Northwestern Medicine®
Feinberg School of Medicine

CONSTRAINED & ADAPTIVE RANDOMIZATION

What is it and how to I make it work with REDCap?

Statistically Speaking Seminar Series

March 9, 2023

Jacob M. Schauer, Department of Preventive Medicine – Biostatistics, Northwestern University

What is this talk about?

Randomized trials are referred to as a "gold standard" for generating rigorous scientific evidence…

- Why do we randomize (when it is ethical/feasible to randomize)?
  - Generate equivalent groups across conditions in a study/trial.
- What can go wrong with randomization?
  - Among other things, chance differences between groups.
- What can we do when randomization goes wrong (or we think it might)?
  - Constraining randomizations to "force" balance between study arms
    - Various approaches: stratified, covariate-constrained, and adaptive
    - Choice between them often depends on the study type and available data

Examples of ongoing randomized trials

Applications of constrained and adaptive randomization

- **NEED-PT**: Does embedding a PT in Emergency Department teams improve outcomes for patients with low back pain (i.e., less pain, less frequent opioid use, etc.)
  - Collaborators: Howard Kim (PI), Danielle McCarthy, Bruce Lambert, Amee Seitz, Kayla Muschong, Ann Kan, Jody Ciolino
  - Two-armed cluster randomized trial (anticipated N > 300)

- **2GETHER 3.0**: Does a tailored educational and counselling intervention reduce the risk of contracting HIV or other STI in MSM?
  - Collaborators: Michael Newcomb (PI), James Carey, Gregory Swann, Daniel Ryan, Cole Price, Jody Ciolino, Marc Broxton
  - Two-armed randomized trial (anticipated N > 5,000)

Randomize to get equivalent study arms

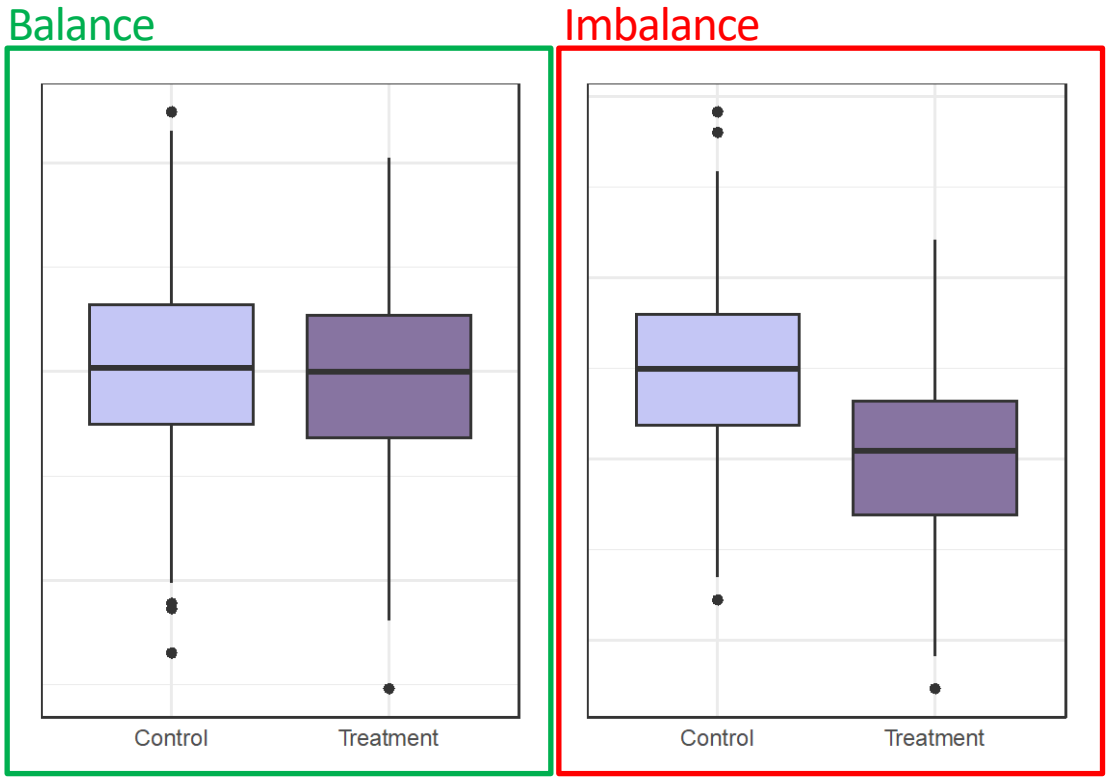
The general idea

- We want study arms to be equivalent on all measured and unmeasured **pre-randomization variables.**
  - At the very least, we'd want equivalence on *measured* pre-randomization variables, a.k.a. **'covariates'**.

Randomize to get equivalent study arms

The general idea

- We want experimental arms (treatment & control groups) to be similar on **pre-randomization variables** a.k.a. **'covariates'**.



Randomize to get equivalent study arms

The general idea

- We want experimental arms (treatment & control groups) to be similar on **pre-randomization variables** a.k.a. **'covariates'**.
- In individually randomized experiments, this means the randomized participants.
- In grouped randomized experiments (i.e., cluster randomized trials) we want the clusters *and* individuals within clusters to be similar.
  - These are similar when measures of the cluster are averages of measures taken on individuals (e.g., % female patients for each physician).
  - These can be distinct ideas, too.

Table 1: Covariate imbalance reporting

Point 1: Including a column for total controls shows distribution of characteristics in the source population. (EV)

Point 2: Stratifying controls by exposure shows potential confounding in the source population (e.g. by education or heart failure) by showing association with exposure in controls. (IV)

Point 3: No column with p-values, as statistical tests are not an appropriate method for assessment of confounding in exposed and unexposed controls, and similarity is not expected between cases and total controls.

Point 4: Showing variables both as collected (e.g. continuous age), and as analyzed (e.g. categorical age), can show potential for measurement error and residual confounding. (IV)

Point 5: To reduce visual clutter, show percentages rounded to nearest whole number, unless more precision is warranted.

Point 6: Including a selection variable (e.g. insurance status), even if not included in final analytic model, allows its distribution to be compared between cases and total controls to make judgements about whether cases reasonably arose from this source population. (IV)

Point 7: Show skewed continuous variables as median (min-max) or (25^{th} – 75^{th} percentile) instead of mean (SD). (IV, EV)

Point 8: Showing potential modifiers (e.g. hypertension), even if not included in the final analytic model, can help readers assess generalizability of findings. (EV)

Table 1. Characteristics of hemorrhagic stroke cases and controls, stratified by exposure (to represent distribution in the source population)

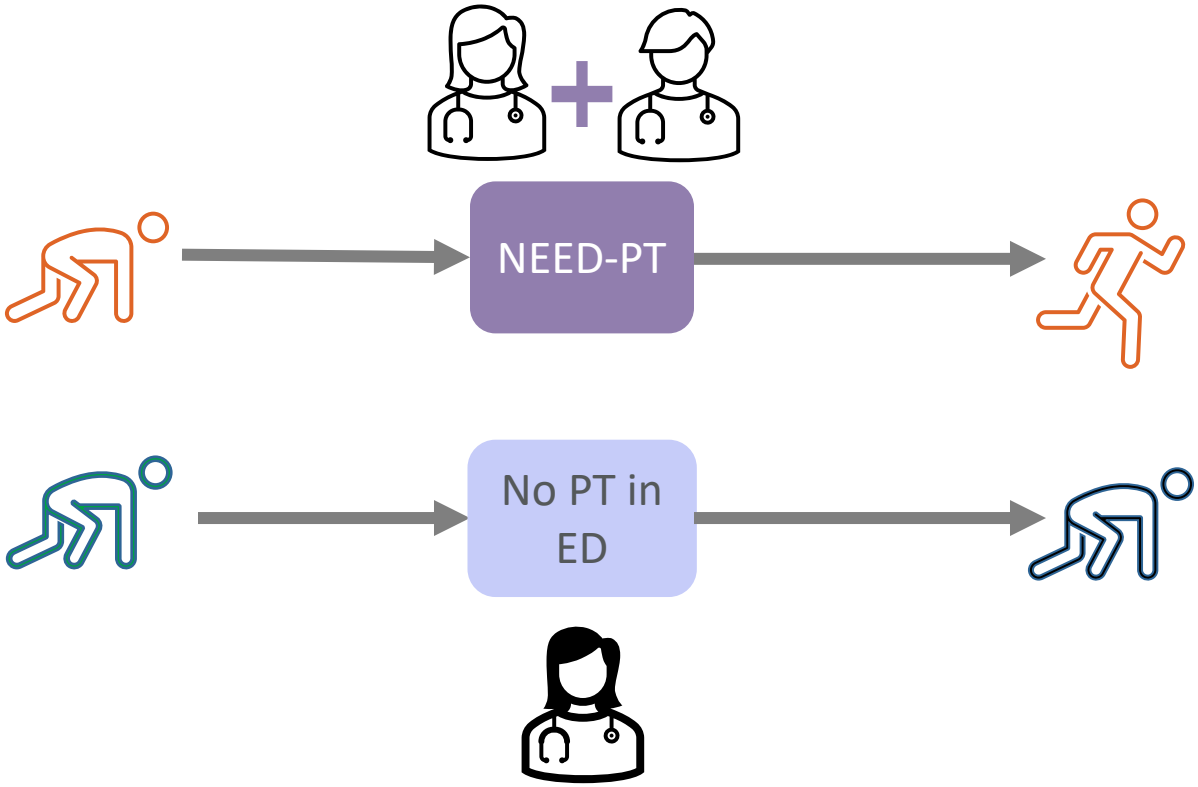
	Cases	Controls		
Sample characteristic[1]	(n=854)	*Total* (n=1708)	*Exposed* (n=332)	*Unexposed* (n=1376)
Diabetes (exposure)[2]	265 (31%)	332 (19%)	-- --	-- --
Demographics				
Male	325 (38%)	637 (37%)	73 (22%)	564 (41%)
Age, years (mean [SD])	69 (10.8)	63 (9.9)	64 (11.7)	63 (13.1)
Age, years				
18-40	34 (4%)	150 (9%)	27 (8%)	124 (9%)
41-60	111 (13%)	242 (14%)	50 (15%)	193 (14%)
61-80	589 (69%)	1244 (73%)	239 (72%)	1004 (73%)
81+	120 (14%)	72 (4%)	16 (5%)	55 (4%)
Education				
< High school	77 (9%)	165 (10%)	13 (4%)	151 (11%)
High school	325 (38%)	735 (43%)	116 (35%)	620 (45%)
Some college	367 (43%)	678 (40%)	170 (51%)	509 (37%)
>=College	85 (10%)	130 (8%)	33 (10%)	96 (7%)
Insurance status				
Public	486 (57%)	926 (54%)	195 (59%)	729 (53%)
Private	248 (29%)	567 (33%)	100 (30%)	468 (34%)
None	120 (14%)	215 (13%)	37 (11%)	179 (13%)
Personal medical history				
CCI, median (min-max)	5 (0-15)	2 (0-10)	3 (0-10)	0 (0-7)
Heart failure	453 (53%)	404 (24%)	60 (18%)	344 (25%)
Atrial fibrillation	265 (31%)	238 (14%)	73 (22%)	165 (12%)
Hypertension	290 (34%)	375 (22%)	86 (26%)	289 (21%)
Pharmacologic agent use				
Sulfonylureas	538 (63%)	692 (40%)	183 (55%)	509 (37%)
Vasodilators	154 (18%)	195 (11%)	43 (13%)	151 (11%)
Diuretics	461 (54%)	728 (43%)	136 (41%)	592 (43%)
Beta blockers	239 (28%)	416 (24%)	113 (34%)	303 (22%)
Statins	325 (38%)	453 (27%)	123 (37%)	330 (24%)
NSAIDs	376 (44%)	731 (43%)	139 (42%)	592 (43%)

[1]Variable distributions are reported as n (%) unless otherwise specified.
[2]Exposure distribution not reported for strata defined by exposure status.
Abbreviations: CCI, Charlson Comorbidity Index; min, minimum; max, maximum; NSAID, non-steroidal anti-inflammatory drug; SD, standard deviation.

