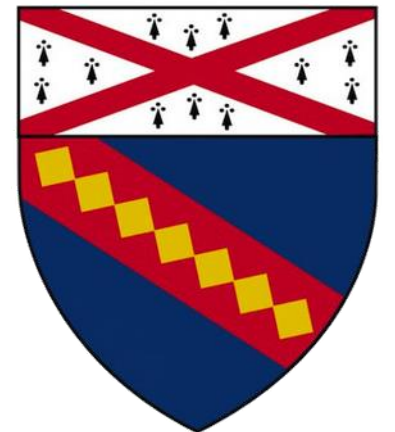


Choosing an Implementation Design After You Know the Strategy

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



Why is Study Design Particularly Important to Biostatisticians?

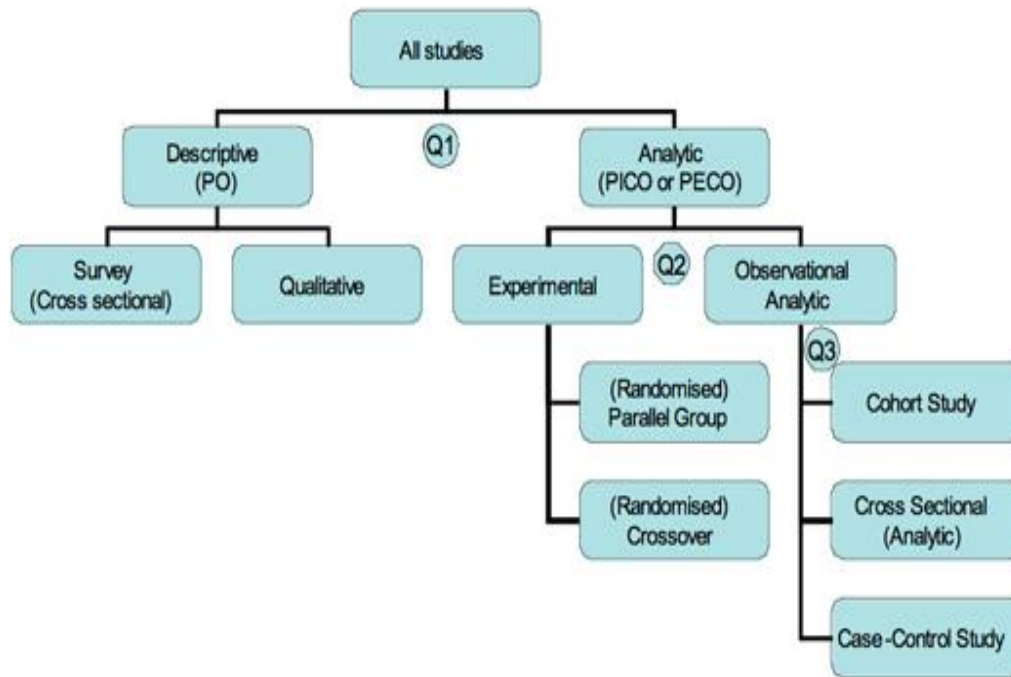
- The study design lays the foundation for collecting, analyzing, and interpreting data in a way that produces valid, reliable, and reproducible results.
- The quality of the study design directly impacts the statistical analysis and ultimately the conclusions that can be drawn from the study.



Your analysis is as good as your data.

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?

“In many ways the design of a study is more important than the analysis. A badly designed study can never be retrieved, whereas a poorly analysed one can usually be reanalysed. Consideration of design is also important because the design of a study will govern how the data are to be analysed.”

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

- **Population**
 - Who or what is the patient, population or problem in question?
- **Intervention**
 - What is the intervention being considered?
- **Control/Comparator**
 - What is the intervention being compared to?
- **Outcome**
 - What is the desired outcome?
- **Time**
 - How long will it take to reach the desired outcome?

Two Basic Design Strategies of Clinical Trials

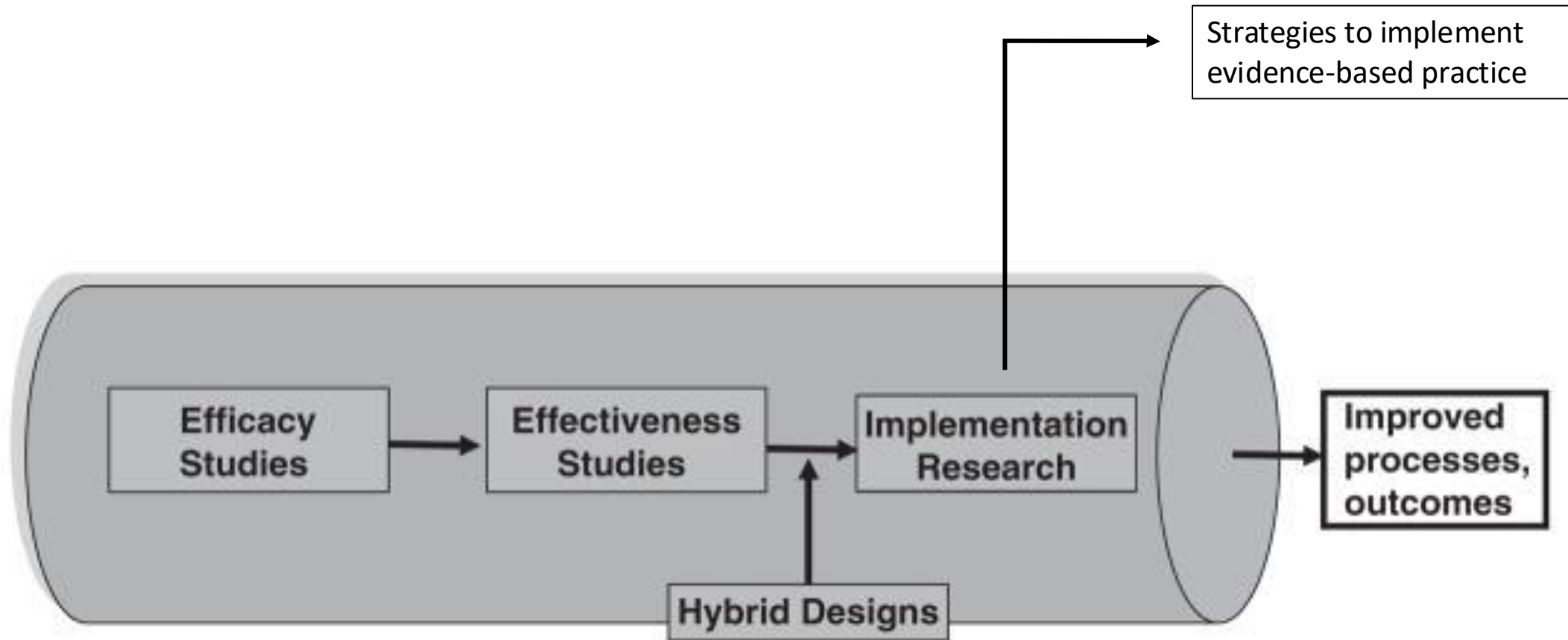
Pragmatic (Effectiveness)

