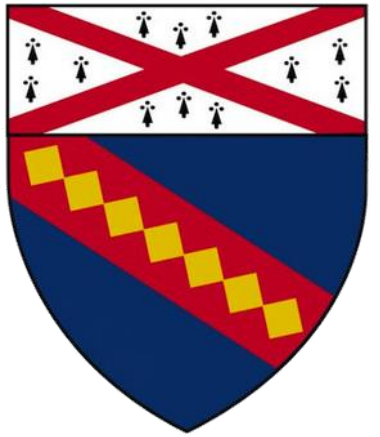
Study Designs and Methods for Implementation Research

Denise Esserman, PhD
Professor of Biostatistics
Yale School of Public Health



Why is Study Design Particularly Important to Biostatisticians?

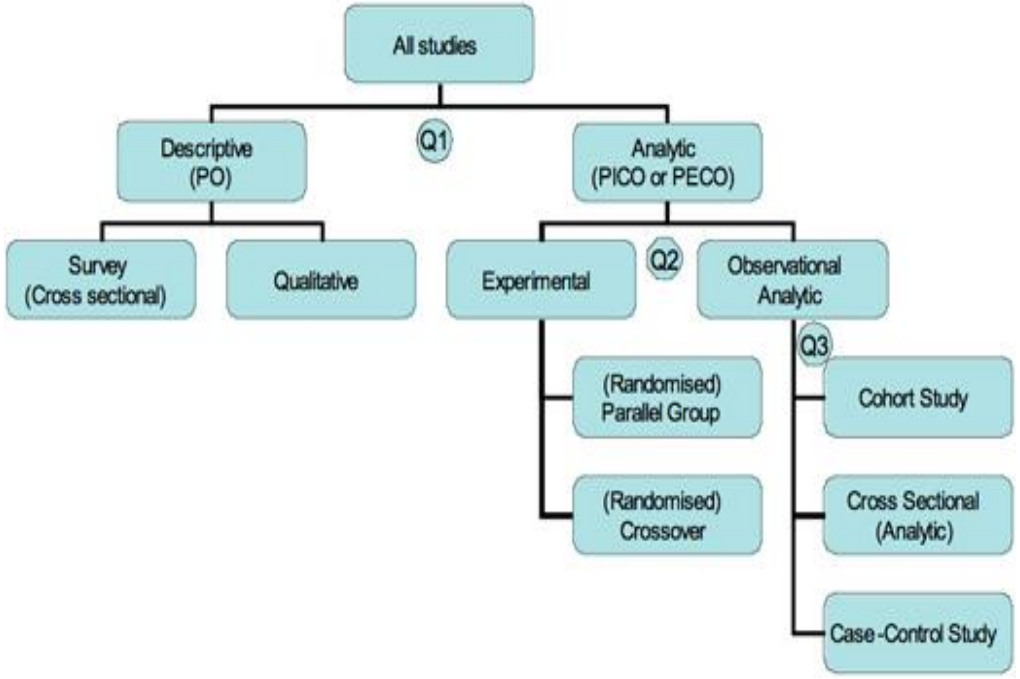
- The study design lays the foundation for collecting, analyzing, and interpreting data in a way that produces valid, reliable, and reproducible results.
- The quality of the study design directly impacts the statistical analysis and ultimately the conclusions that can be drawn from the study.

Your analysis is as good as your data.

Image source: OpenAI. (2024). Generated by ChatGPT. https://www.openai.com/

What Type of Design Should You Use?

"The type of question that you have will often lead to the type of research that will best answer the question."



- Q1: What is the aim of the study?
- Q2: Can/will the intervention be randomly assigned?

Intervention/Prevention: RCT>Cohort> Case-Control>Case Series
Therapy: RCT> Cohort > Case-Control > Case Series
Prognosis/Prediction: Cohort > Case-Control > Case Series
Diagnosis/Diagnostic: Prospective study
Etiology: RCT>Cohort> Case-Control>Case Series

- Q3: When will the outcome be determined?

https://library.louisville.edu/kornhauser/EBP/types-evidence

Study designs and statistical considerations

All design and statistical considerations are driven by the **research question and aims.**

PICO (T)				
Population	Intervention	Comparator	Outcome	Time

Review of PICO-T

- **P**opulation
  - Who or what is the patient, population or problem in question?
- **I**ntervention
  - What is the intervention being considered?
- **C**ontrol/**C**omparator
  - What is the intervention being compared to?
- **O**utcome
  - What is the desired outcome?
- **T**ime
  - How long will it take to reach the desired outcome?

Two Basic Design Strategies of Clinical Trials

Explanatory (Efficacy)

- Optimal conditions
- Academic/specialized centers
- Smaller sample size
- Tighter (homogeneous) eligibility criteria
- Complex treatment protocols
- Tightly monitored
- Hard endpoints, e.g., mortality, stroke, etc.
- Typical study: double-blind, placebo-controlled drug trial

Pragmatic (Effectiveness)

- Real world
- Community based patients
- Larger sample size
- Broader (heterogeneous) eligibility criteria
- Easily administered treatments
- Little monitoring of adherence and fidelity
- Patient centered outcomes: Mortality, morbidity, QoL
- Typical study: comparison of an intervention vs usual care

Research Pipeline

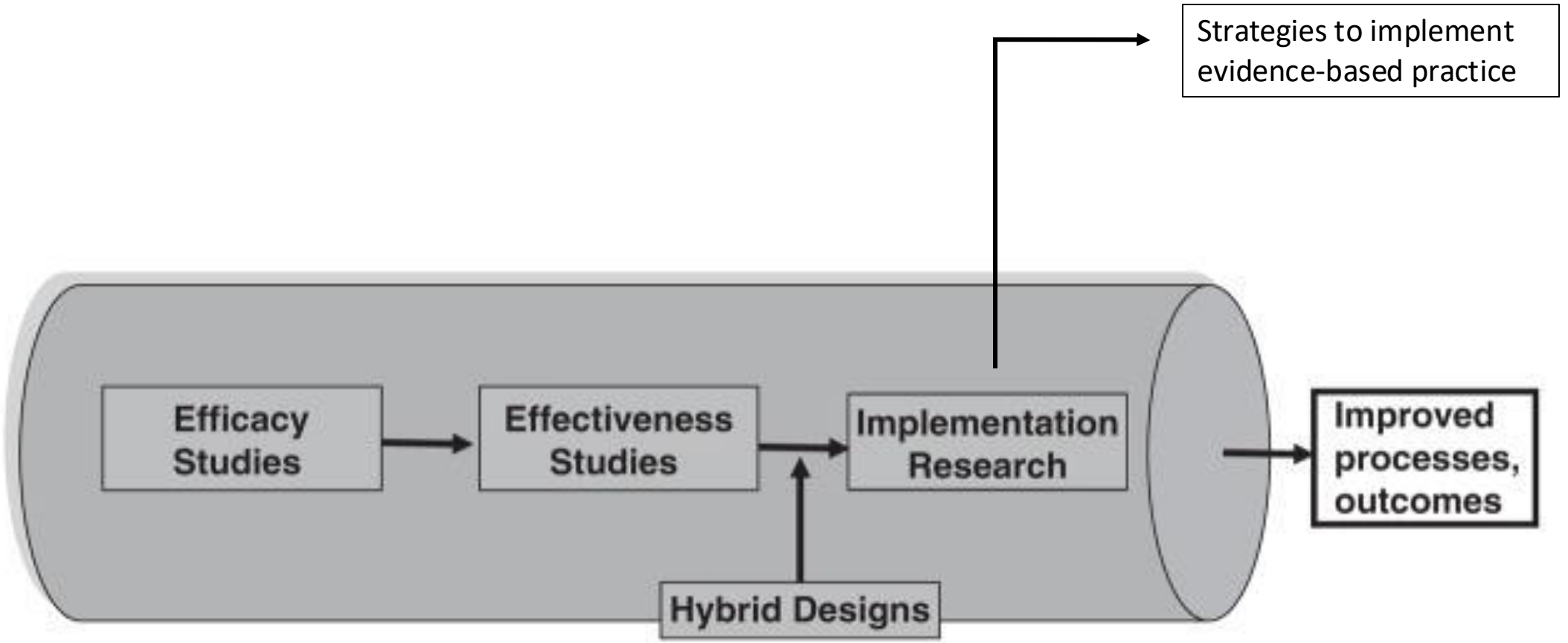
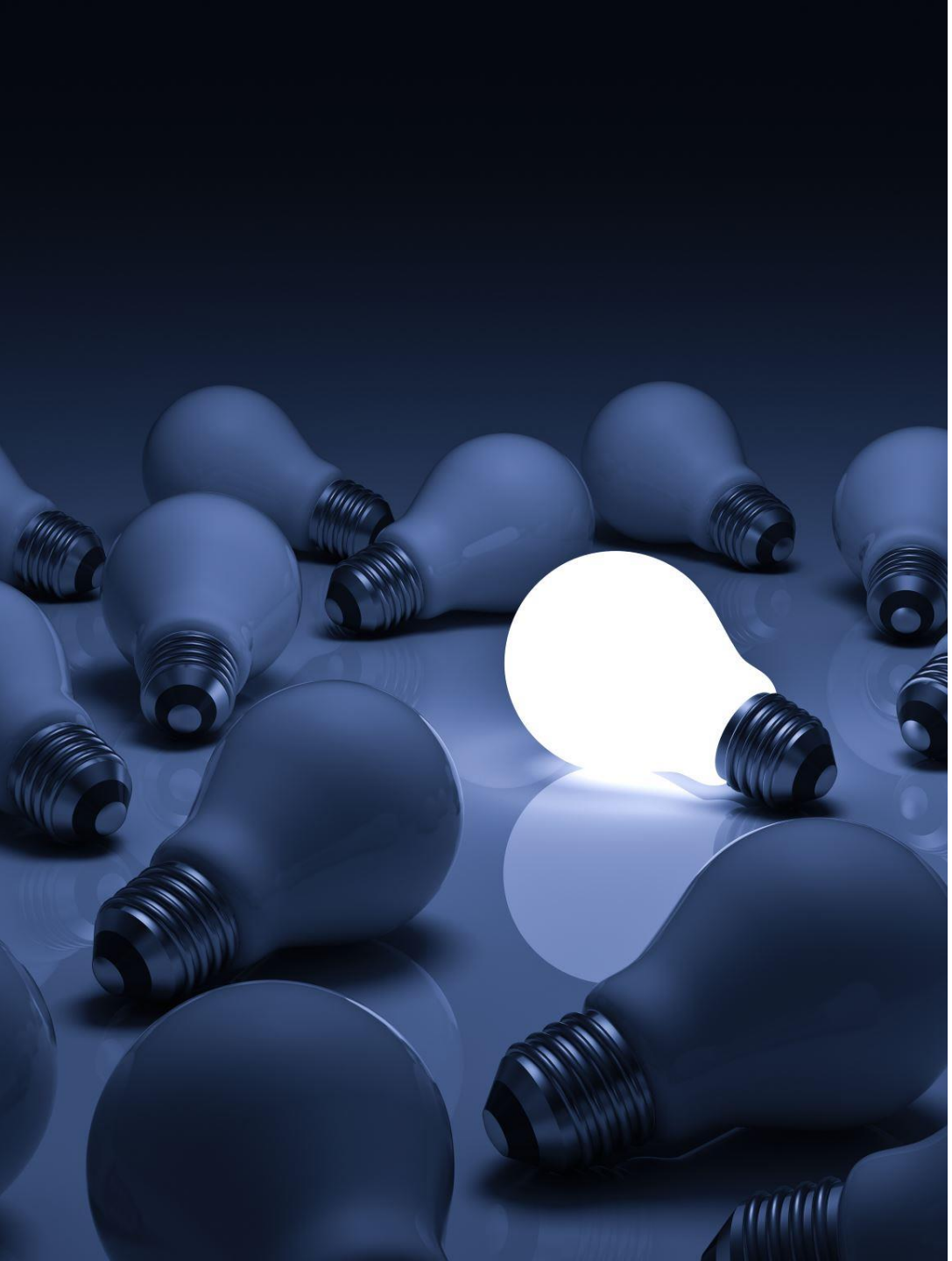