Hayes-Larson et al., 2019

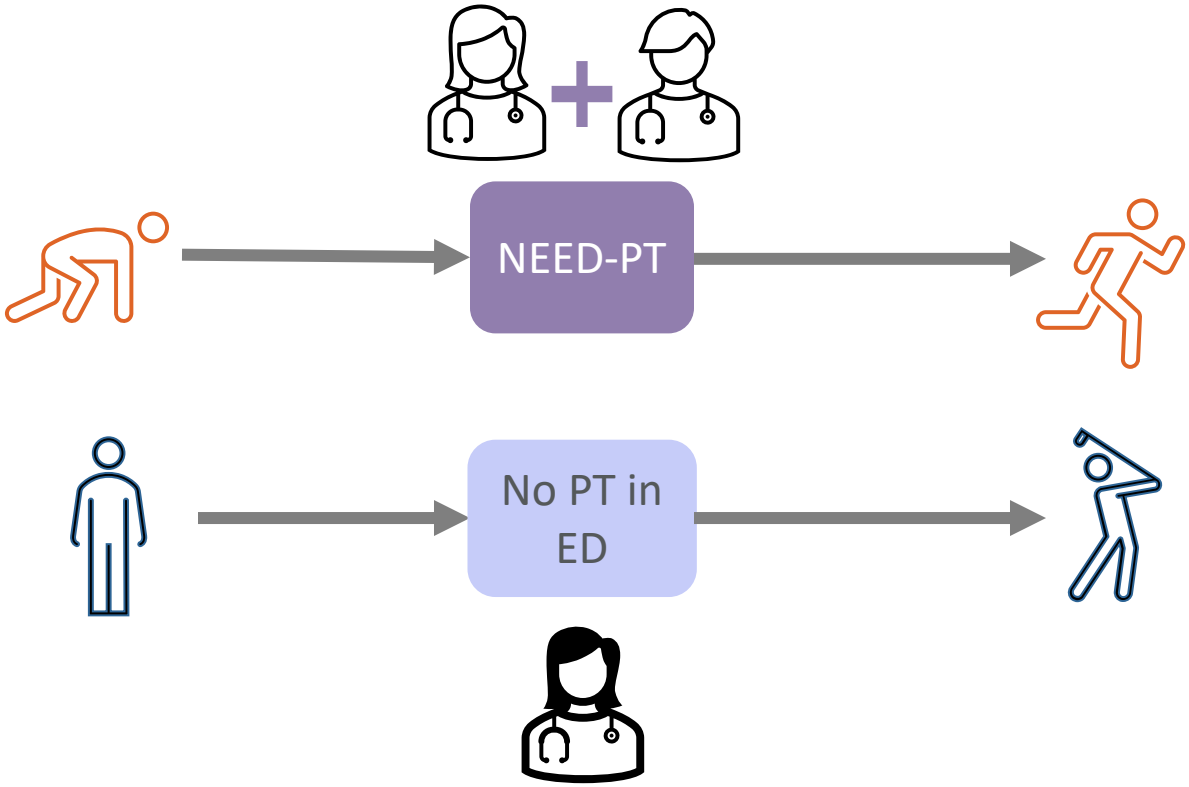
Face validity: comparing similar groups is important

NEED-PT Example



Face validity

NEED-PT Example



Balance is statistically important

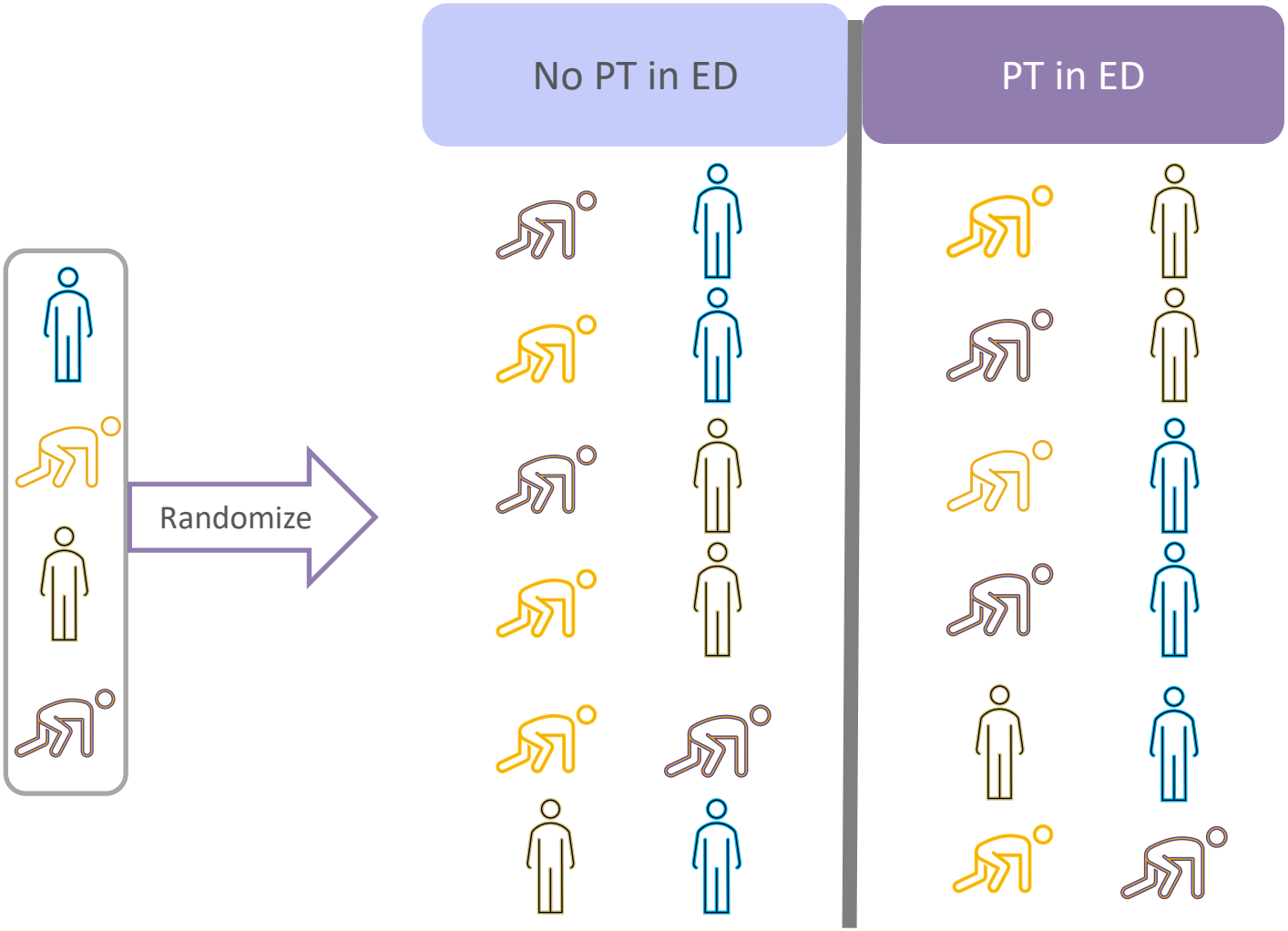
- ***Unbiased*** estimation of causal estimates assume treatment assignment is *independent* of both
  - **measured (observed)** participant attributes $X = [X_1, \ldots, X_p]$, and
  - **unmeasured (unobserved)** participant attributes $W = [W_1, \ldots, W_p]$
- Imbalance can reduce the power of statistical analysis.
- Imbalance (not properly adjusted for) can bias conditional power analysis or potential study stopping criteria (Ciolino et al., 2011)

Northwestern Medicine®
Feinberg School of Medicine

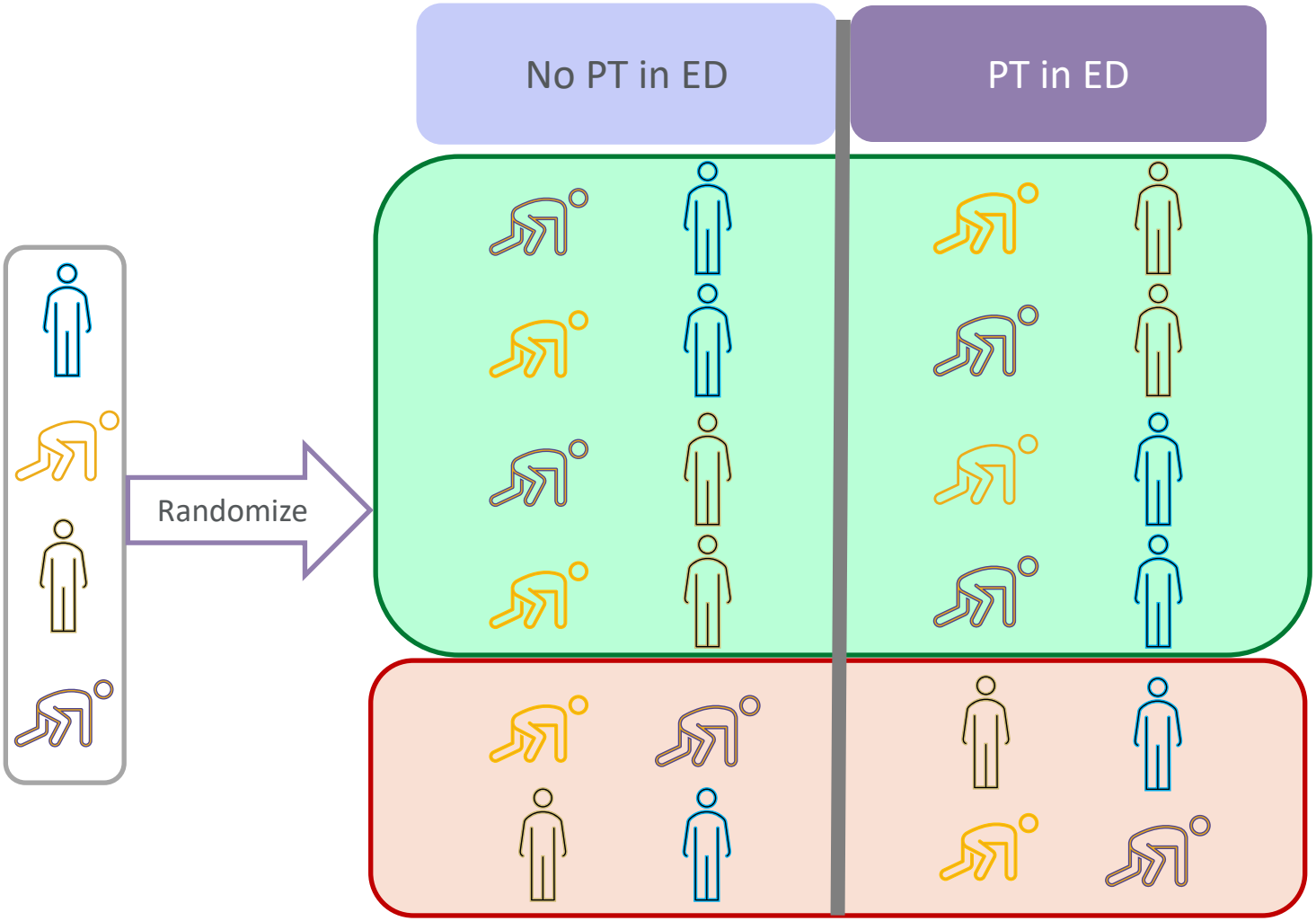
What can go wrong?

Who's afraid of a bad randomization?