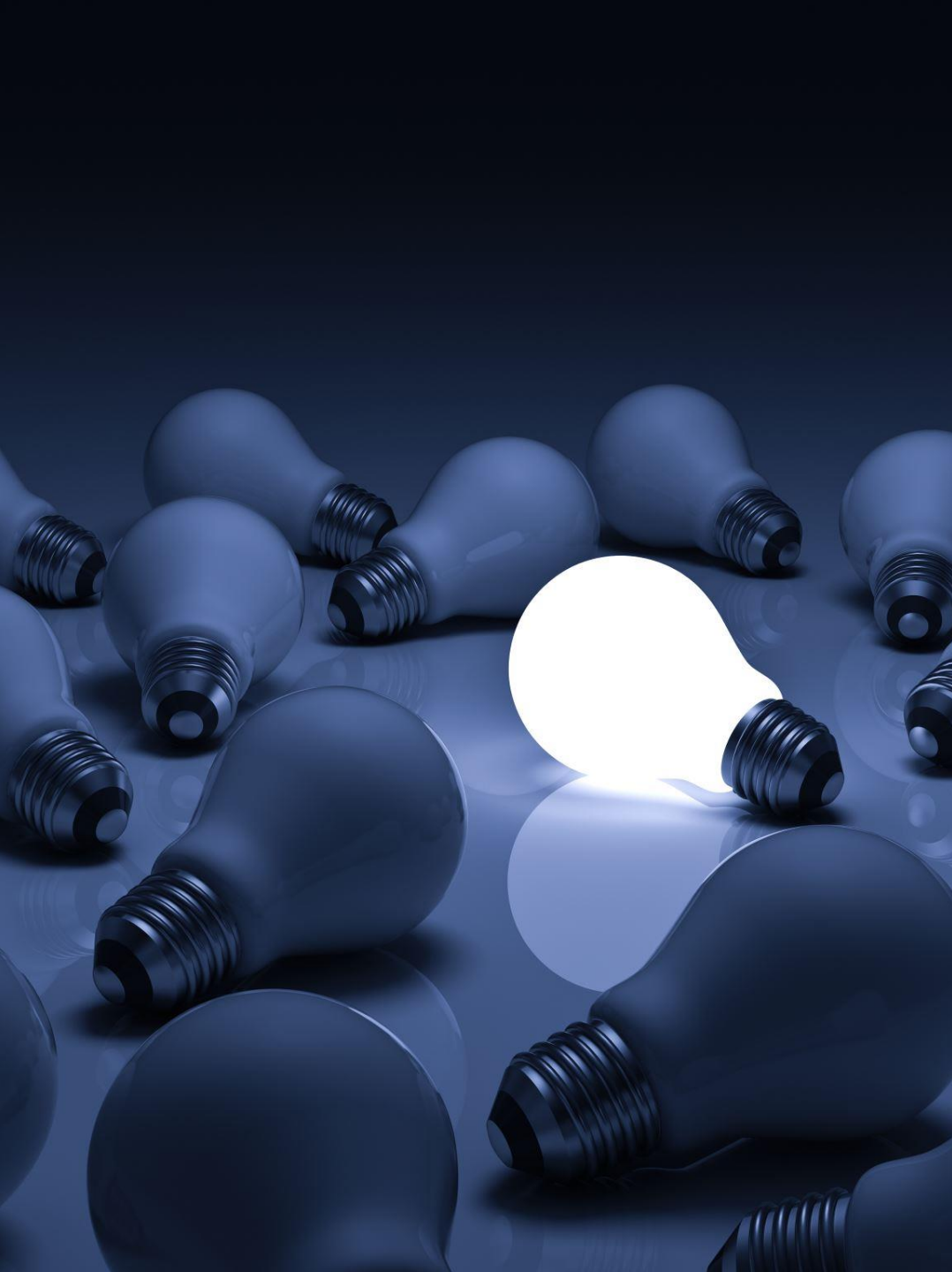
- Real world
- Community based patients
- Larger sample size
- Broader (heterogeneous) eligibility criteria
- Easily administered treatments
- Little monitoring of adherence and fidelity
- Patient centered outcomes: Mortality, morbidity, QoL
- Typical study: comparison of an intervention vs usual care

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

Research Pipeline





Implementation Science

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

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)

Threats to Internal Validity

Type of Bias	What is it?	Control for Bias
Selection	Knowing which patient/cluster is going to receive which treatment (and allocation)	Randomization with allocation concealment (i.e., blinded randomization)
Evaluator bias (soft outcome measures)	Knowledge of treatment when assessing outcome	Blinded evaluators
Information (e.g., ascertainment bias)	Differential ascertainment of outcomes in treatment arms	Comparable follow-up schedules; outcomes that are not subject to differential ascertainment

Bias

- Exists in all research and is difficult to fully eliminate
- Bias can occur at any stage of the research process
- Bias impacts validity and reliability and can lead to misinterpretation of the data and findings

Missing Data are a Threat to Valid Results



Bias in Clinical Studies

“Every clinical study has some bias. Nevertheless, their presence does not necessarily imply that a study should be disregarded. Rather, it is important to identify and assess the potential impact of biases on the conclusions of the study. “

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

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

Pragmatic Trial Designs (PCTs)

“Although PCTs do not necessarily require a specific design approach, both the kinds of questions PCTs are designed to answer and the settings in which they take place may tend to favor certain approaches, such as cluster randomization. The nature of the interventions PCTs seek to test, which may involved healthcare delivery changes, might be better implemented through randomization at the practice, clinic, or even hospital level.”