Figure 1: Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012 Mar;50(3):217-26. doi: 10.1097/MLR.0b013e3182408812. PMID: 22310560; PMCID: PMC3731143.

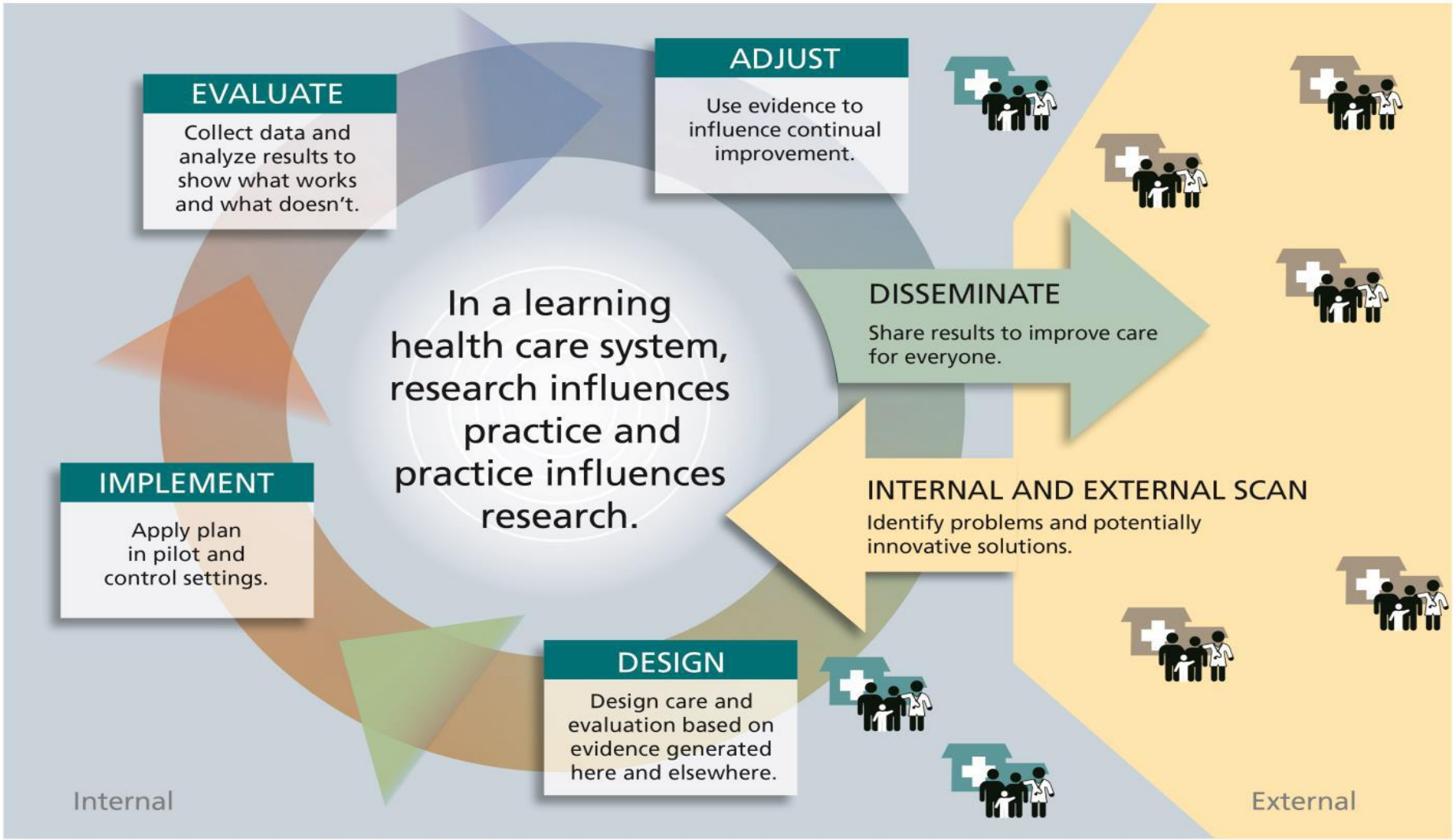


Implementation Science

How do we get "what works" to the people who need it, with greater speed, fidelity, efficiency, quality, and relevant coverage?

https://impsciuw.org/implementation-science/learn/implementation-science-overview/

Promotion of learning health system



Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med. 2012 Aug 7;157(3):207-10. doi: 10.7326/0003-4819-157-3-201208070-00012. PMID: 22868839.
https://rethinkingclinicaltrials.org/

Focus of Implementation Designs

- External validity
- Implementation-barriers and facilitators
  - What is preventing the use and sustainability of evidence-based interventions?
- Studying factors that lead to uptake of evidence-based practice
- Understanding factors that may moderate/limit robustness
  - Setting – (e.g., implementation; geographic)
  - Demographic characteristics (e.g., race/ethnicity; age)

Hybrid Design Types

1

Primary aim is effectiveness; secondary aim is implementation

2

Co-primary aims are effectiveness and implementation

3

Primary aim is implementation (strategies); secondary aim effectiveness/clinical outcomes

Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012 Mar;50(3):217-26. doi: 10.1097/MLR.0b013e3182408812. PMID: 22310560; PMCID: PMC3731143.

Design Characteristics

Design Characteristic	Clinical Effectiveness Trial	Implementation Trial
Test	“Clinical” Intervention	Implementation intervention or strategy
Typical Unit of Randomization*	Patient, clinical unit	Provider, clinical unit, or system
Typical Unit of Analysis	Patient	Provider, clinical unit, or system
Summative outcomes	Health outcomes; cost	Adoption/uptake; process/quality measures

Recreated Table 2: Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012 Mar;50(3):217-26. doi: 10.1097/MLR.0b013e3182408812. PMID: 22310560; PMCID: PMC3731143.