Equivalent groups to answer causal questions



Equivalent groups to answer causal questions



What do we mean by **equivalent/imbalanced?**

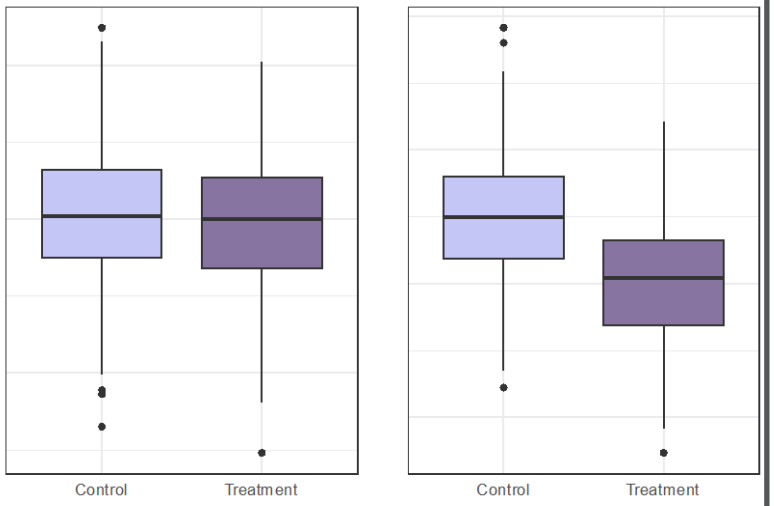
Multiple covariates to consider

- Typically, we want our study arms to be "the same" on average for **several** baseline variables/covariates:
  - Baseline measure of outcome (e.g., pain, opioid use, presence of STI)
  - Patient characteristics (e.g., age, height, weight, sex, relationship status etc.)

What do we mean by **equivalent/imbalanced?**

Metrics of imbalance

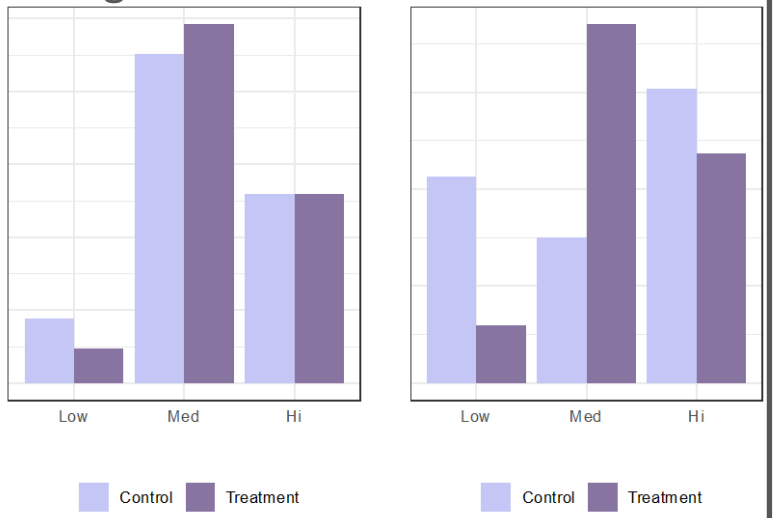
Continuous covariates



Raw mean differences $\bar{X}_T - \bar{X}_C$

Standardized mean differences $\frac{\bar{X}_T - \bar{X}_C}{\sigma_X}$

Categorical covariates



Raw differences in proportions

Odds ratios/risk ratios by category

- Difference in group averages or proportions should close to zero.
- Some use hypothesis tests to evaluate differences between groups (but that's not a good idea).

When does randomization not give **equivalent groups?**

- **Smaller sample sizes:** Theoretical discussion suggests that larger sample sizes (N>200-300) are less likely to result in imbalance than randomization with smaller sample sizes (Altman, 1985; Senn, 1989; Pocock et al., 2002).

- **Large # of covariates:** Chance imbalance is likely to happen if we consider larger numbers of baseline measures/covariates.

# of Covariates	Probability of "significant" imbalance on ≥1 covariate under random assignment
2	9.75%
5	22.6%
10	40.1%
20	64.1%
50	92.3%

What can you do if you get a "bad" randomization?

Depends on your study…

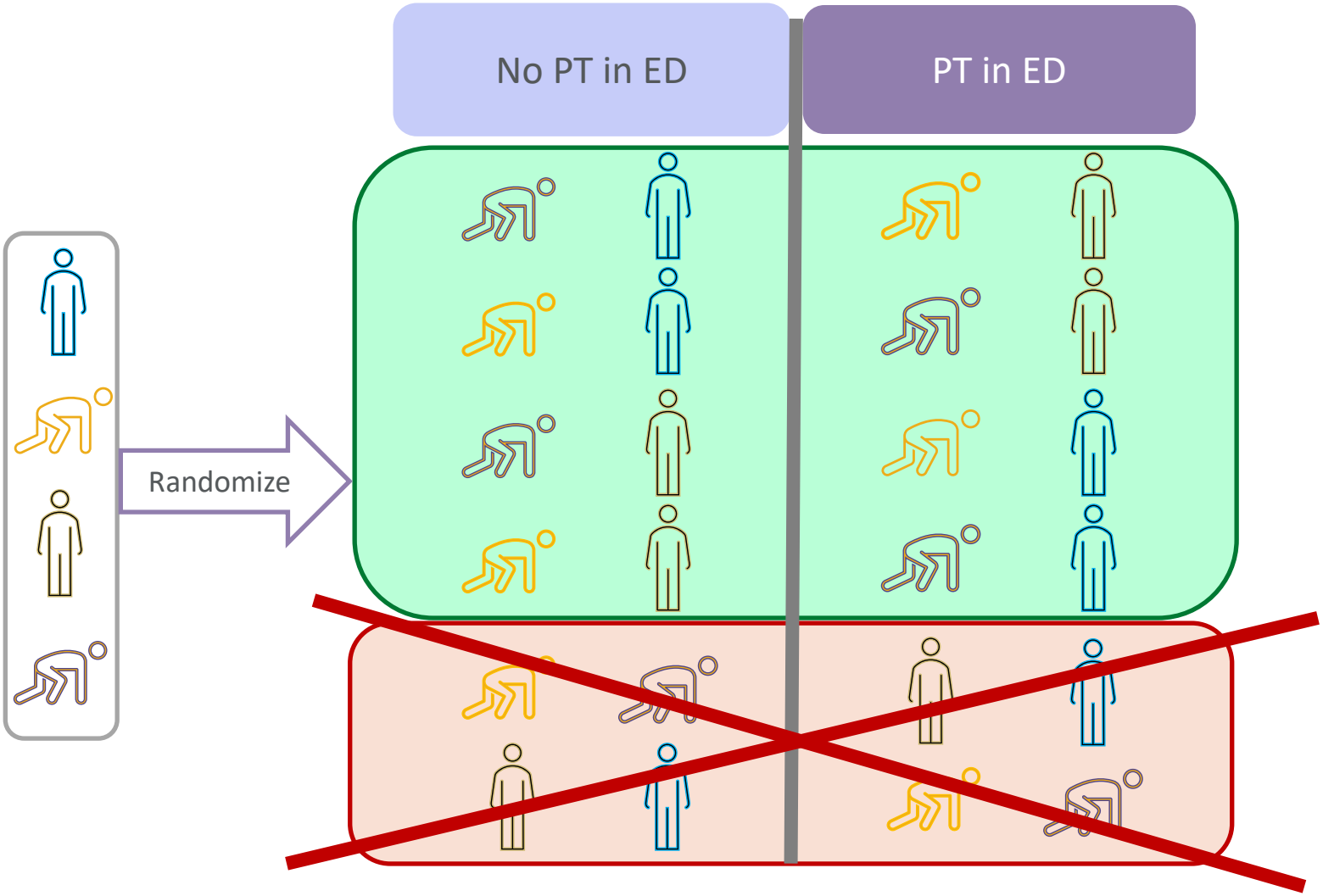
- If you have all covariate data prior to randomization (and can check balance immediately after randomization), then you can re-randomize if metrics for "bad" randomization are defined *a priori* (Morgan & Rubin, 2012).
  - Proponents: Fisher, Neyman, Gossett, Yates, Savage, Cochran, Rubin, …
- If patients arrive sequentially over time and are randomized, you may not know how bad your randomization is until its too late!
  - In that case, you ***hope*** that statistical adjustments can fix any induced bias.

Northwestern Medicine®
Feinberg School of Medicine

Constraining randomization

I'm afraid of a bad randomization…

Constrained randomization



Constrained randomization

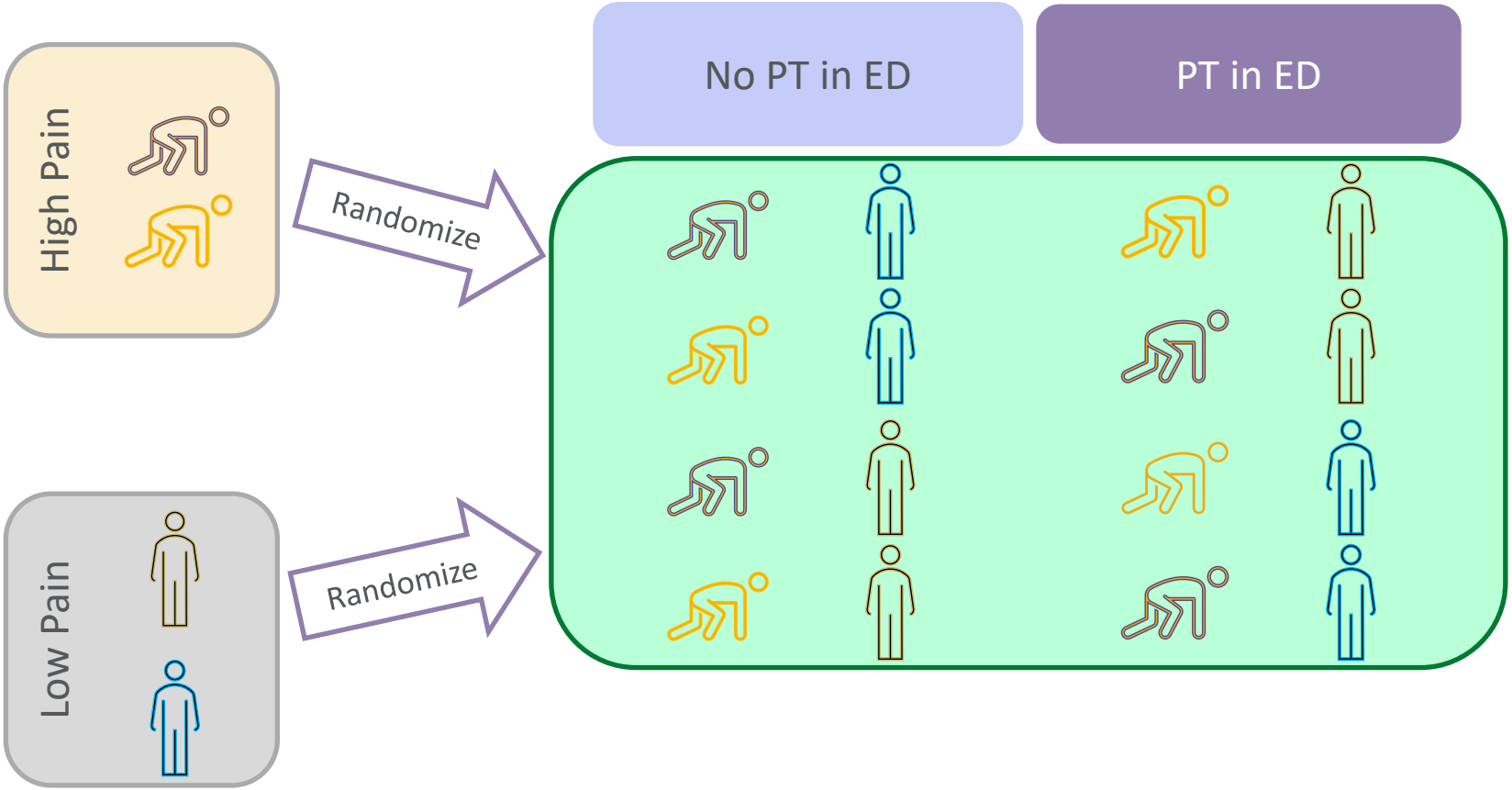
The general idea

- We want covariates to have similar distributions in each study arm.
- Idea: Why don't we only select from randomization where that happens?
  - Which covariates?
  - How *similar* and on what metric?
  - When do you measure covariates?
  - When can/should patients be randomized?