Hybrid Design Types

1

Primary aim is effectiveness; secondary aim is implementation

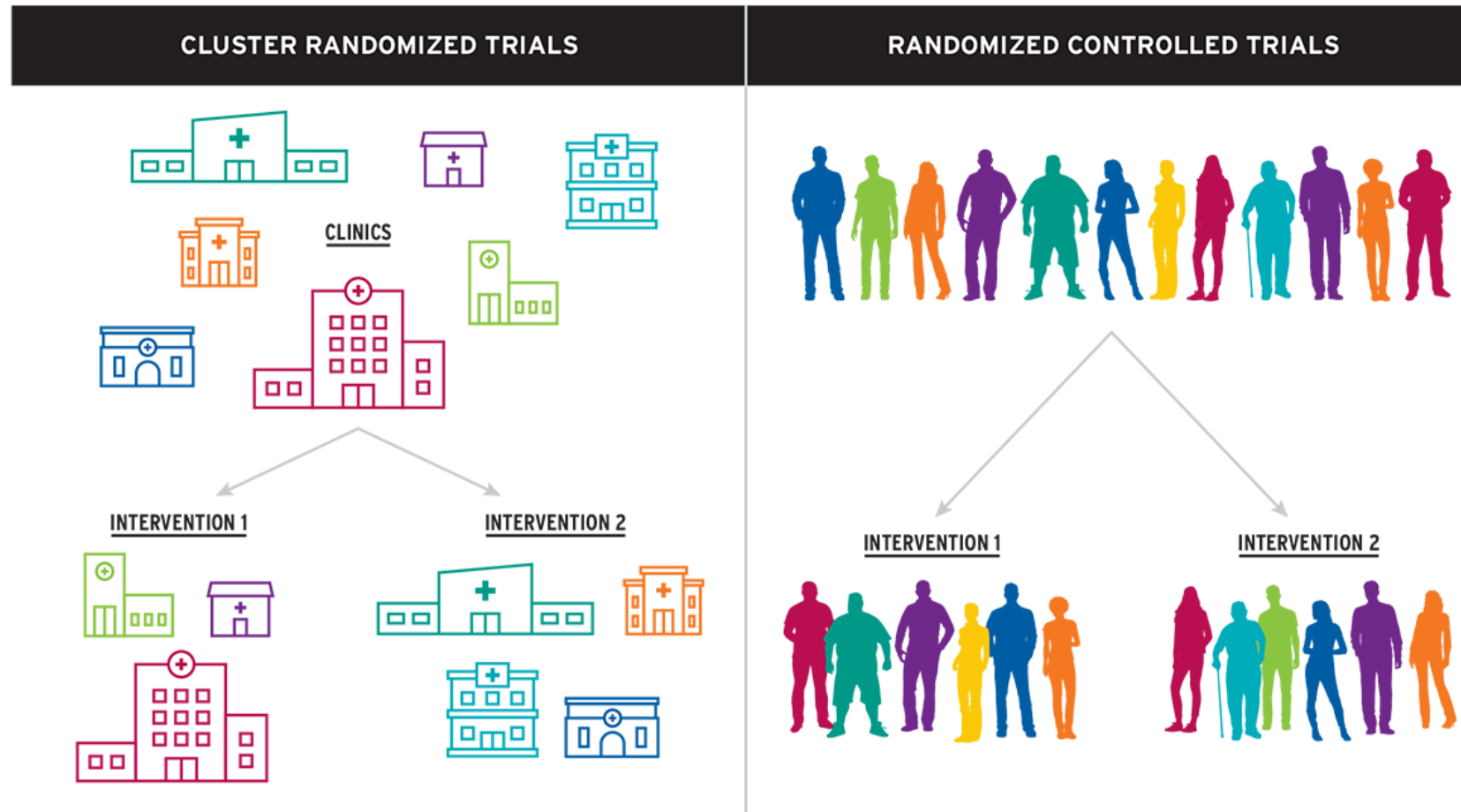
2

Co-primary aims are effectiveness and implementation

3

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

Experimental Designs



Quasi-Experimental Designs

Pre-Post Designs

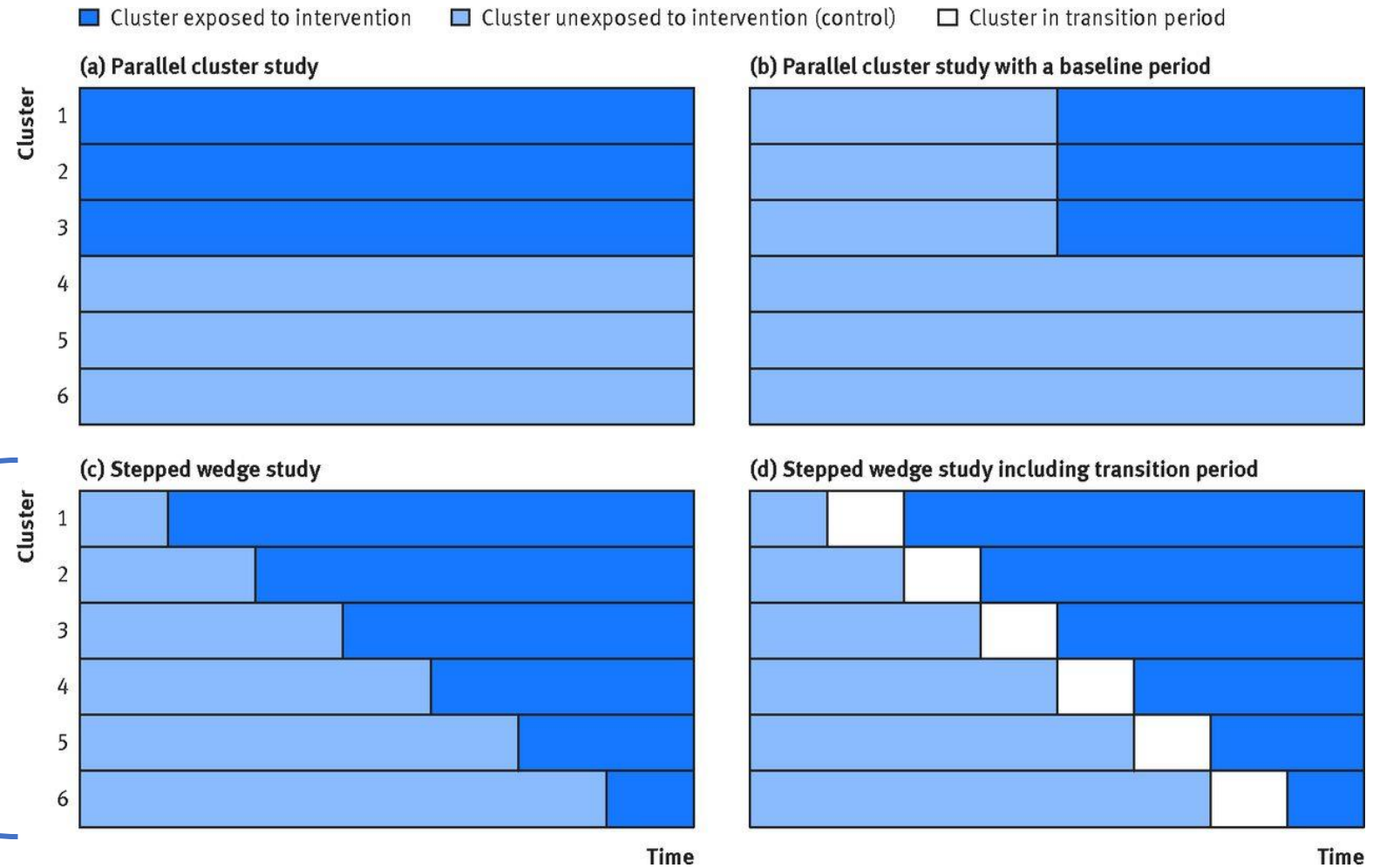
➤ Non-equivalent Control group

- 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

Cluster Randomized Design

Advantages

- Evaluation of interventions that cannot be delivered at the individual level
- Address potential contamination
- Mirror more closely how interventions are implemented at scale.