Design Considerations

- Testable hypothesis
- Valid and measurable endpoints
- Delivery of the intervention
- Randomization, Stratification and Blinding
- Sample size/power consideration
- Interim monitoring
  - Safety, efficacy, futility
- Analysis
- Data collection
- Ethics

Goal is to obtain an unbiased assessment of the “treatment” effect and carryout valid hypothesis tests

Some Additional Considerations

Experimental vs. quasi-experimental design

Types of comparisons

- Within-site
  - Quasi-experimental (e.g., pre-post; interrupted time series)
- Between-site
  - Randomization at the level of implementation (avoid contamination) (e.g., cluster)
- Within- and between-site
  - Cross-over type designs (e.g., Stepped Wedge; Waitlist control)

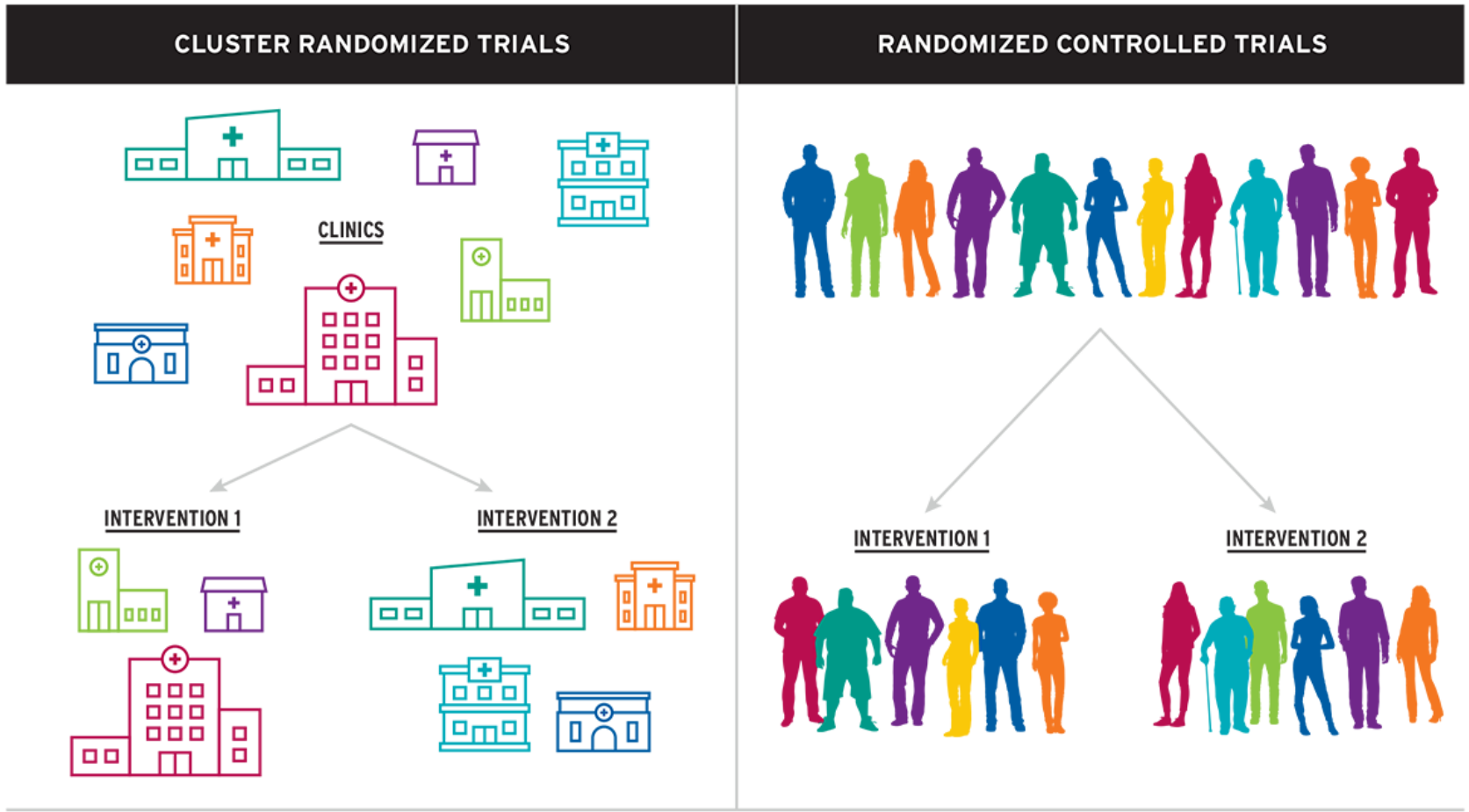
Types of randomization

- Simple
- Stratified
- Covariate Constrained
- Individually Randomized vs. Cluster Randomized vs. Partially Clustered

Treatment

- Combination of treatment vs. Monotherapy
- Sequenced treatments vs. Un-sequenced
- Patient Preference for treatment vs. No Allowance

Experimental Designs



Clinical Evidence Generation— Why and How | FMCNA
fmcna.com

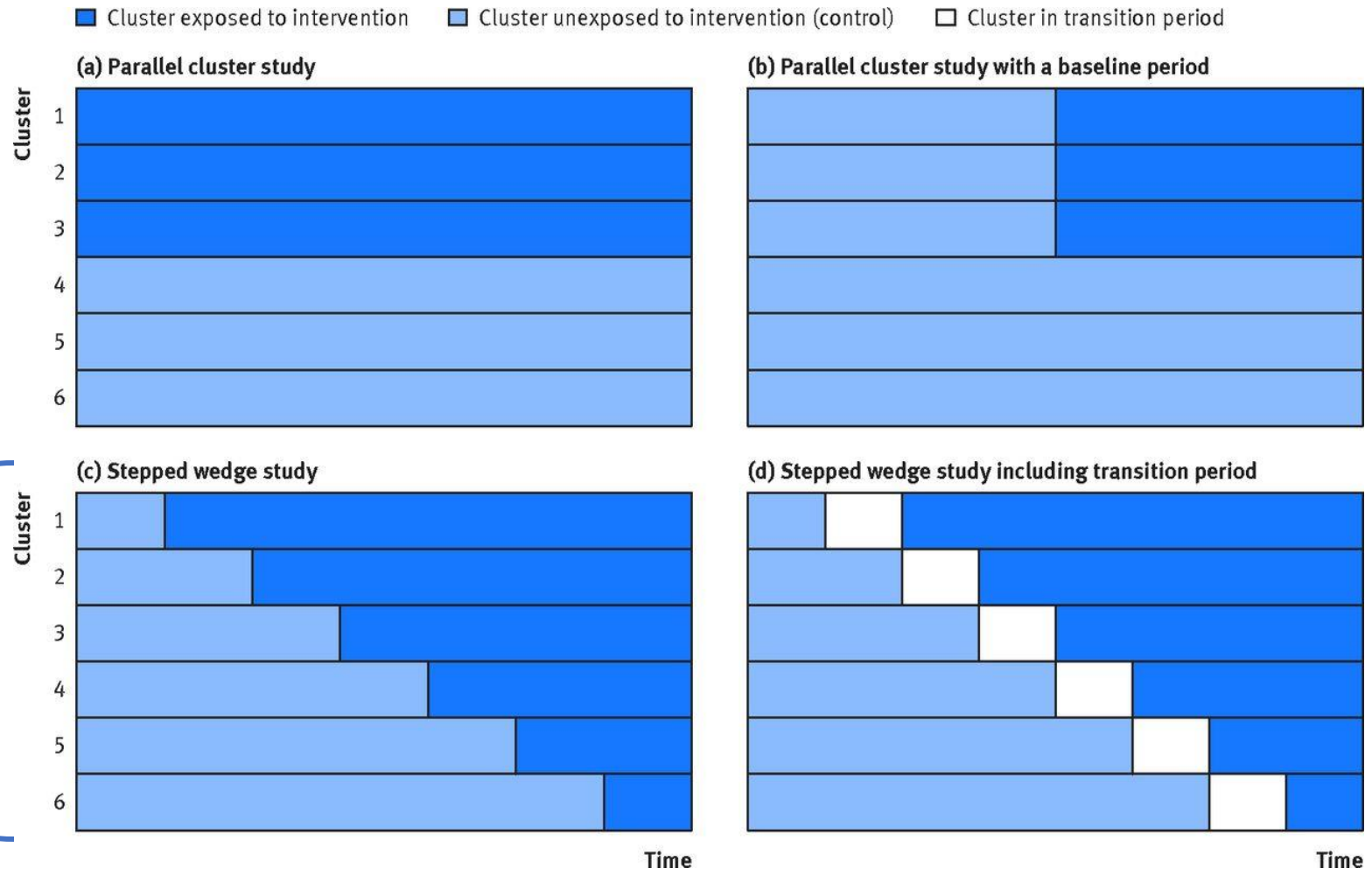
Quasi-Experimental Designs

Pre-Post Designs

- Non-equivalent Control group
  - Capitalize on naturally occurring experiments
  - Threats to internal validity – no “equivalent” control group
    - Mitigate through selection of control, matching, propensity score methods
  - Regression to mean (e.g., if a site was selected because it was underperforming then would see improvement regardless of the intervention)
  - Qualitative methods can help to uncover factors affecting results
- Interrupted Time Series
  - Good when cannot identify a comparable control group (e.g., policy mandates)
  - Pre- and post-intervention data are collected
  - Defining pre- and post-intervention period can be challenging
  - Need to collect sufficient data points across time
  - Need to worry about confounding from other interventions that may also be implemented

Cluster Designs

ONE-WAY cross-over design



https://www.bmj.com/content/350/bmj.h391

Stepped-Wedge Cluster Design

DESIGN CHOICES

- ❑Closed Cohort
  - All subjects participate from beginning to end
- ❑Open Cohort
  - New subjects can enter, and others can leave (time of exposure varies)
- ❑Cross-Sectional
  - Subjects are only assigned to treatment or control

Stepped-Wedge Cluster Design

ADVANTAGES

- All units receive the intervention
- Helps with resource allocation
- Used to reconcile robust evaluation with political or logistical constraints
- If substantial cluster-level effects present or have large clusters, stepped wedge will be more powerful than parallel design (more cost efficient)

CHALLENGES

- Contamination
- Carry-over effects
- Temporal trends (confounding effect of time)
- Blinding
- Multiple data collection points are required

https://trialsjournal.biomedcentral.com/counter/pdf/10.1186/s13063-015-0842-7.pdf

Other Possible Design Choices

Factorial or Fractional-factorial designs

- Multi-component implementation strategies
- Randomize to different combinations – can evaluate effectiveness of individual components and/or combinations

Adaptative Implementation Designs

- SMART (Sequential, Multiple Assignment Randomized Trial)
  - Optimizing treatment sequences over time
  - Some or all participants are randomized more than once
- Stepped Care Designs

Single Subject Experimental Designs

- On-Off-On
- Requires reversibility of the intervention

Miller CJ, Smith SN, Pugatch M. Experimental and quasi-experimental designs in implementation research. Psychiatry Res. 2020 Jan;283:112452. doi: 10.1016/j.psychres.2019.06.027. Epub 2019 Jun 20. PMID: 31255320; PMCID: PMC6923620.