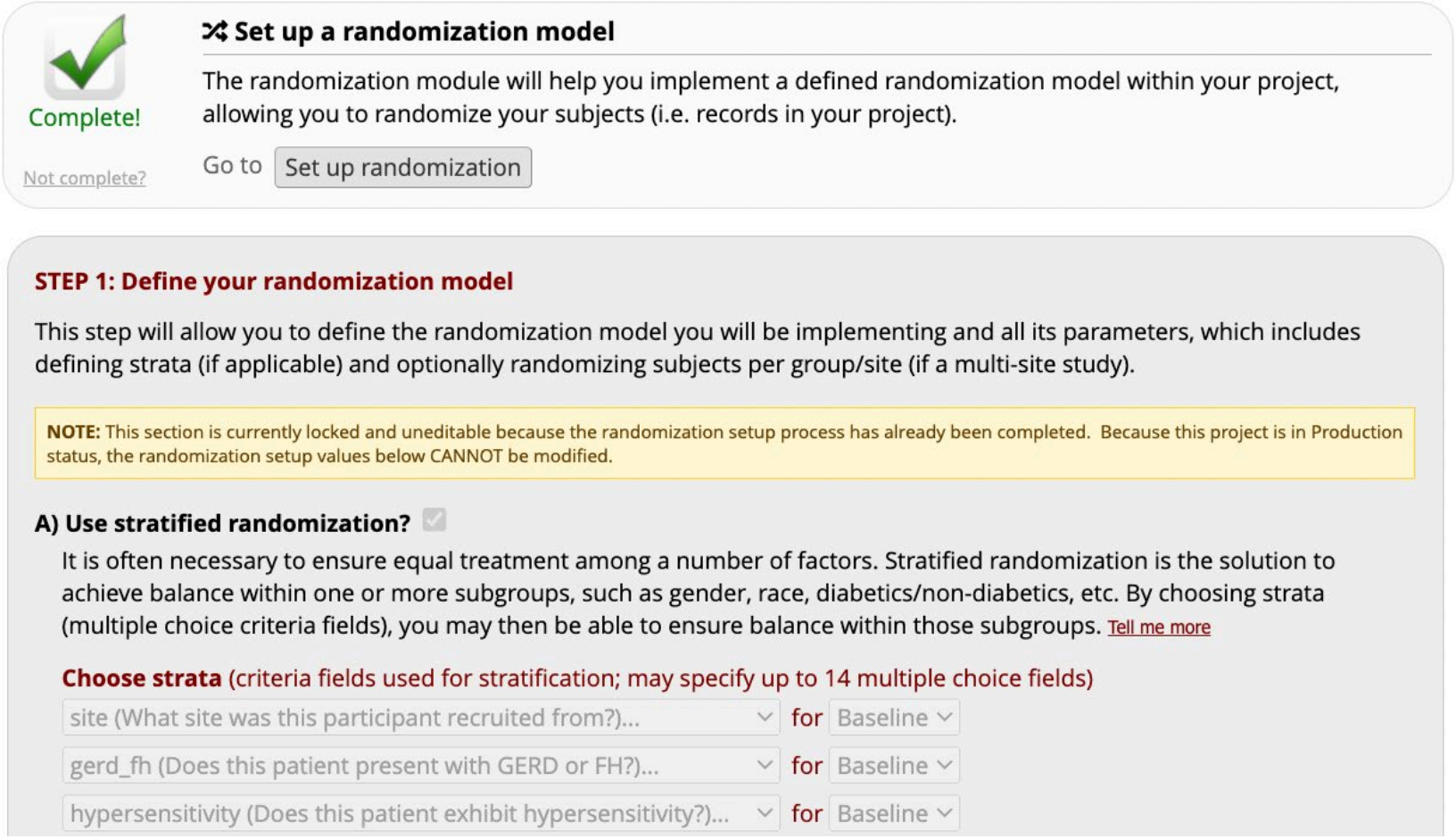
Northwestern Medicine®
Feinberg School of Medicine

Stratified randomization

Stratified randomization (blocking)



Stratified randomization in REDCap



Stratified randomization (blocking)

- Considerations:
  - Strata must have size N≥2 (ideally more unless you're doing direct matching)
  - Stratification variables can be **measured prior to randomization**
  - Analysis can (but doesn't have to) adjust for stratum.
    - Can also include stratum-by-treatment interaction.

Benefits:

1. Relatively simple
2. Allows for evaluation of effect heterogeneity by stratum
3. Can improve precision

Drawbacks:

1. Limited in the number and type of variables
   a. Discretize continuous variables (e.g., low-med-hi pain score)
   b. Example: 4 variables each with 3 levels = 3^4 = 81 strata!

Northwestern Medicine®
Feinberg School of Medicine

Covariate-constrained randomization

Useful if you have data on all units prior to randomization

Covariate-constrained randomization

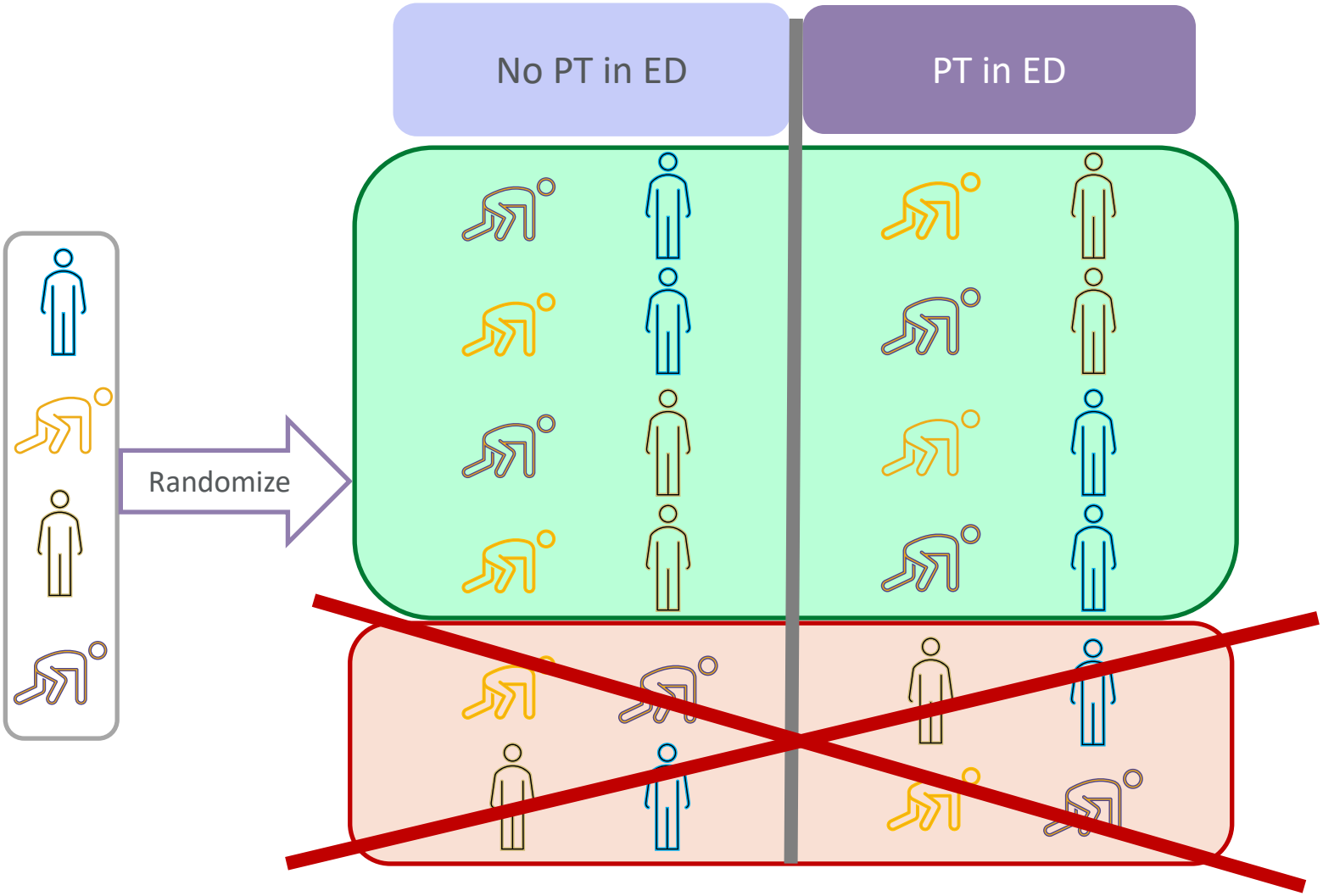
General approach

1. Enumerate all possible randomizations.
   - For large samples, this may be computationally intractable, so we simulate a large number (>1,000) of possible randomizations.
2. Compute balance/differences between study arms for each covariate of interest.
3. Identify randomization allocations with adequate balance (small differences).
   - Easier said than done.
4. Select one of those allocations at random.

Note that this is similar to the idea of "re-randomizing" if you get a bad randomization (Morgan & Rubin, 2012)

(Raab & Butcher, 2001; Moulton, 2004)

Covariate-constrained randomization



Covariate-constrained randomization

Considerations

- Defining "adequate" balance should be done *a priori*.
- Careful attention should be paid to *which* covariates you consider. (Li et al., 2016)
- Analysis is subject to minor statistical disagreement (Li et al., 2016):
  - Permutation tests match the process of data generation and are suggested as precise analyses.
  - Statistical results suggest model-based estimation (ANOVA, t-tests, etc.) is unbiased and sacrifices only a minor precision in large sample sizes.
- Only feasible if you have relevant information on **all units to be randomized** prior to randomization.
  - This is somewhat rare in individual randomized trials but is more common in cluster randomized trials.
  - CCR is better than SR even if information isn't 100% accurate (Organ et al., 2021)
- In cluster randomized trials, it ensures randomized clusters are comparable but not necessarily individuals within clusters.