Disadvantages

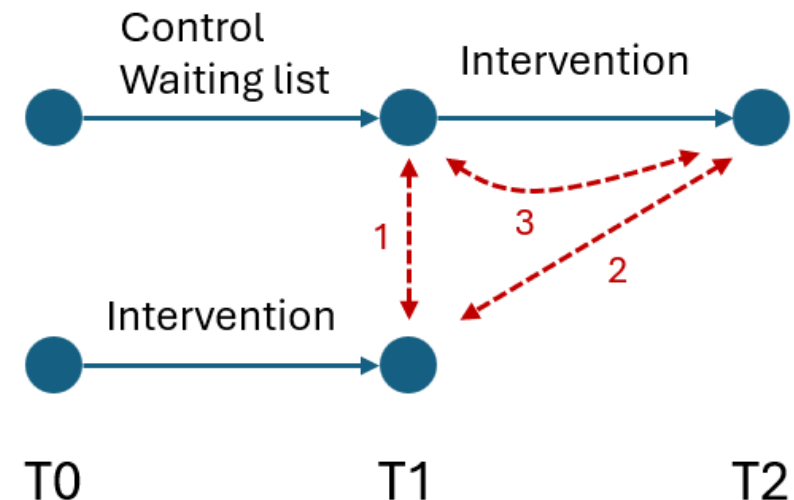
- Greater design (and analysis) complexity
- Larger sample size
- Potential for baseline imbalance (depending on sample size)

Contamination

- Contamination: Participants in the control condition likely to receive the intervention (or components of the intervention)
- Cluster randomization is often advocated to minimize contamination
 - Should use individual randomization if possible!
- Need to assess whether it is real or just theoretically possible
- Impact of contamination
 - Biases the treatment effect towards the null (increases chance of Type II error)
 - Assess impact – tradeoff between increased sample size due to reduced effect size versus potentially increased complexity AND sample size due to cluster randomization

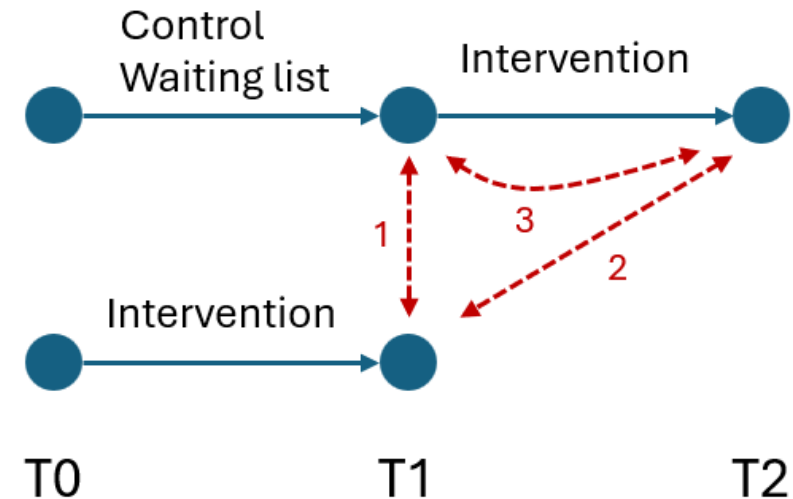
Waitlist Control

- Participants/clusters are originally randomized to not receive the experimental treatment
- Given treatment after the treatment group has completed the study.
- Why might one use this type of design?



Some Hypotheses to Test

1. Difference in means between intervention and control at time T1.
2. Difference in means between control at T2 and intervention at T1
3. Difference in means between control at T1 and T2 (within-group)



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

Partial Cluster Designs

- Also known as Individually Randomized Group Treatment (IRGT) trials or partially nested trials
- Individual level randomization, but clustering is introduced after randomization

Examples: Behavioral interventions, surgical trials

- Clustering could be introduced in one or all the arms in the study

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

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

- FRACTIONAL FACTORIAL

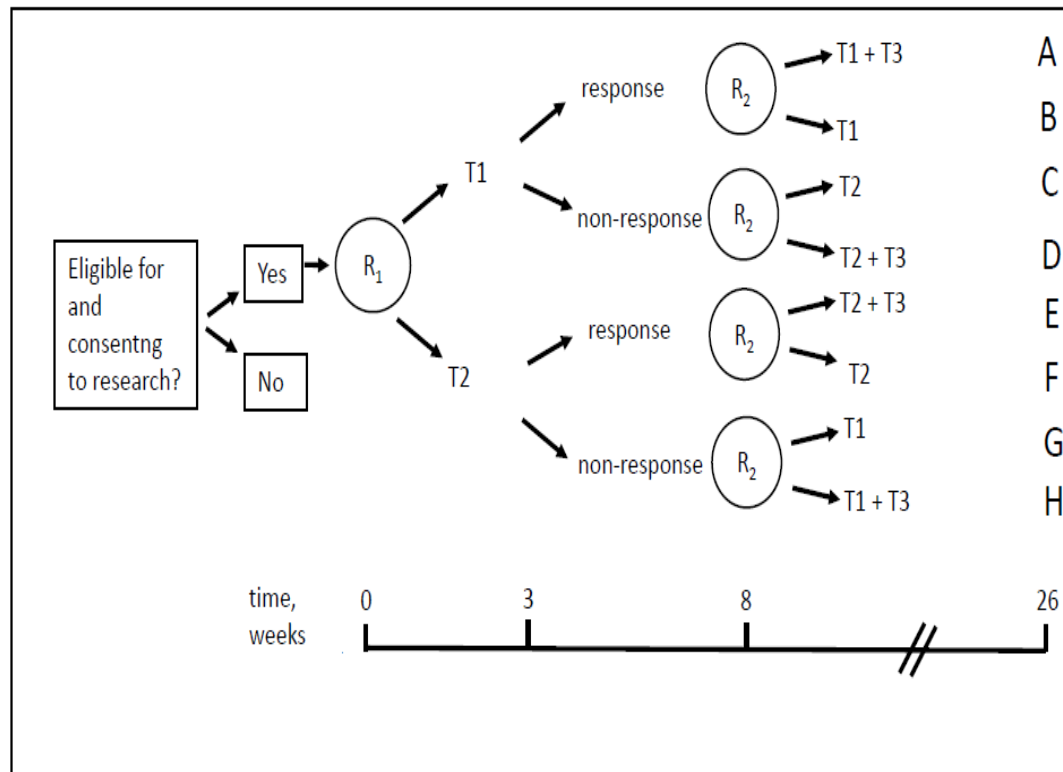
Run	Treatment	Factor		
		A	B	C
1	c	-1	-1	+1
2	a	+1	-1	-1
3	b	-1	+1	-1
4	abc	+1	+1	+1