Factorial Design Example: Three Treatments

- FULL FACTORIAL

Run	Treatment	Factors		
		A	B	C
1	I	-1	-1	-1
2	a	+1	-1	-1
3	b	-1	+1	-1
4	ab	+1	+1	-1
5	c	-1	-1	+1
6	ac	+1	-1	+1
7	bc	-1	+1	+1
8	abc	+1	+1	+1

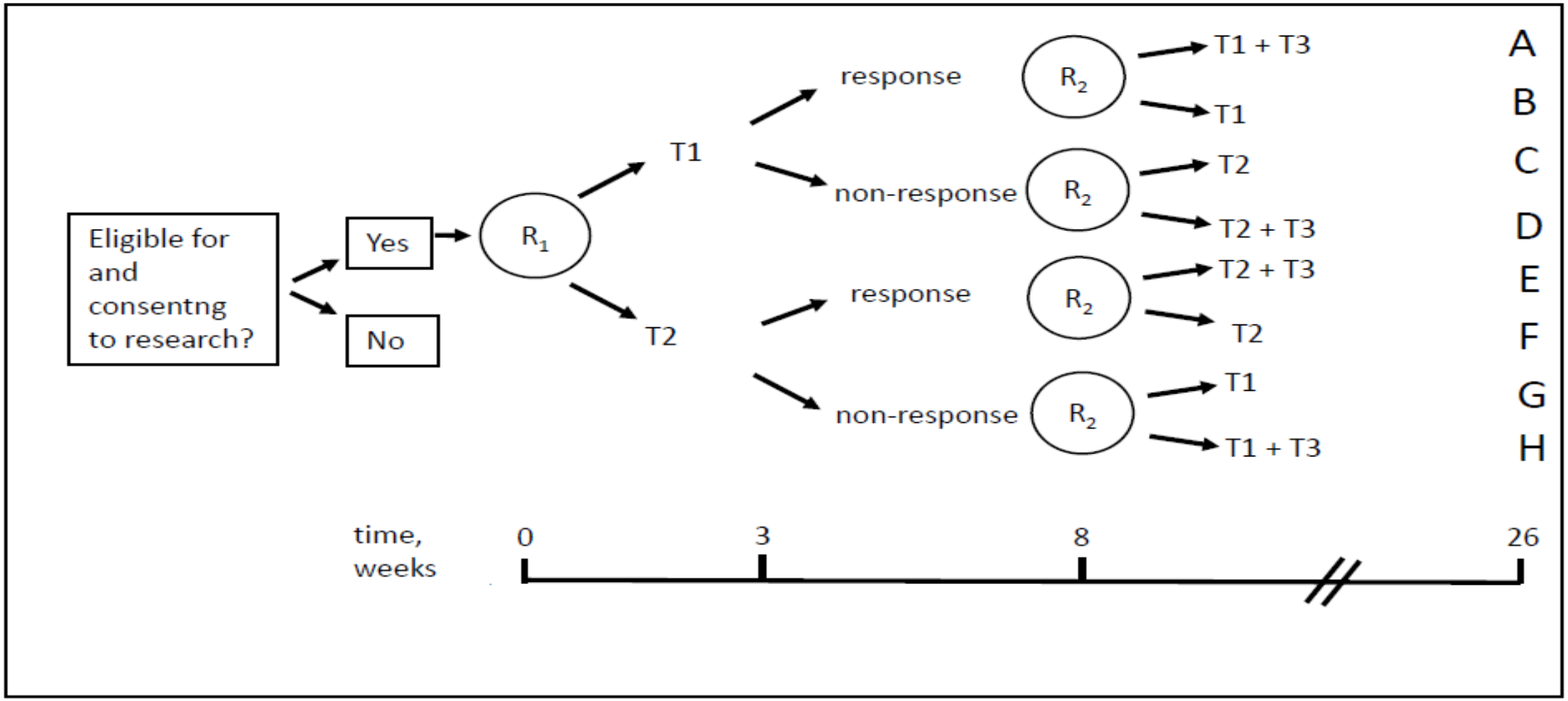
https://www.leansigmacorporation.com/full-factorial-doe-with-minitab/

- FRACTIONAL FACTORIAL

Run	Treatment	Factor		
		A	B	C
1	c	-1	-1	+1
2	a	+1	-1	-1
3	b	-1	+1	-1
4	abc	+1	+1	+1

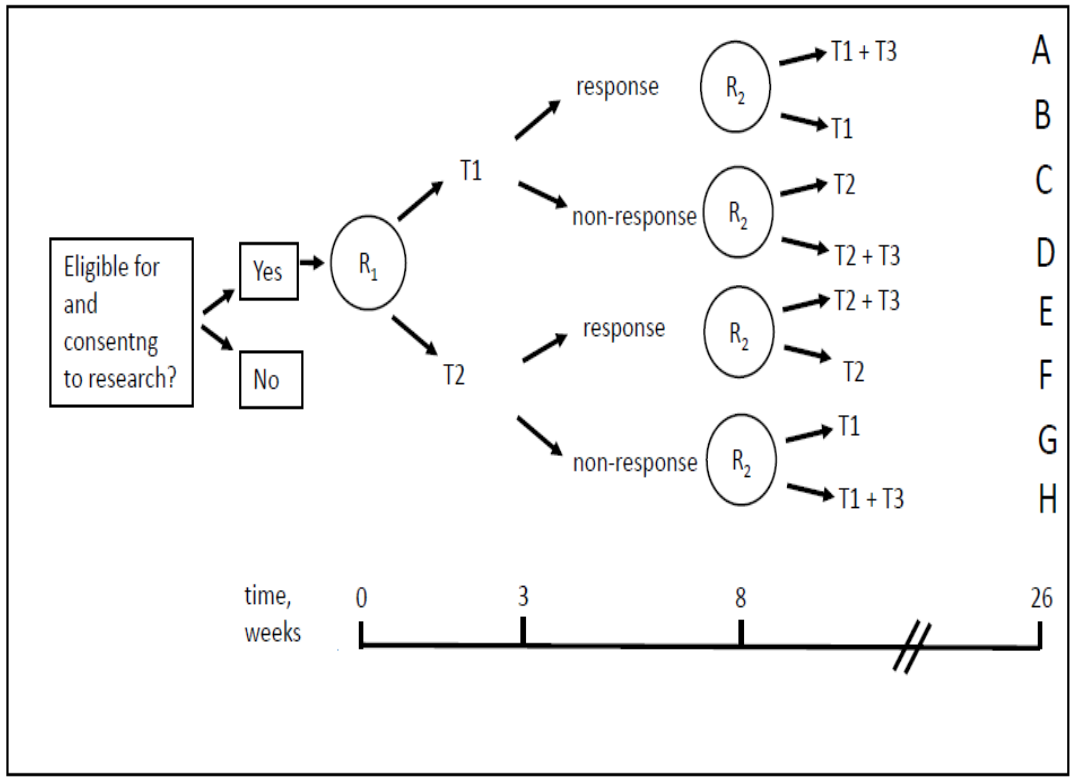
https://www.leansigmacorporation.com/fractional-factorial-designs-with-jmp/

Example of a SMART DESIGN



Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. Annu Rev Clin Psychol. 2012;8:21-48. doi: 10.1146/annurev-clinpsy-032511-143152. Epub 2011 Dec 12. PMID: 22224838; PMCID: PMC3887122.