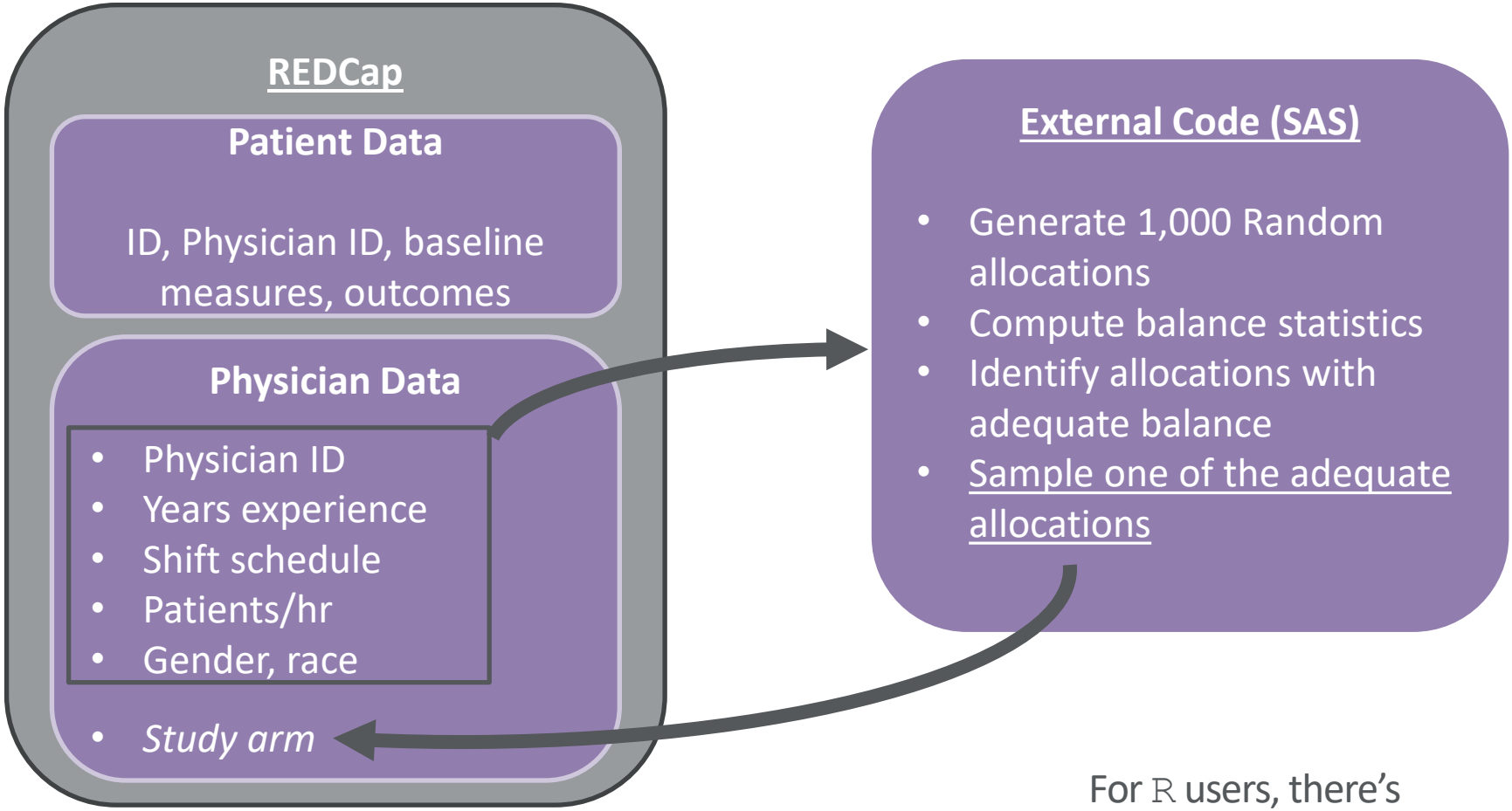
Covariate-constrained randomization

Key considerations: Assessing and quantifying balance

- Defining "adequate" balance should be done *a priori*.
- Strategies:
  - Variable-by-variable: "allowable" differences between study arms for each metric (Moulton, 2004).
    - Difference in physician experience should be <2 years, day shifts <1.
  - Composite imbalance metrics (Raab & Butcher, 2001)
    - $B_{L2}: \sum_j w_j \left(\bar{X}_{jT} - \bar{X}_{jC}\right)^2$
- Trade-offs:
  - The tighter the constraints, fewer allocations will be deemed adequate.
  - Implications of over-constraining (e.g., so there's only 1 adequate allocation) are not well-known.
    - Simulations suggest that even if permissible allocations make up just 1% of the total possible randomizations, there is limited impact on analysis (Morgan & Rubin, 2012; Li et al., 2016).

Covariate-constrained randomization

Not *really* in REDCap, but we can make it work



For `R` users, there's `cvcrand`!

Covariate-constrained randomization

Balance without constraints

Mean and percent differences between groups in simulated allocations under no, mild, and heavy randomization constraints.

	No Constraint (N=1,000)		Mild Constraint (115/1,000)		Heavy Constraint (16/1,000)	
Variable	75%	Max	75%	Max	75%	Max
Experience (years)	3.39	9	2.65	7.09	0.74	0.83
Orange Zone Shifts/Mo.	0.83	1.91	0.16	0.39	0.13	0.22
Patients/Hr.	0.13	0.36	0.08	0.22	0.06	0.10
Day Shifts/Mo.	0.81	2.43	0.38	0.85	0.25	0.35
Race: Minority*	4	8	2	6	2	2
Sex: Male*	3	13	3	5	3	3

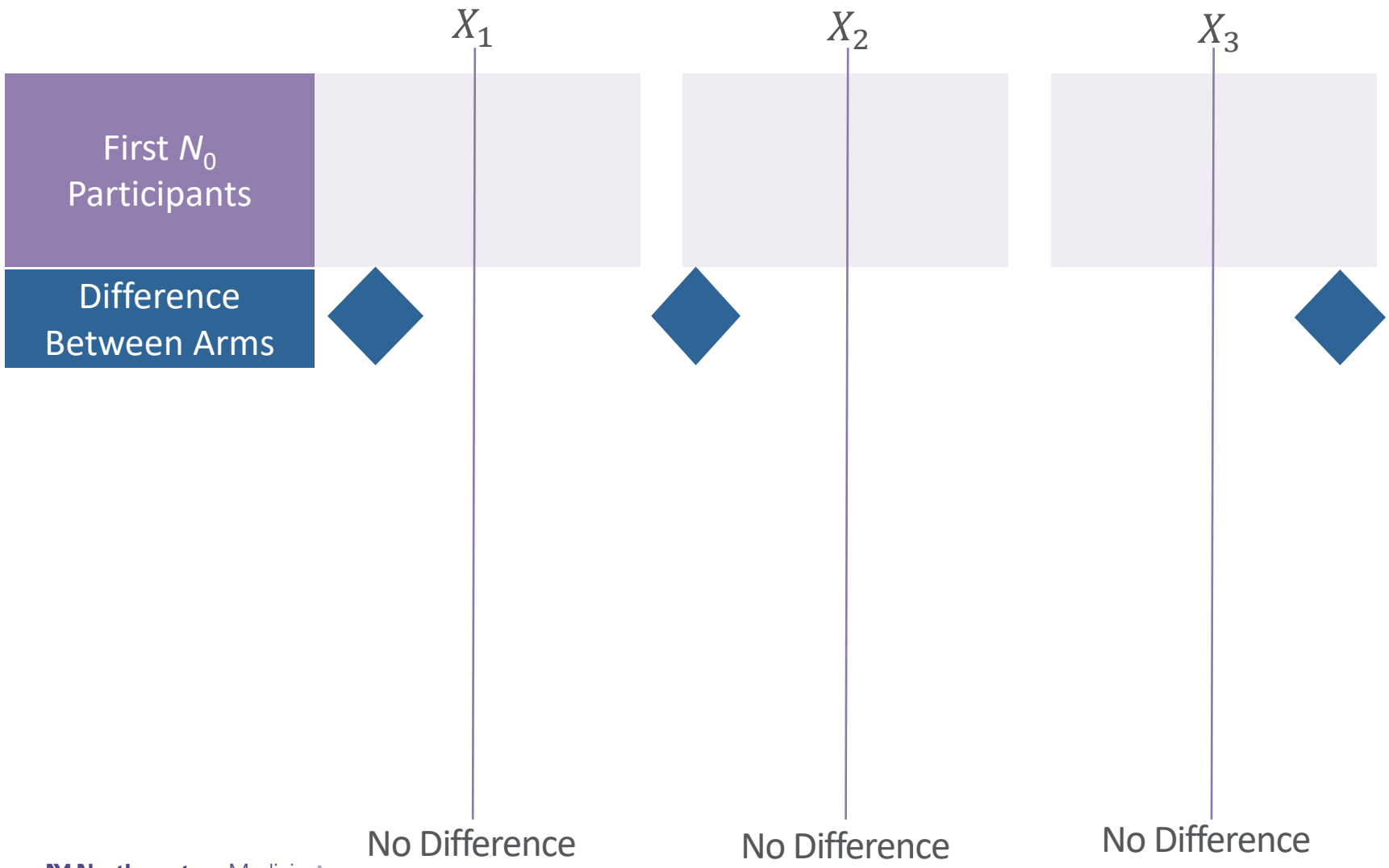
* Values indicate raw count imbalance (e.g., difference in # of male physicians across arms).