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

Examples of Adaptive Designs

- SMART (Sequential, Multiple Assignment Randomized Trial)
 - Optimizing treatment sequences over time
 - Some or all participants are randomized more than once
- Stepped Care Designs
 - Only one level of randomization
 - Participants that do not achieve the interim outcome are stepped up to get more

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	

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

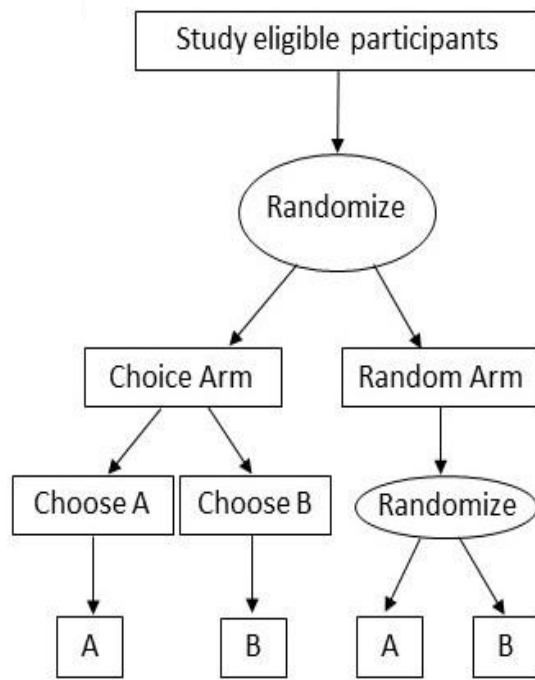
Can get evidence regarding interactions among multiple treatments

CHALLENGES/LIMITATIONS

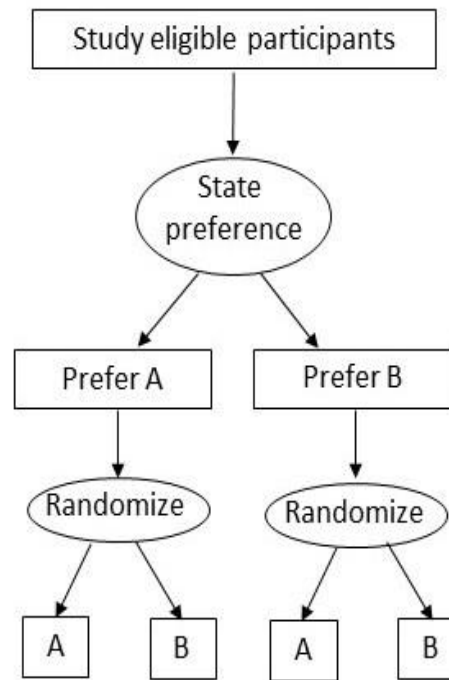
- Requires more planning
 - Disruptions can have a major impact on the findings
- Generally, does not provide enough evidence to definitively establish effectiveness

Patient Preference Designs

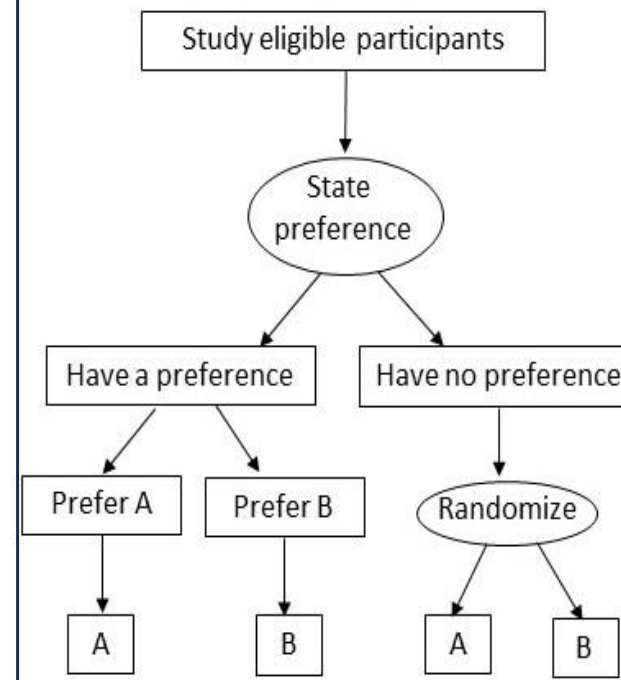
(a) Two-stage randomized clinical trial



(b) Fully randomized design



(c) Partially randomized preference design



Unit of Randomization vs Unit of Measure

- Unit of randomization
 - Level at which random assignment occurs
 - Examples:
 - Individual, household, classroom, school
- Unit of Measure
 - Level at which data are collected and analyzed
 - Examples
 - Individual (blood pressure), household (income), classroom (average test score), school (attendance)
- There can be a mismatch between the unit of randomization and the unit of measure

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

Endpoint vs outcome

- Outcome: Measured variable
 - Examples: Blood Pressure; Quality of Life; Depression
- Endpoint: Analyzed parameter
 - Example: Change from baseline in mean Health Related Quality of Life at 6-months
- Good characteristics
 - Relevant
 - Hard (objective)
 - Easily Observed
 - Observed without Error
- Determining the best endpoint to establish the effectiveness of an intervention may not be a trivial task
 - True especially if trying to use routinely collected data in the context of clinical practice.

How do we decide best endpoint and outcomes?

- For pragmatic trials
 - Meaningful to key stakeholders (e.g., participants, funders, communities, clinicians)
 - Example: Progression free survival may be of little importance to cancer patients if overall survival is not improved
 - Relatively easy to collect using routinely collected healthcare data
 - Example: Hospitalizations
 - Available in routinely collected healthcare data
 - Not designed for research purposes – clinical care and reimbursement
- Things to think about:
 - What challenges do you anticipate trying to ascertain the endpoint?
 - How might you address this challenge?
- A “less pragmatic” trial on outcome might be necessary