Example of a SMART DESIGN



Embedded adaptive intervention	First-stage treatment options	Status at end of first-stage	Second-stage treatment options	Subgroup	Strategy
1	T1	Responder	T1 + T3 (augment)	A	A + C
		Nonresponder	T2 (switch)	C	
2	T1	Responder	T1 + T3 (augment)	A	A + D
		Nonresponder	T2+ T3 (switch + augment)	D	
3	T1	Responder	T1 (maintenance)	B	B + C
		Nonresponder	T2 (switch)	C	
4	T1	Responder	T1 (maintenance)	B	B + D
		Nonresponder	T2+ T3 (switch + augment)	D	
5	T2	Responder	T2+ T3 (augment)	E	E + G
		Nonresponder	T1 (switch)	G	
6	T2	Responder	T2 + T3 (augment)	E	E + H
		Nonresponder	T1+ T3 (switch + augment)	H	
7	T2	Responder	T2 (maintenance)	F	F + G
		Nonresponder	T1 (switch)	G	
8	T2	Responder	T2 (maintenance)	F	F + H
		Nonresponder	T1+ T3 (switch + augment)	H	

Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. Annu Rev Clin Psychol. 2012;8:21-48. doi: 10.1146/annurev-clinpsy-032511-143152. Epub 2011 Dec 12. PMID: 22224838; PMCID: PMC3887122.

Scientific Questions with SMART

Scientific question	Contrast/analysis of interest
Which initial treatment produces better outcomes at 6 months?	A+B+C+D versus E+F+G+H
Among non-responders to initial treatment, is it better to switch treatment or switch treatment and augment?	C+G versus D+H
Among responders to initial treatment, is it better to maintain treatment or augment treatment?	A+E versus B+F
Which of the 8 embedded adaptive interventions leads to greatest change in the outcome of interest?	Identify best among: A+C, A+D, B+C, B+D, E+G, E+H, F+G, F+H
Can we develop more deeply tailored adaptive interventions?	

SMART Design

ADVANTAGES

More representative of practice

- Allow adaptations

Can answer multiple research questions

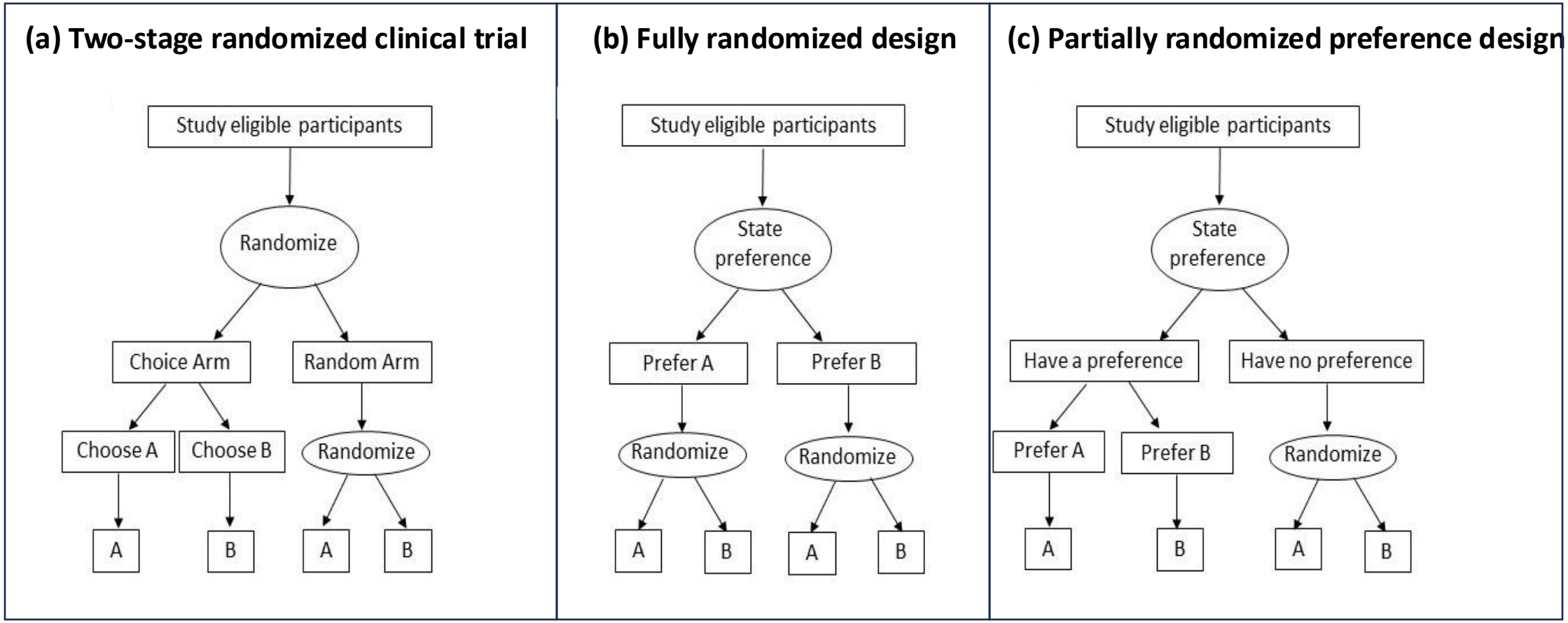
Can get evidence regarding interactions among multiple treatments

CHALLENGES/LIMITATIONS

- Requires more planning
  - Disruptions can have a major impact on the findings
- Generally, does not provide enough evidence to definitively establish effectiveness

https://d3c.isr.umich.edu/experimental-designs/sequential-multiple-assignment-randomized-trials-smarts/; Kidwell KM, Almirall D. Sequential, Multiple Assignment, Randomized Trial Designs. *JAMA.* 2023;329(4):336–337. doi:10.1001/jama.2022.24324

Patient Preference Designs



Covariate Constrained Randomization

- Useful in cluster randomized trials with small number of clusters
  - Higher probability of a “bad” randomization
- Allocate treatment so that balance on key covariates
- Can use exact or caliper criteria
- Randomly sample from a list of acceptable allocations
  - Need to make sure there is a sufficient set of possibilities

Moulton LH. Covariate-based constrained randomization of group-randomized trials. Clin Trials. 2004;1(3):297-305. doi: 10.1191/1740774504cn024oa. PMID: 16279255.

Analytic Considerations

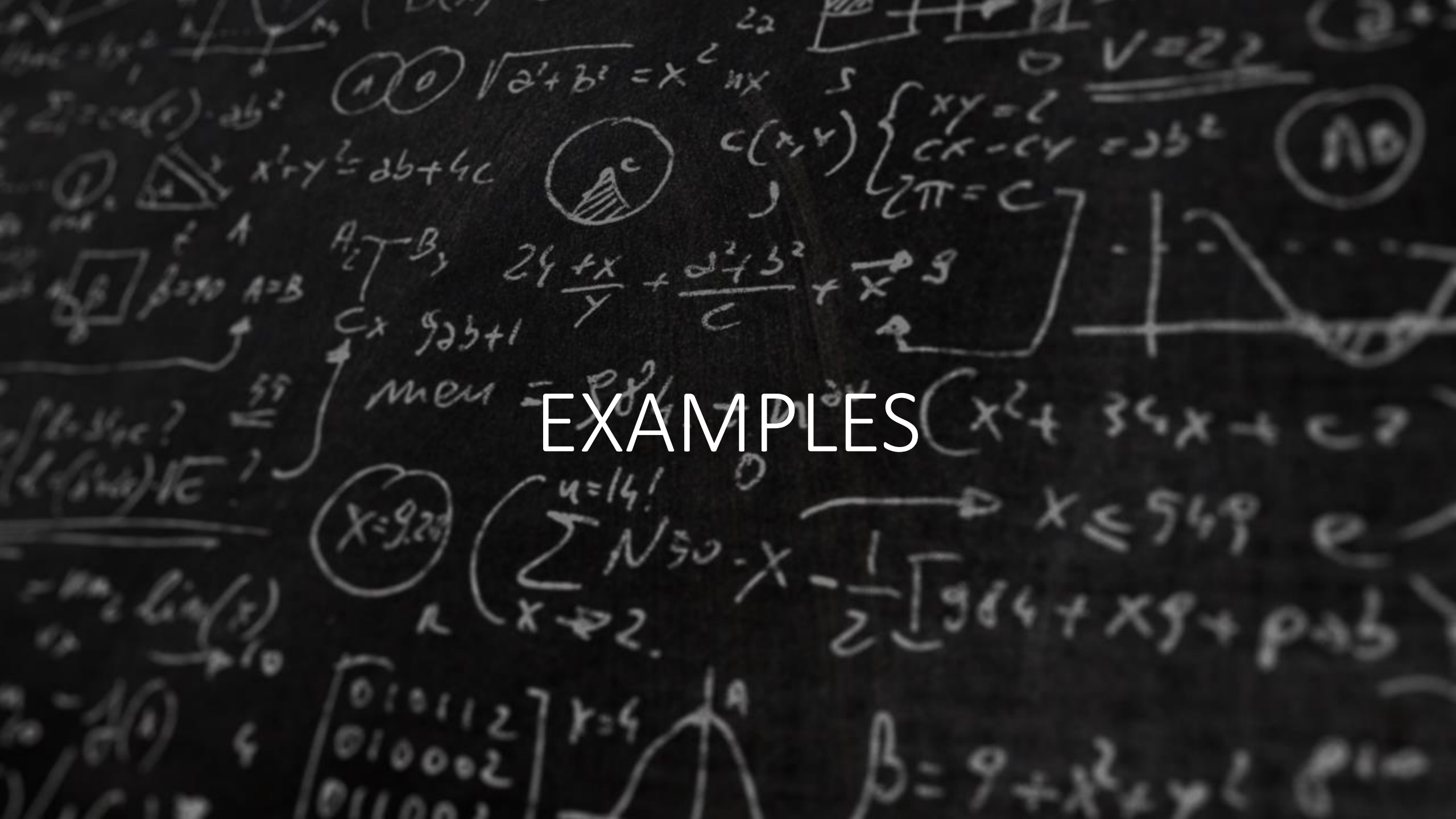
- Implementation studies are often multi-level
  - To evaluate need to understand levels at which intervention are implemented and measured
- Need to account for clustering
  - Hierarchical generalized linear mixed effect models (GLMM)
  - General estimating equations (GEE)
  - Frailty models (survival end points)
- Segmented Regression
  - Model pre-post intervention trends
- Propensity Score Methods
  - Matching, stratification, weighting

Mazzucca S, Tabak RG, Pilar M, Ramsey AT, Baumann AA, Kryzer E, Lewis EM, doi: 10.3389/fpubh.2018.00032. PMID: 29515989; PMCID: PMC5826311.