Adaptive randomization

Useful if participants arrive and are randomized sequentially over time

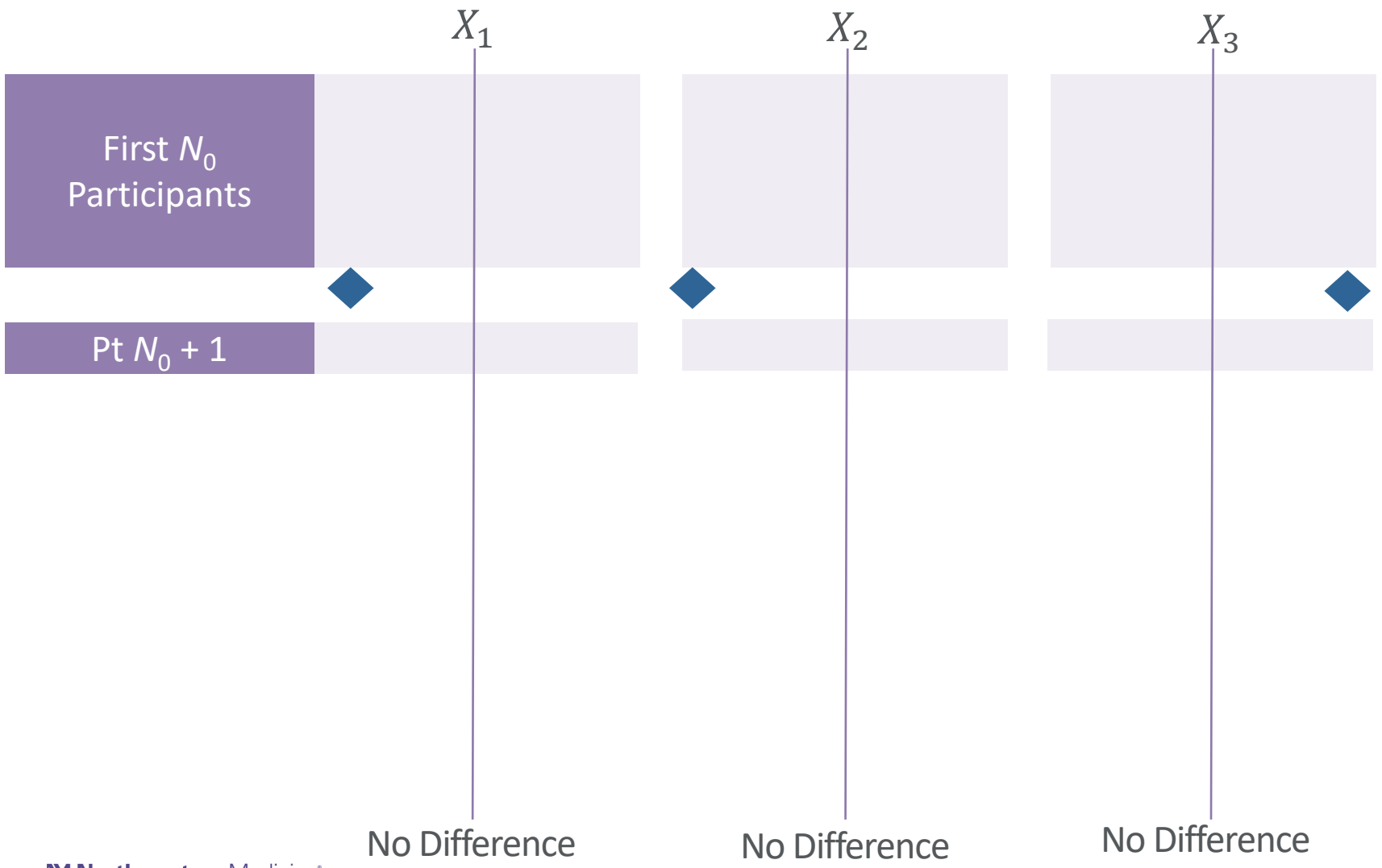
Adaptive randomization

Participants arriving sequentially over time



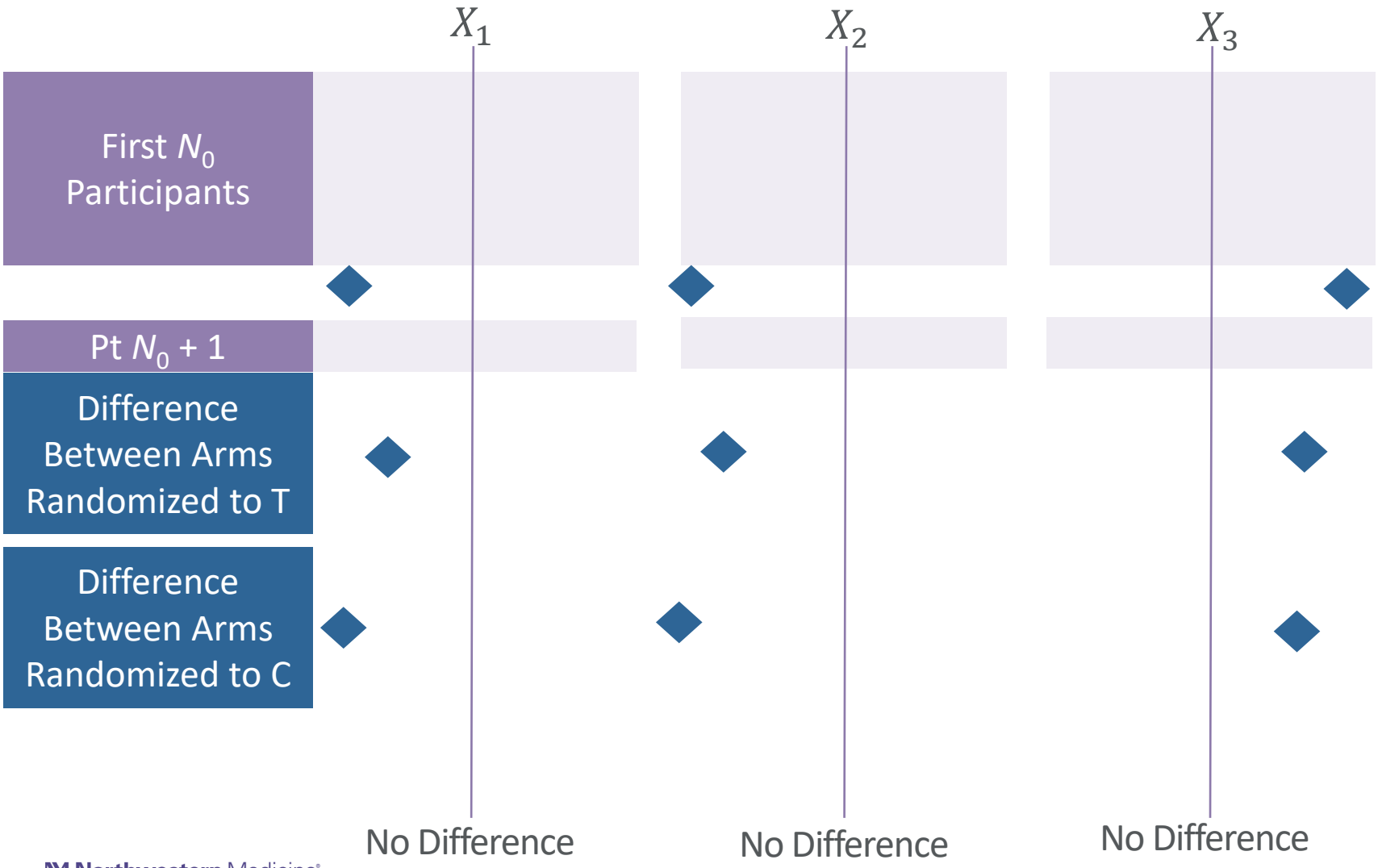
Adaptive randomization

Participants arriving sequentially over time



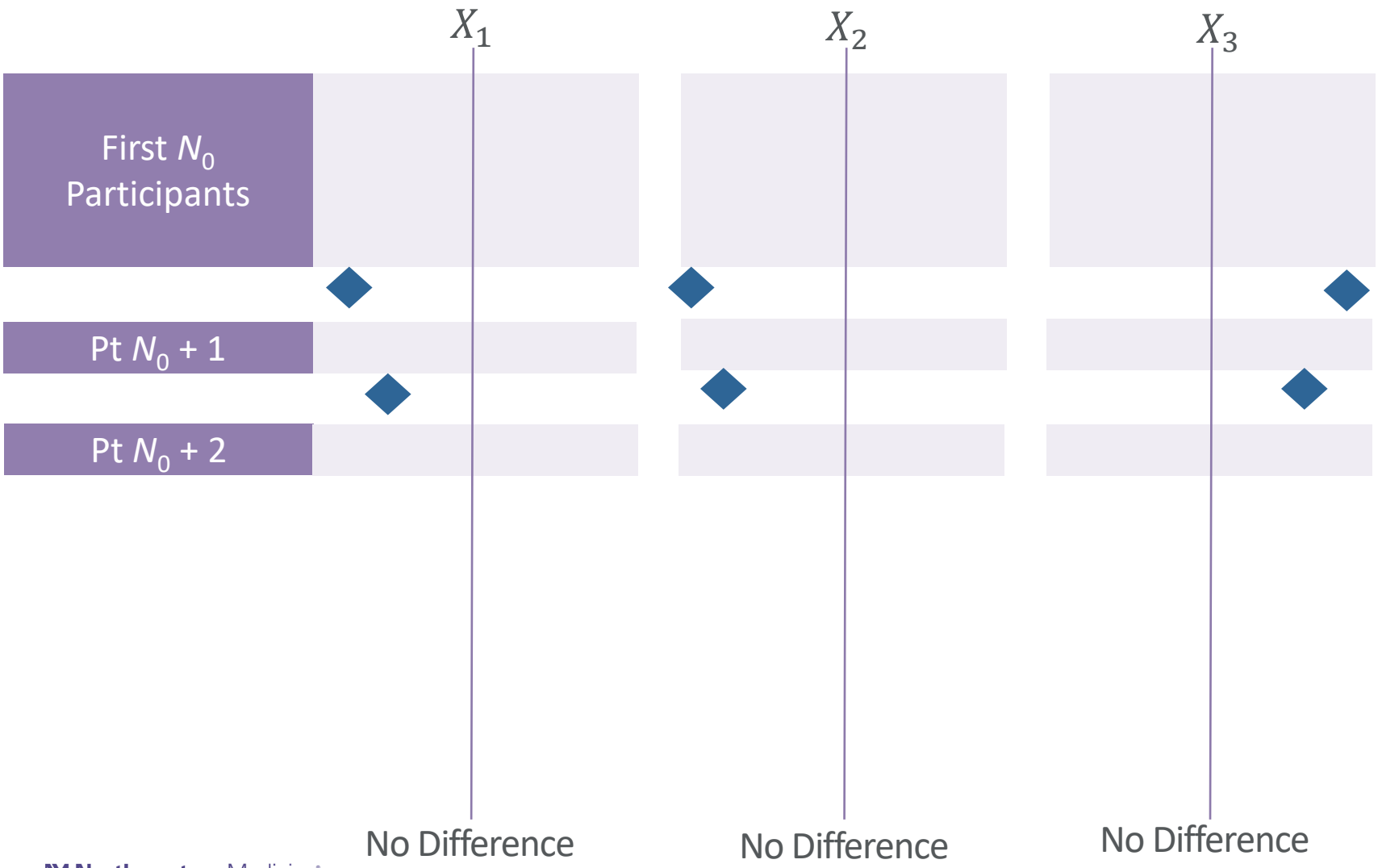
Adaptive randomization

Participants arriving sequentially over time



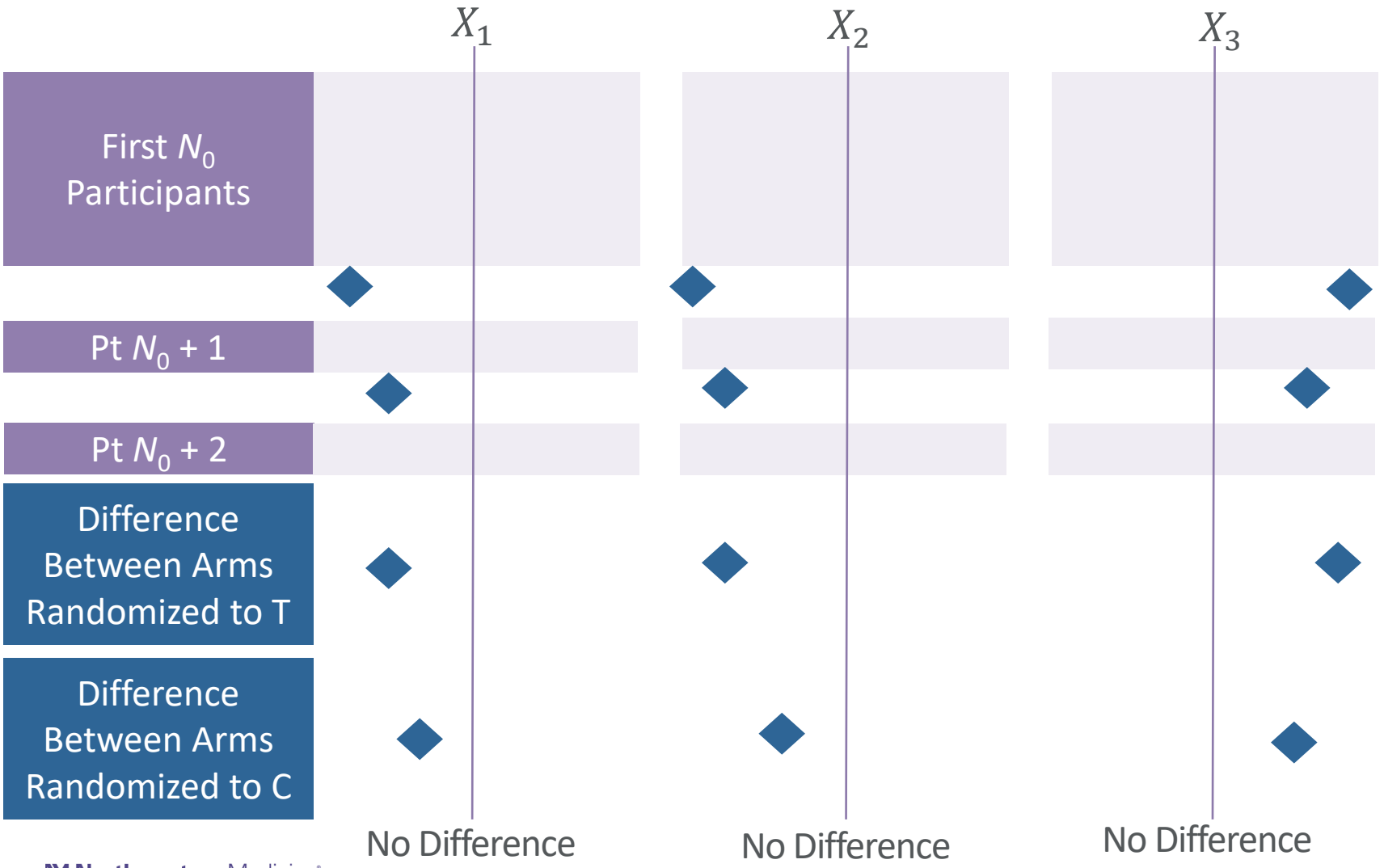
Adaptive randomization

Participants arriving sequentially over time



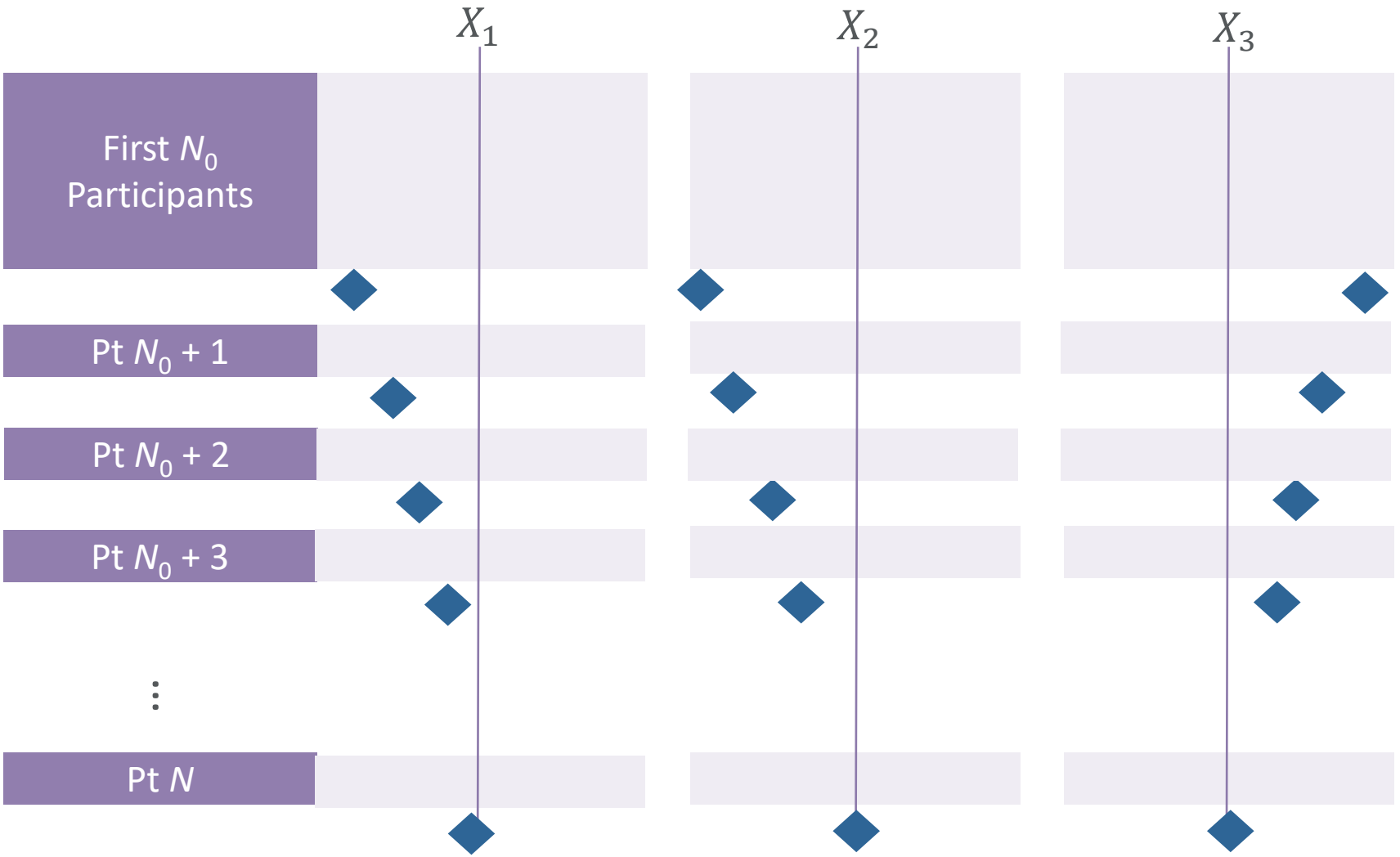
Adaptive randomization

Participants arriving sequentially over time



Adaptive randomization

Participants arriving sequentially over time



Adaptive randomization algorithms

- Minimization (Taves, 1974; Pocock & Simon, 1975)
- Optimal Biased Coin Randomization (Atkinson, 1982)
- Optimal Biased Coin with Covariates Randomization (Atkinson, 1999)
- **Minimal Sufficient Balance** (Zhao, Hill, & Palesch, 2015)
- Common-Scale Minimal Sufficient Balance (Johns et al., 2022)

Minimal Sufficient Balance

Algorithm

1. Enroll N_0 participants.
2. Randomize with equal probability for each arm (50%)
3. New participant *i* with measures X_{ij} on covariates $X_1, \ldots, X_p$
4. For each covariate X_j, we compute a "vote" for whether balance would be improved if participant *i* were in **T** or **C**
   a. If balance is not improved regardless of assignment, vote "neutral"
5. Count votes across covariates.
6. Randomize as follows:

$$P[Assign\ person\ i\ to\ T] = \begin{cases} \pi & if\ T\ gets\ the\ most\ votes \\ 1 - \pi & if\ C\ gets\ the\ most\ votes \\ 0.5 & otherwise. \end{cases}$$