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

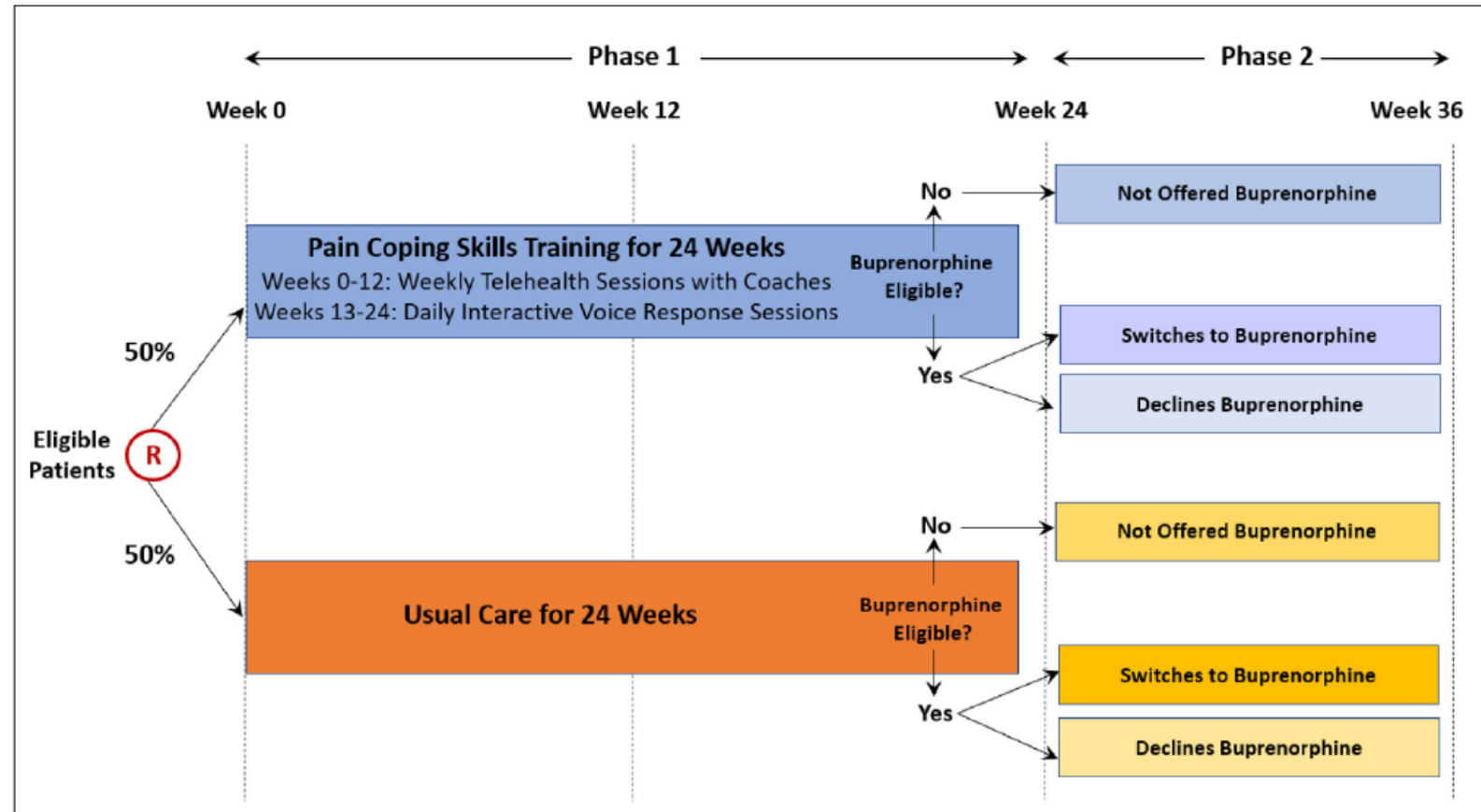
Sample size considerations

- 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

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)

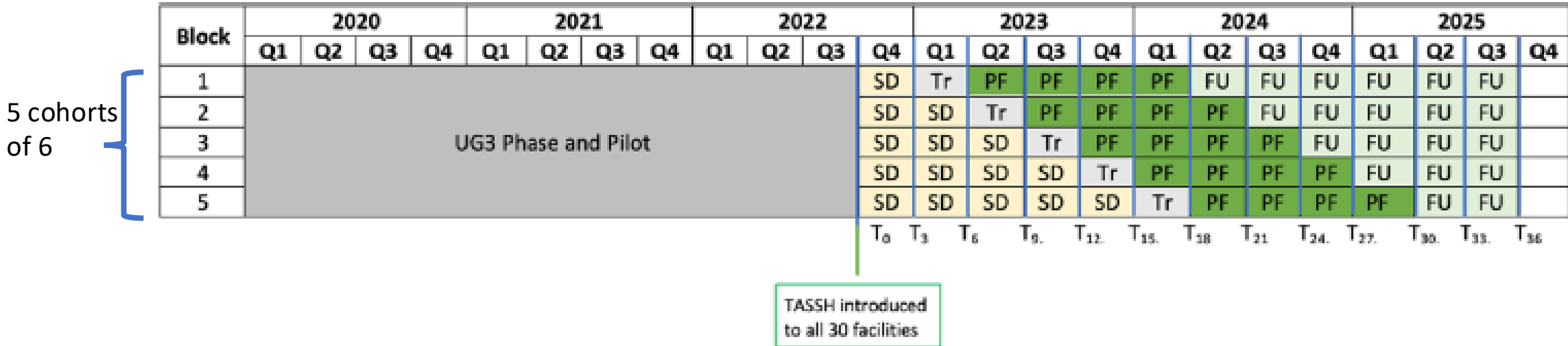




Managing Hypertension Among People Living with HIV: an Integrated Model (MAP-IT) trial

- 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



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’.