Sample size considerations

- Need to consider impact of “real-world” when hypothesizing an effect size
  - What is meaningful from an implementation perspective?
  - Small improvements may have large impacts (e.g., cost)
- For a cluster randomized trial, the sample size is the number of clusters → impacts choice of design
- Often can be more complicated because of the multi-level nature of the design

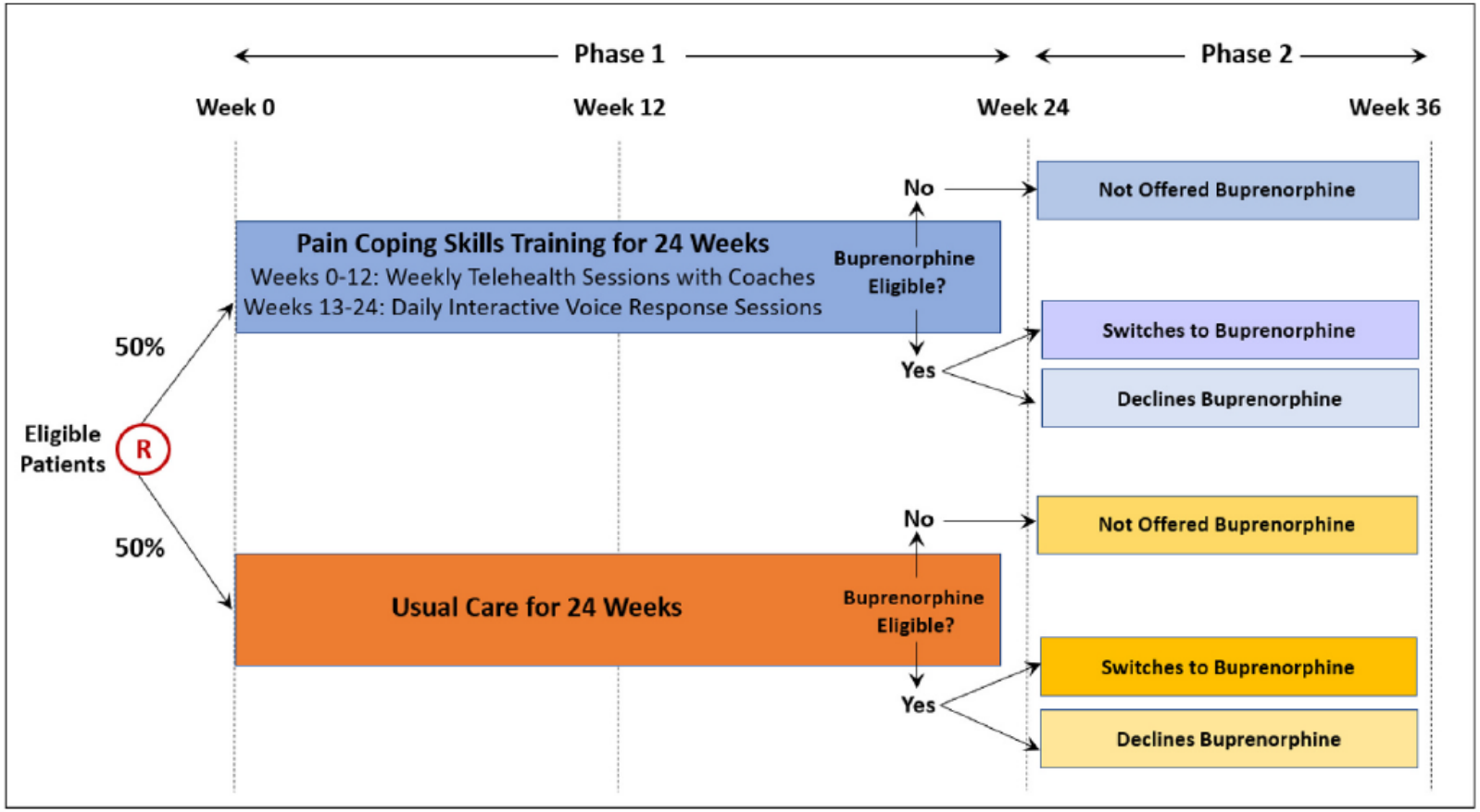
https://www.bmj.com/content/372/bmj.m3721



EXAMPLES

HOPE Consortium Trial

- Goal: Reduce pain and opioid use in hemodialysis (HOPE Trial)
- Multicenter randomized trial addressing chronic pain among patients receiving hemodialysis for end-stage kidney disease
- Design: Sequential, multiple assignment design with randomized component (Phase 1) and non-randomized component for subset of participants (Phase 2)



Dember LM, Hsu JY, Bernardo L, Cavanaugh KL, Charytan DM, Crowley ST, Cukor D, Doorenbos AZ, Edwards DA, Esserman D, FischerMJ, Jhamb M, Joffe S, Johansen KL, Kalim S, Keefe FJ, Kimmel PL, Krebs EE, Kuzla N, Mehrotra R, Mishra P, Pellegrino B, Steel JL, Unruh ML, White DM, Yabes JG, Becker WC; HOPE Consortium. The design and baseline characteristics for the HOPE Consortium Trial to reduce pain and opioid use in hemodialysis. Contemp Clin Trials. 2024 Jan;136:107409. doi: 10.1016/j.cct.2023.107409. Epub 2023 Dec 10. PMID: 38086444; PMCID: PMC10922728.

Managing Hypertension Among People Living with HIV: an Integrated Model (MAP-IT) trial

- Stepped wedge cluster randomized trial
- Test adoption and sustainability of a proven-effective implementation strategy for integration of hypertension control in HIV treatment
  - Apply to low-resource setting
- Evaluate effectiveness of practice facilitation (PF) on the adoption of a task-strengthening strategy for hypertension control (TASSH) for people living with HIV (PLWH)
- Enrolled 30 primary care clinics in Nigeria – plan to enroll 960 PLWH with uncontrolled hypertension
- Primary outcome: Utilization of the components of TASSH (Uptake) – measured at clinic level on monthly basis
- Secondary outcomes: BP control, implementation fidelity, and TASSH sustainability