Minimal Sufficient Balance

Computing the "vote" for a covariate

Continuous covariates

1. Conduct a *t*-test (or U-test) for balance on X_j for participants already randomized.
   a. Let S_j be the test statistic and p_j be the *p*-value of the test.
2. If $p_j < \alpha$:
   a. Vote for **T** if [$\bar{X}_{jC} > \bar{X}_{jT}$ and $X_{ij} > \bar{X}_{jC}$] OR [$\bar{X}_{jC} < \bar{X}_{jT}$ and $X_{ij} < \bar{X}_{jC}$]
   b. Vote for **C** if [$\bar{X}_{jC} < \bar{X}_{jT}$ and $X_{ij} > \bar{X}_{jT}$] OR [$\bar{X}_{jC} > \bar{X}_{jT}$ and $X_{ij} < \bar{X}_{jT}$]
3. Vote "Neutral" otherwise

For each covariate you get a vote for T, C, or Neutral

Note that this adapts in response to *p*-values!

MSB in REDCap

2Gether 3.0

- N_0 the initial number of completely random assignments.
  - N_0 = 100 in 2Gether
- Covariates used for the MSB algorithm
  - Race, age, relationship status, gender identity, PReP usage
- Assignment probability
  - 2Gether used $\pi = 70\%$
- Test of differences between groups and “significant” p-value
  - 2Gether used $\alpha = 0.3$

How much can MSB reduce chance imbalance?

Simulations

- Want balance on:
  - Age ($|d| < 0.1$)
  - Race (% diff < 5%)
  - Gender (% diff < 5%)
  - Relationship status (% diff < 5%)
  - PReP use (% diff < 5%)

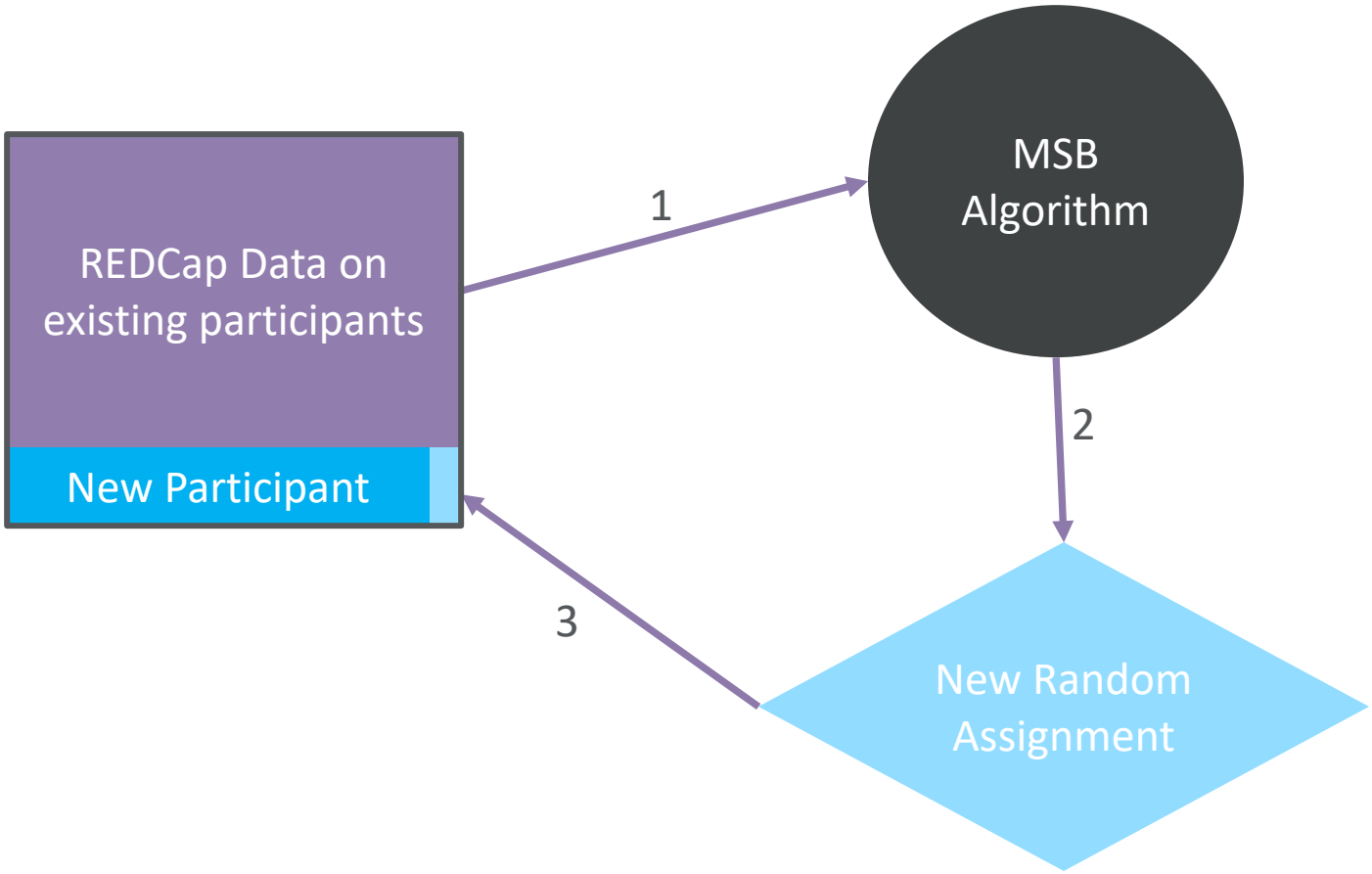
How much can MSB reduce chance imbalance?

Simulations

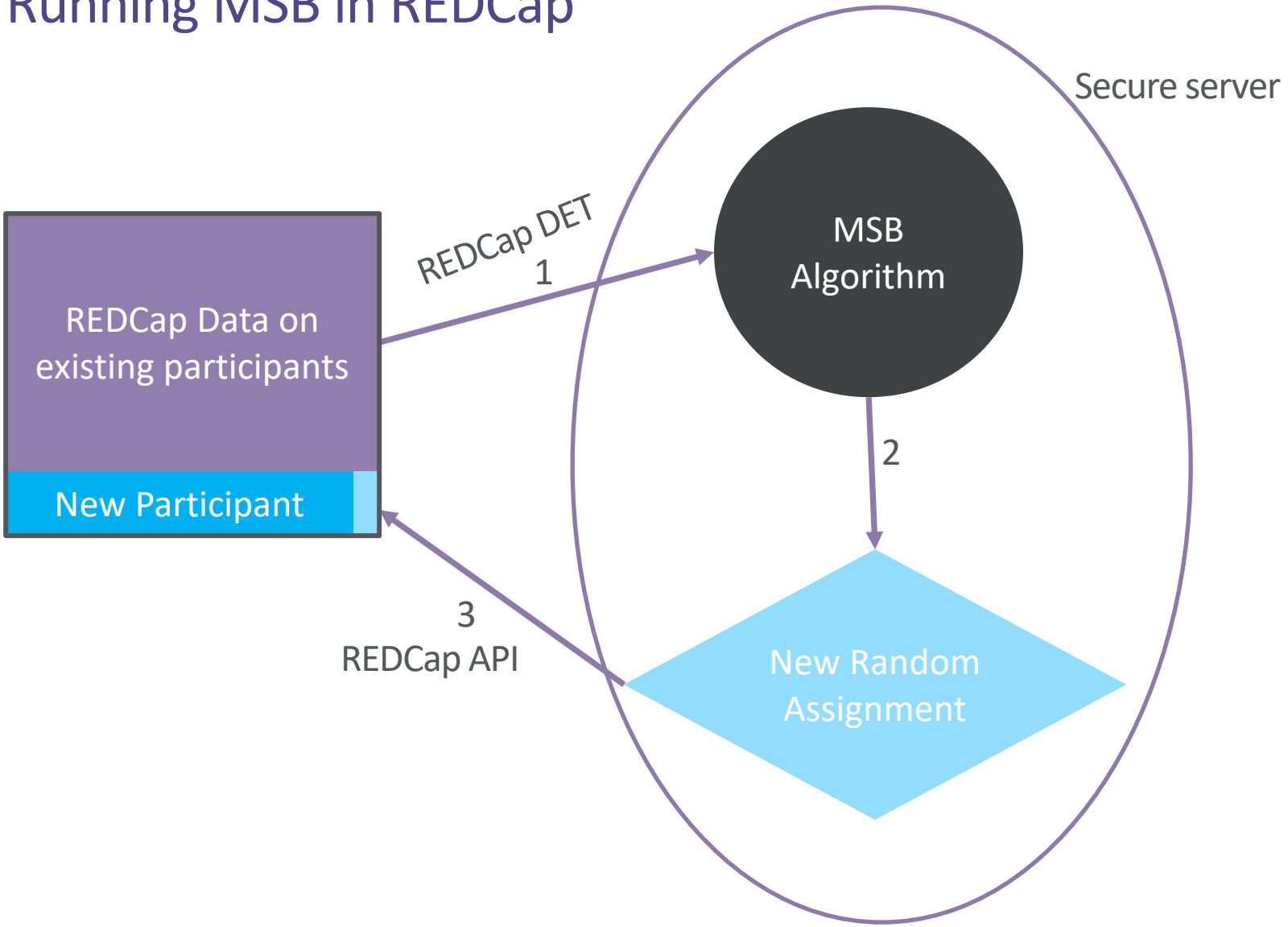
- Want balance on:
  - Age ($|d| < 0.1$)
  - Race (% diff < 5%)
  - Gender (% diff < 5%)
  - Relationship status (% diff < 5%)
  - PReP use (% diff < 5%)