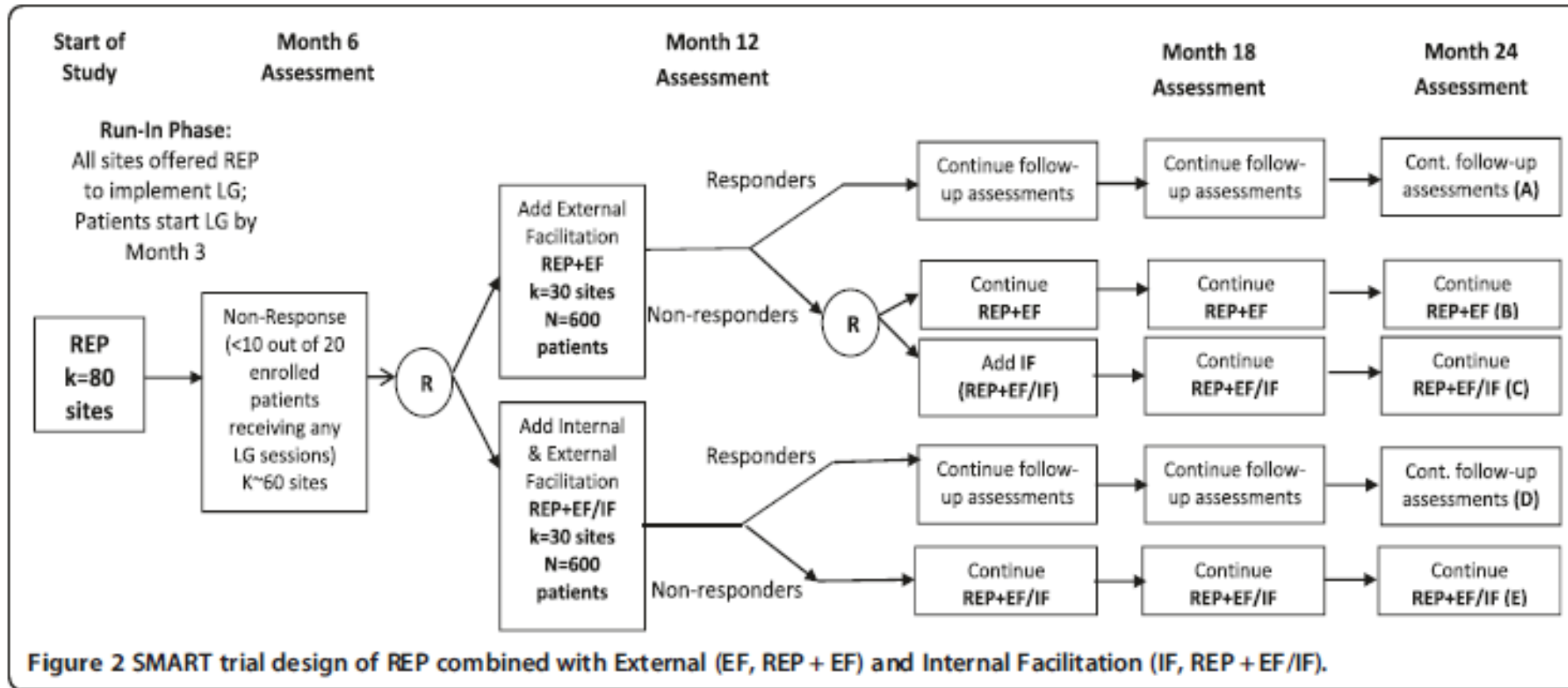
Fig. 2 Stepped wedge study design



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

Cluster Randomized SMART



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

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.

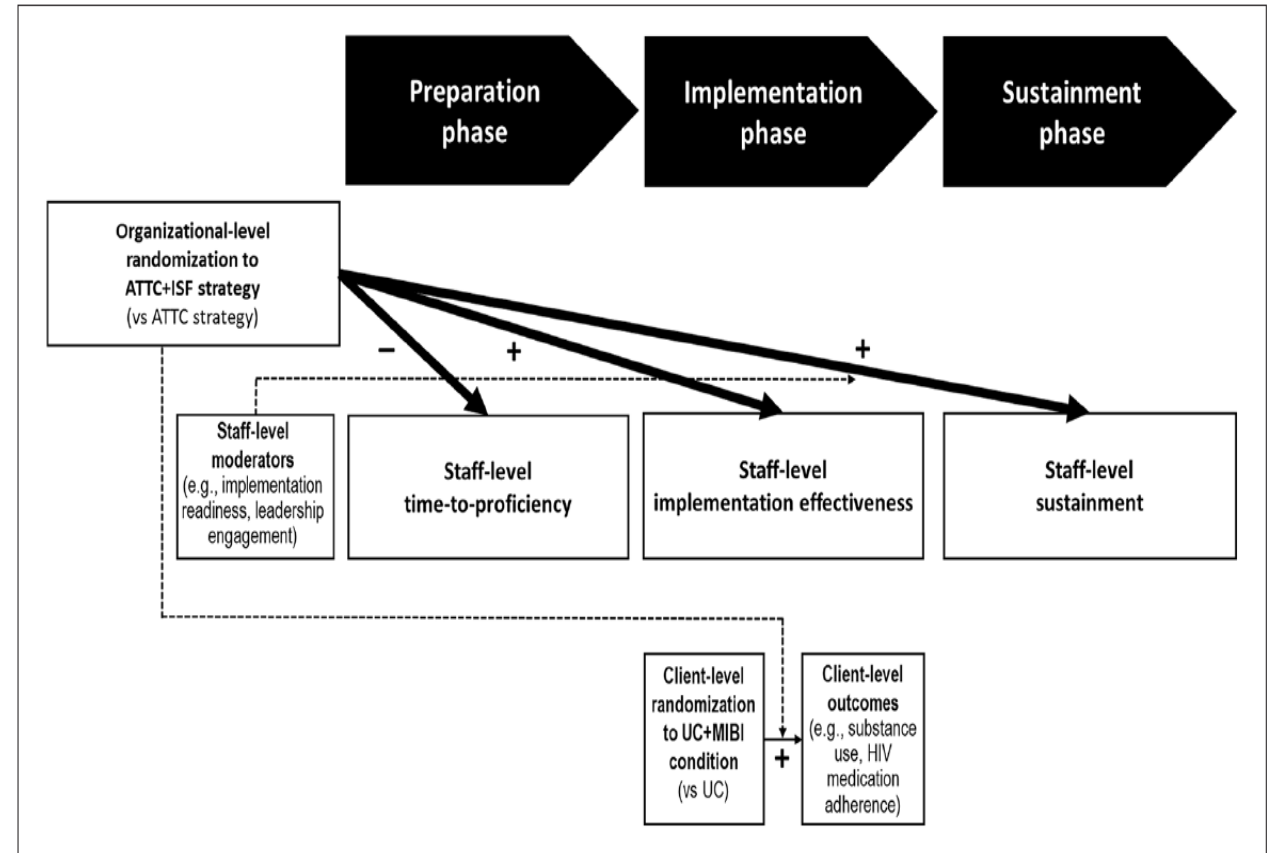
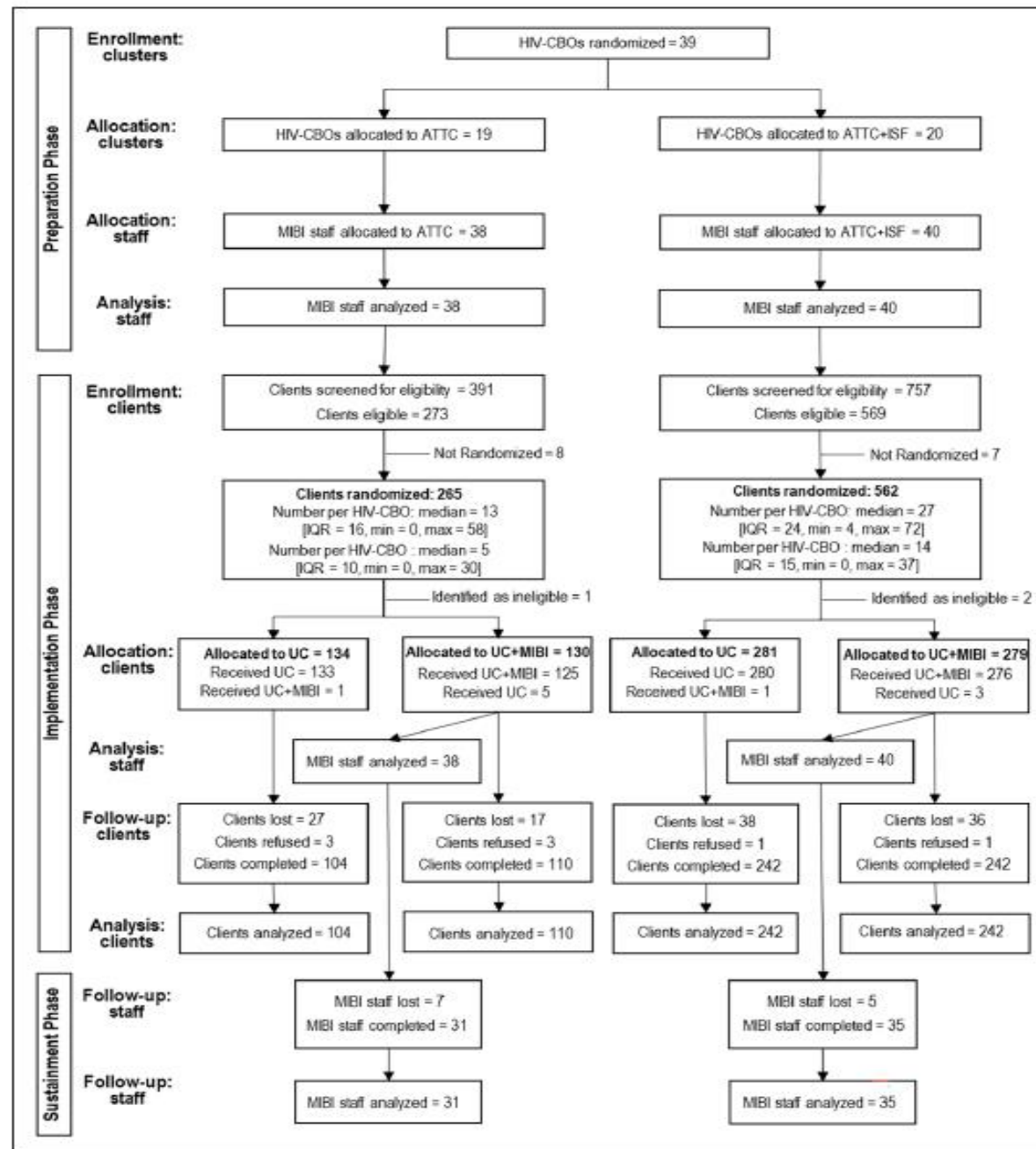