MAP-IT Stepped Wedge Trial

5 cohorts of 6

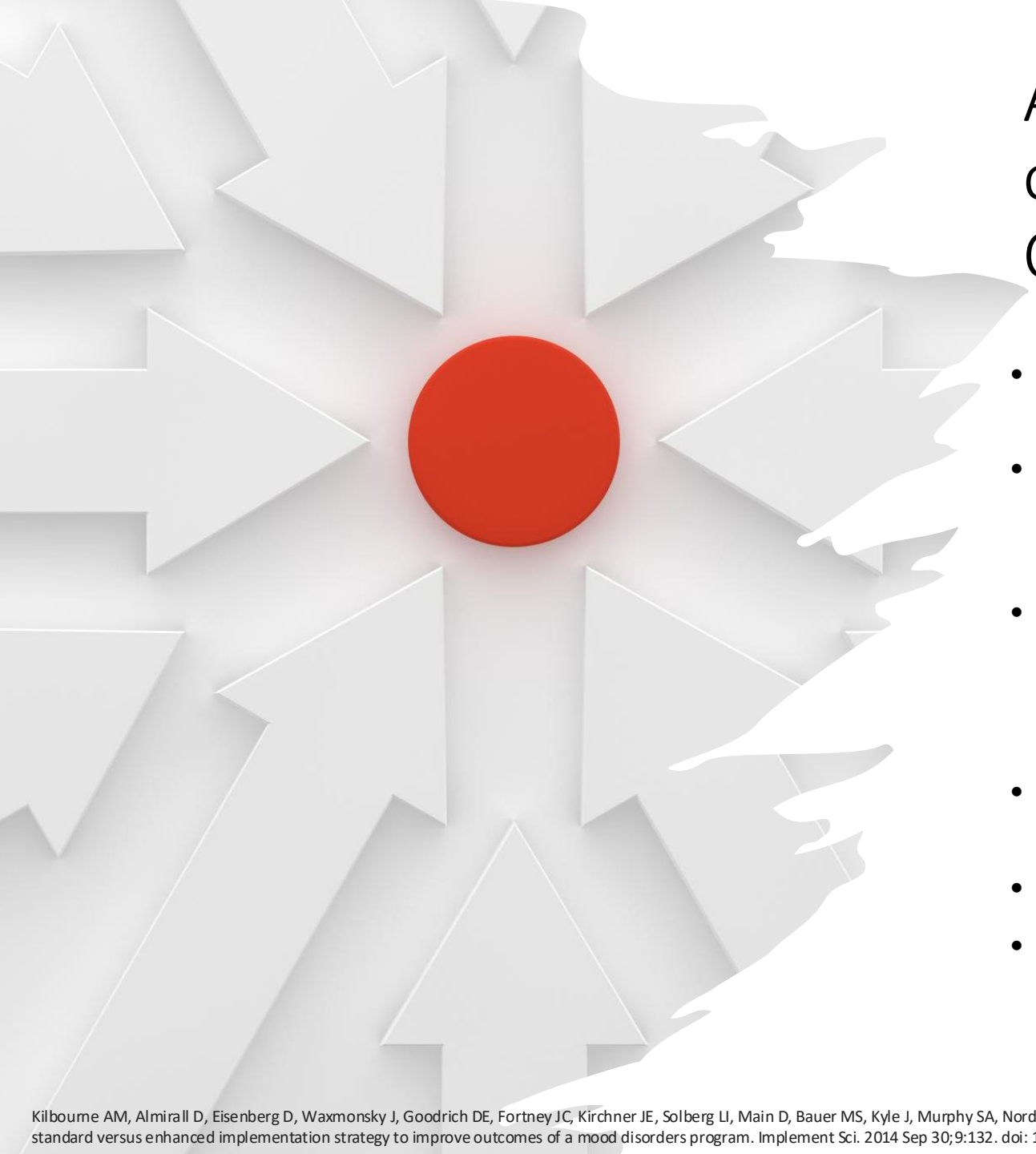
Block	2020				2021				2022				2023				2024				2025			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	UG3 Phase and Pilot											SD	Tr	PF	PF	PF	PF	FU	FU	FU	FU	FU	FU	
2												SD	SD	Tr	PF	PF	PF	PF	FU	FU	FU	FU	FU	
3												SD	SD	SD	Tr	PF	PF	PF	PF	FU	FU	FU	FU	
4												SD	SD	SD	SD	Tr	PF	PF	PF	PF	FU	FU	FU	
5												SD	SD	SD	SD	SD	Tr	PF	PF	PF	PF	FU	FU	
												T_0	T_3	T_6	T_9	T_{12}	T_{15}	T_{18}	T_{21}	T_{24}	T_{27}	T_{30}	T_{33}	T_{36}

TASSH introduced to all 30 facilities

Abbreviations: The control group or **"SD"** means the 'self-directed condition (TASSH only)'; the treatment group or **"PF"** stands for the 'practice facilitation plus TASSH condition (PF+TASSH)'; **"FU"** means the 'follow-up period'; and **"Tr"** is for the 'transition period when PF training occurs".

Fig. 2 Stepped wedge study design

Aifah, A.A., Hade, E.M., Colvin, C. *et al.* Study design and protocol of a stepped wedge cluster randomized trial using a practical implementation strategy as a model for hypertension-HIV integration — the MAP-IT trial. *Implementation Sci* **18**, 14 (2023). https://doi.org/10.1186/s13012-023-01272-5



Adaptive Implementation of Effective Programs Clinical Trial (ADEPT)

- Life Goals (LG) is an effective evidence-based practice for persons with mental disorders
- Replicating Effective Programs [REP] – effective implementation strategy
  - Sites not responding to this strategy
- Compare implementation strategies to improve outcomes of mood disorders
  - Compare REP + External Facilitator (EF) to REP + EF + IF (Internal Facilitator)
- 80 community-based outpatient clinics (N=1600 patients) across US
- Primary outcome: Mental-health related quality of life
- Secondary outcomes: receipt of LG sessions, mood symptoms, implementation costs, and organizational change

Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, Fortney JC, Kirchner JE, Solberg LI, Main D, Bauer MS, Kyle J, Murphy SA, Nord KM, Thomas MR. Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. Implement Sci. 2014 Sep 30;9:132. doi: 10.1186/s13012-014-0132-x. PMID: 25267385; PMCID: PMC4189548

Cluster Randomized SMART

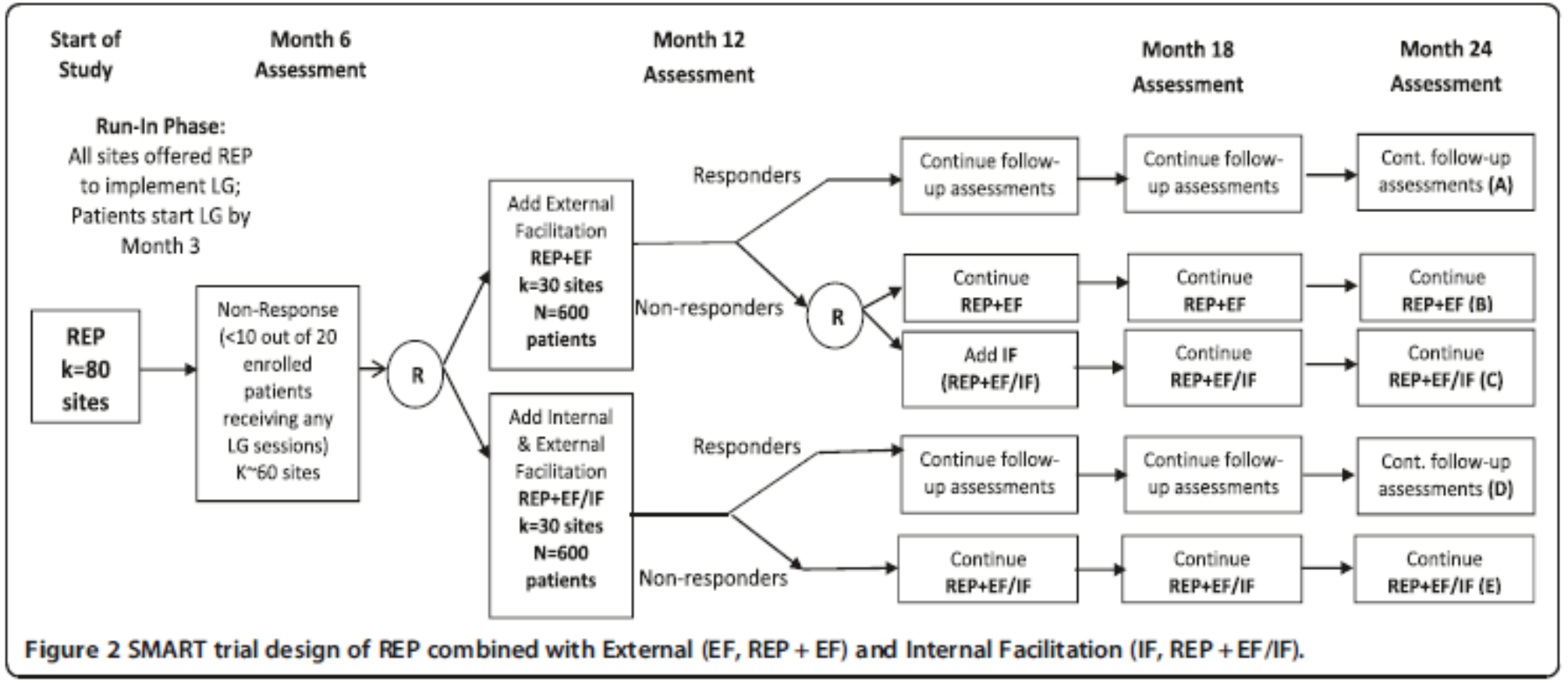


Figure 2 SMART trial design of REP combined with External (EF, REP + EF) and Internal Facilitation (IF, REP + EF/IF).

Main finding: Simpler EF-only augmentation to REP was as good as if not better than the higher intensity EF/IF augmentation for both clinic and patient outcome

-Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, Fortney JC, Kirchner JE, Solberg LI, Main D, Bauer MS, Kyle J, Murphy SA, Nord KM, Thomas MR. Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. Implement Sci. 2014 Sep 30;9:132. doi: 10.1186/s13012-014-0132-x. PMID: 25267385; PMCID: PMC4189548

-Smith SN, Almirall D, Prenovost K, Liebrecht C, Kyle J, Eisenberg D, Bauer MS, Kilbourne AM. Change in Patient Outcomes After Augmenting a Low-level Implementation Strategy in Community Practices That Are Slow to Adopt a Collaborative Chronic Care Model: A Cluster Randomized Implementation Trial. Med Care. 2019 Jul;57(7):503-511. doi: 10.1097/MLR.0000000000001138. PMID: 31135692; PMCID: PMC6684247.

Type 2, Hybrid Cluster Randomized Trial

Test the Implementation and Sustainment Facilitation (ISF) strategy as an adjunct to the Addiction Technology Transfer Center (ATTC) strategy for integrating a motivational interviewing-based brief intervention (MIBI) for substance use disorder (SUDs) within HIV community-based organizations.