	% of Imbalanced Allocations	
Sample Size	Traditional	MSB
200	82.6% (0.02)	74% (0.02)
500	61.2% (0.02)	56% (0.02)
1,000	27.2% (0.02)	12.0% (.01)
5,000	1% (0.001)	0% (0.001)

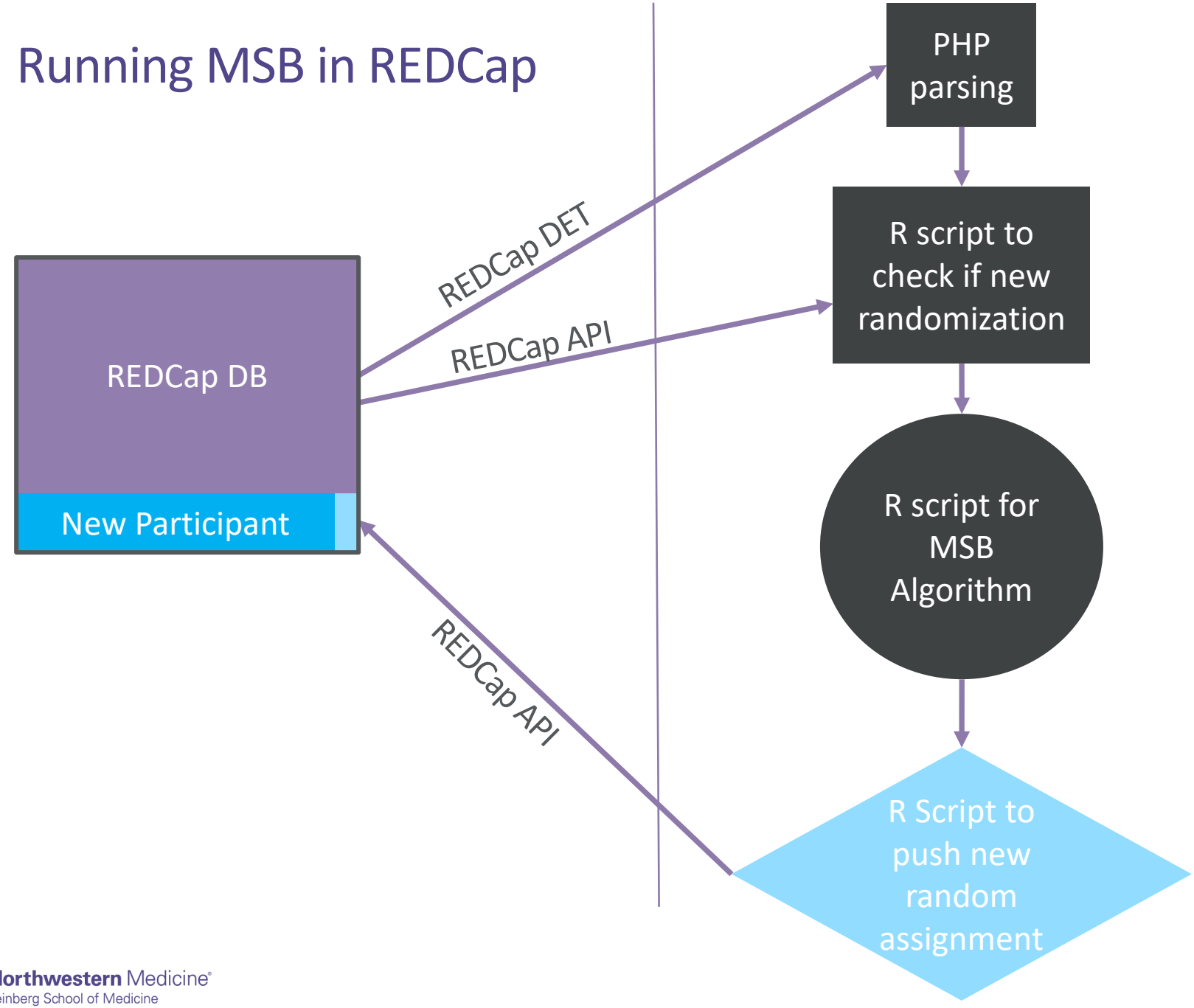
Running MSB in REDCap



Running MSB in REDCap



Running MSB in REDCap



What does DET + MSB look like in practice?

- Off to R and REDCap!

Northwestern Medicine®
Feinberg School of Medicine

Some take-aways

Summary

What/why are constrained randomization methods?

- Chance imbalance is something we can and should guard against in randomized experiments.
  - Throws off our statistics and limits our trust in the validity of our results.
- Constraining randomization through
  - Stratified randomization
  - Covariate-constrained randomization
  - Adaptive randomization
- Each of these has tradeoffs and myriad ways to implement!

Summary

Which one should I use?

- Specify relevant covariates on which you think balance is critical.
  - If there are only a few, maybe stratified randomization will suffice.
  - If there are many, consider the constrained/adaptive randomization algorithms.
  - Consider *when* you'll have covariate data, and *when* it would be optimal to randomize
    - E.g., immediately after baseline data collection? The following day? Week?

Summary

What approach/tuning parameters should I use?

- Implementations range in complexity
- Key point: All constrained randomizations improve balance on the covariates you specify (and not necessarily the ones you don't specify) in the *manner* you specify.
  - Controlling balance on one covariate does not necessarily induce balance on other covariates.
  - Adaptive algorithms that "adapt" in response to *p*-values do not necessarily reduce absolute differences (they reduce the probability of large *p*-values).

Summary

How do I make it work with my study?

- Talk to your friendly neighborhood statistician
  - This can be written into grants (reviewers like this).
  - Write this into your ClinicalTrials.gov registration.
  - It makes for an interesting paragraph in your protocol paper!
  - It can be programmed for your study.
- Existing code repositories can ease future implementations

Northwestern Medicine®
Feinberg School of Medicine

Thank you!

Resources

- Git repository for MSB + DET code
- Tutorials on Stratified Randomization in REDCap
- Tutorials on Covariate-Constrained Randomization
- Overview on constraining randomization to ensure balance

Northwestern Medicine®
Feinberg School of Medicine

Bonus slides

Minimal Sufficient Balance

Computing the "vote" for a covariate

Continuous covariates

1. Compute *d* = the standardized mean difference between arms for covariate X_j (Treatment – Control)
2. If $|d| > \delta_{max}$:
   a. Vote for **T** if [$\bar{X}_{jC} > \bar{X}_{jT}$ and $X_{ij} > \bar{X}_{jC}$] OR [$\bar{X}_{jC} < \bar{X}_{jT}$ and $X_{ij} < \bar{X}_{jC}$]
   b. Vote for **C** if [$\bar{X}_{jC} < \bar{X}_{jT}$ and $X_{ij} > \bar{X}_{jT}$] OR [$\bar{X}_{jC} > \bar{X}_{jT}$ and $X_{ij} < \bar{X}_{jT}$]
3. Vote "Neutral" otherwise

For each covariate you get a vote for T, C, or Neutral

Note that this adapts in response to *absolute differences*!

How much can MSB reduce chance imbalance?

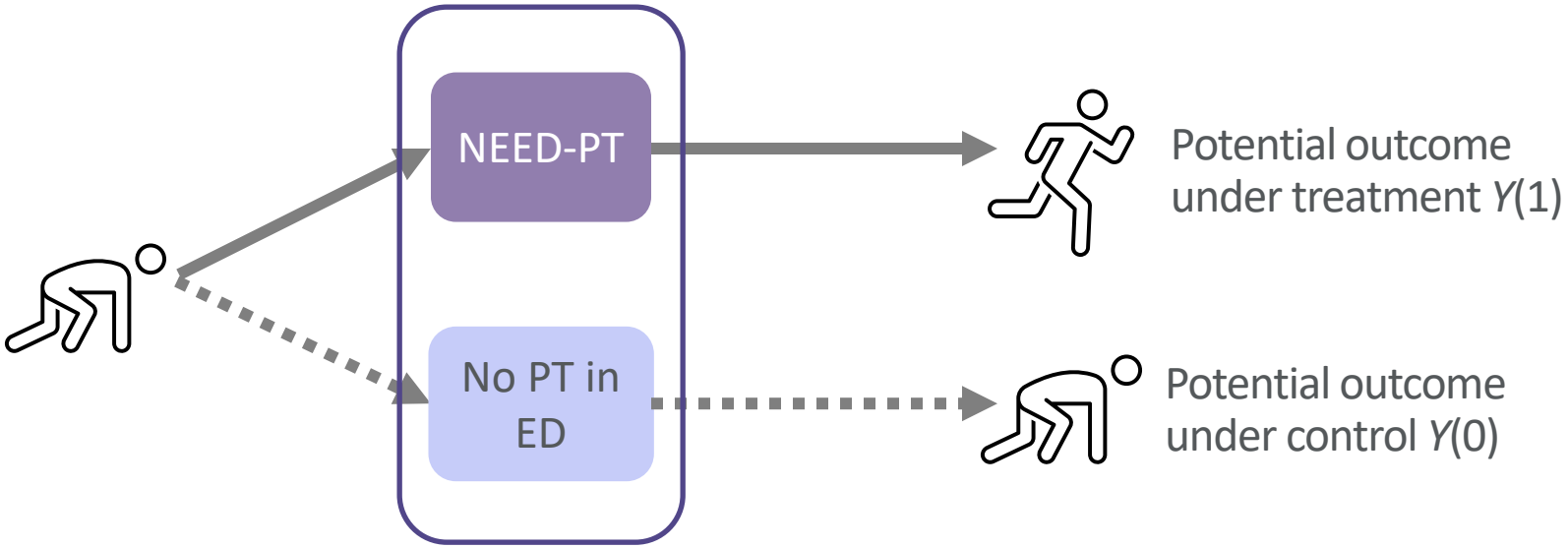
Simulations

- Want balance on:
  - Age ($d < 0.1$)
  - Race (% diff < 5%)
  - Gender (% diff < 5%)
  - Relationship status (% diff < 5%)
  - PReP use (% diff < 5%)

	% of Imbalanced Allocations		
Sample Size	Traditional	MSB-*p*	MSB-*d*
200	83% (0.02)	74% (0.02)	65% (0.02)
500	61% (0.02)	56% (0.02)	24% (0.02)
1,000	27% (0.02)	12% (.01)	1% (0.004)
5,000	<1% (0.001)	0% (0.001)	0% (0.001)

Causality

Conceptual & statistical model



Treatment Assigned

$Z = 1 \Longrightarrow$ PT

$Z = 0 \Longrightarrow$ No PT

Observed outcome $Y = Y(1)Z + Y(0)(1 - Z)$