Figure 1. Aims and hypotheses.

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.



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

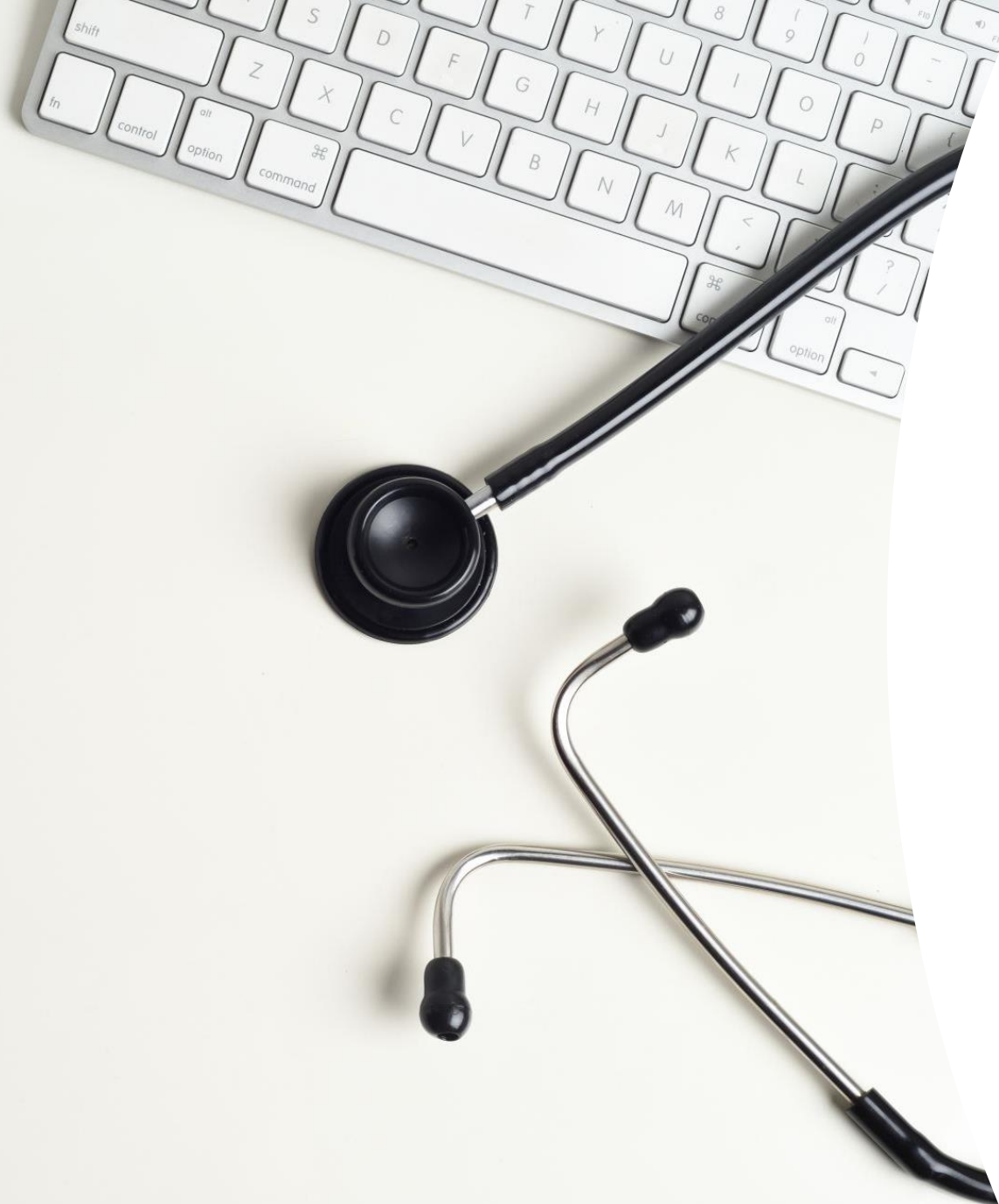
- Time-to-proficiency
- Implementation Effectiveness
- Level-of-Sustainment

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?