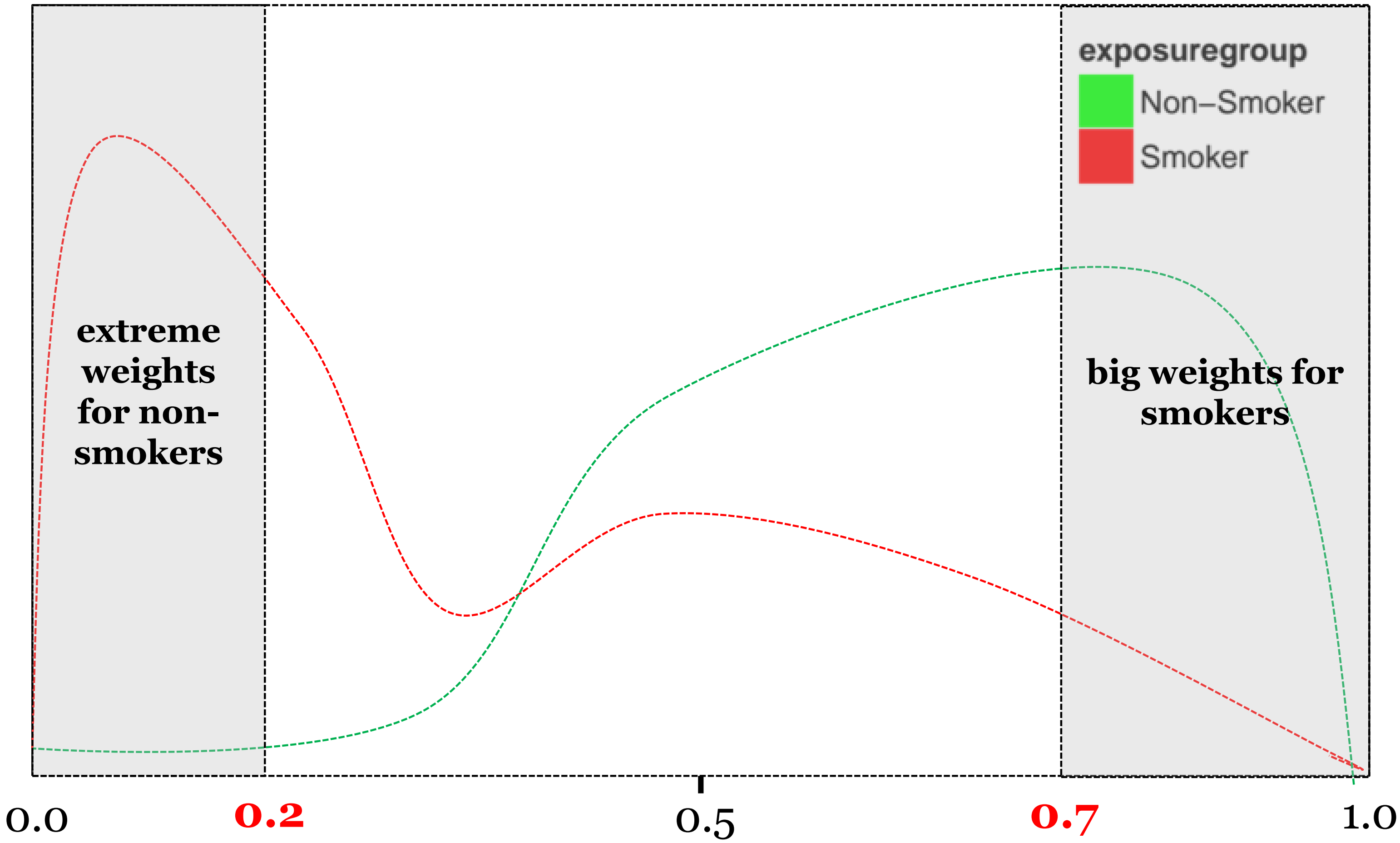
```
     t              t    t   tt  ttt  tt  ttt  ttt  t    tt    t    t  t  t
0  ———————————————————————————————————————————————————————————————————————————  1
   c c  c    c  c    c    c  c ccc c  c cc c  c c  c
                                                                      c
```

inverse-probability STABILIZED weights

- Let Z denote treatment assignment.
- Let $e = \Pr(Z = 1|\boldsymbol{X})$; that is e is the propensity score.
- The inverse probability stabilized treatment weight is defined as:

$$w = \frac{Z * \Pr(Z = 1)}{e} + \frac{(1 - Z) * \Pr(Z = 0)}{1 - e}$$

- $\Pr(Z = 1)$ and $\Pr(Z = 0)$ represent the marginal probabilities of treatment and control respectively, (i.e., unconditional probabilities).
- These “shrink” the values back to the overall rate.



P(non-S|**X**) = Probability of being a non-smoker, given covariates

extreme inverse-probability weights

- You can remove extreme values.
- Truncating is also an option:
  - All observations below 0.20 become 0.20.
  - All observations above 0.70 become 0.70.

inference: inverse-probability weights

- Interpretation: IPW creates artificial populations (e.g., treated and control) in which baseline covariates are independent of each other. (link)
- Instead of thinking of random assignment to treatment and control, IPW uses the analogy of sampling the observations from the treated population or control population.

inference: inverse-probability weights

- There are sandwich-type variance estimators that account for the estimated pscores. (link)
- Alternatively, people often use bootstrap based methods. (link – p 27)
- People also use models with population weightings.
  - In R, inside of `lm()` and `glm()`, there is a term `weights=` where you can place your estimated pscores.

fin.