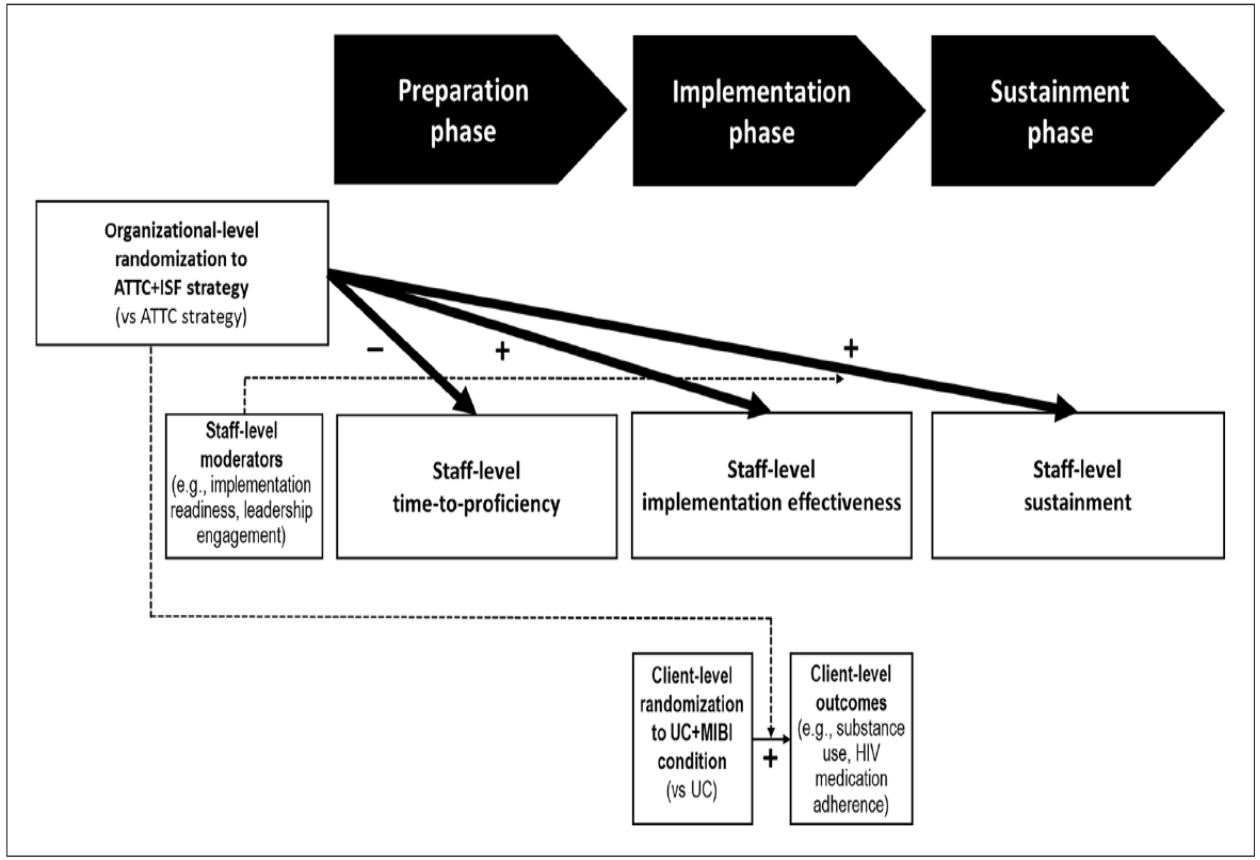
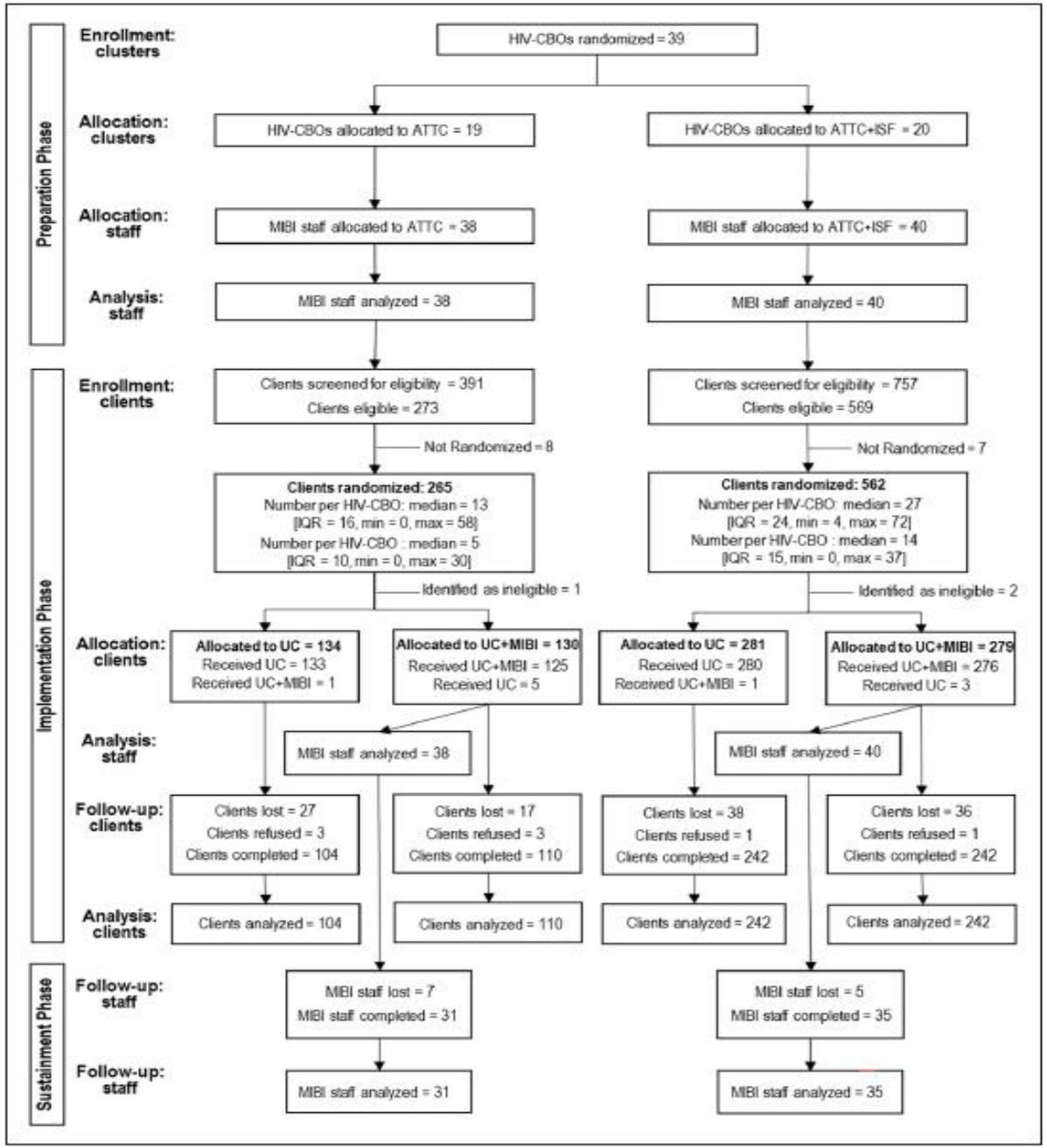


Figure 1. Aims and hypotheses.
Note. ATTC = Addiction Technology Transfer Center; ISF = implementation and sustainment facilitation; MIBI = motivational interviewing-based brief intervention; UC = usual care. Bolded lines indicate primary aim and hypotheses; thin line indicates other aim; and dashed lines indicate hypothesized moderators.

Garner, B. R., Gotham, H. J., Chaple, M., Martino, S., Ford, J. H., Roosa, M. R., Speck, K. J., Vandersloot, D., Bradshaw, M., Ball, E. L., Toro, A. K., Griggs, C., & Tueller, S. J. (2020). The implementation and sustainment facilitation strategy improved implementation effectiveness and intervention effectiveness: Results from a cluster-randomized, type 2 hybrid trial. Implementation Research and Practice, 1. https://doi.org/10.1177/2633489520948073



Targeted Sample size:

- 78 MIBI staff nested within 39 HIV organizations
- 1872 clients within 78 MIBI staff nested with 39 HIV organizations

Garner, B. R., Gotham, H. J., Chaple, M., Martino, S., Ford, J. H., Roosa, M. R., Speck, K. J., Vandersloot, D., Bradshaw, M., Ball, E. L., Toro, A. K., Griggs, C., & Tueller, S. J. (2020). The implementation and sustainment facilitation strategy improved implementation effectiveness and intervention effectiveness: Results from a cluster-randomized, type 2 hybrid trial. Implementation Research and Practice, 1. https://doi.org/10.1177/2633489520948073

Figure 2. Participant flow.
Note. ATTC = Addiction Technology Transfer Center; IQR = interquartile range; ISF = implementation and sustainment facilitation; MIBI = motivational interviewing-based brief intervention; UC = usual care.

Outcomes

STAFF-LEVEL OUTCOMES

- Time-to-proficiency
- Implementation Effectiveness
- Level-of-Sustainment

CLIENT-LEVEL OUTCOMES

- Days of primary substance use
- Number of substance-related problems,
- Times engaging in risky behaviors
- Days of substance use treatment
- Days of medication non-adherence

Considerations

- **Design driven by the research question**
- May consider a variety of different designs
  - Examples: Cluster randomized, individually randomized, stepped-wedge, quasi-experimental
- Need to consider impact of “real-world” when hypothesizing an effect size

“If we want more evidence-based practice, we need more practice-based evidence.”

-Lawrence W. Green (2006)

QUESTIONS?