The Estimands Framework

Aligning Objectives, Endpoints, and Analyses

Moderator:
Liz Turner, PhD
Duke University



NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®

Session Goals

- Define the estimands framework and its importance for pragmatic trials.
- Address populations in ePCTs.
- Discuss strategies for handling intercurrent events.
- Share examples of applying the estimands framework in NIH Collaboratory Trials.

Participants



Fan Li
Yale School of Public Health



Hulin Wu
UTHealth Houston



Elizabeth Colantuoni
Johns Hopkins University



Richard Skolasky, Jr
Johns Hopkins University

NIH Collaboratory Trial Designs

Trial Design

14 parallel group with individual randomization

16 parallel group with cluster randomization

7 stepped-wedge cluster randomization

Intervention Type



13 therapeutic



14 operational/ educational



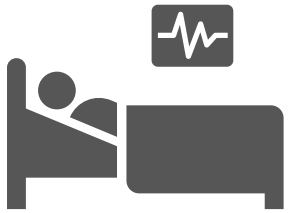
10 mix

Intervention Target

22 patient

9 clinician

6 patient and clinician





NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Estimands Framework: A Brief Introduction

Fan Li, Yale School of Public Health
NIH Collaboratory Workshop, September 8, 2026

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Estimands make the treatment question explicit

Estimand = a precise description of the treatment effect a study aims to quantify.

Why it matters (Cro et al. 2020, BMJ)

98%
did not describe what the reported treatment effect represented

54%
had methods from which the exact estimand could not be deduced

- **A statistical method does not uniquely define the scientific question.**
- Methods can implicitly target a subset or a hypothetical world; prespecification aligns the design and analysis.

State the question first—do not make readers reverse-engineer it from the model.

A brief history

1. **Long-standing idea**
   The causal target—the quantity the study is trying to estimate.
2. **2019: ICH E9(R1)**
   Formalized a structured five-attribute framework, primarily for individually randomized trials.
3. **2026: CRT extension**
   A consensus framework adapts the same structure to cluster-randomized trials.

Sources: Kahan et al., Am J Epidemiol 2023; Kahan et al., BMJ 2026.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Five attributes for an individually randomized trial

For each trial outcome, specify all five attributes before choosing the estimator.

Attribute	What it makes explicit
1 **Population**	The individuals for whom the treatment effect is defined—not necessarily identical to the eligibility or analysis population.
2 **Treatment conditions**	The treatment versions being compared, including how discontinuation, switching, rescue therapy, and adherence are conceptualized.
3 **Endpoint / outcome variable**	The outcome, measurement scale, and time frame that answer the scientific question.
4 **Intercurrent-event strategy**	How each post-baseline event that changes the endpoint's interpretation or existence is handled.
5 **Population-level summary measure**	How outcomes are summarized and contrasted across treatment conditions (for example, a mean difference, risk ratio, or odds ratio).

The estimand describes the question; the estimator is the statistical method used to estimate it.

Adapted from ICH E9(R1) and Kahan et al., Am J Epidemiol 2023.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Intercurrent events change the treatment question

Intercurrent event: an event after baseline that affects the interpretation or existence of the endpoint—such as discontinuation, switching, rescue therapy, or death.

Strategy	Scientific question	How the endpoint enters
Treatment policy	What is the effect regardless of the event?	Use the endpoint whether or not the event occurs.
Composite	Should the event itself count as part of the outcome?	Incorporate the event into the endpoint definition.
While on treatment / while alive	What happened up to the event?	Use the endpoint before the event occurs.
Hypothetical	What would the effect be if the event did not occur?	Use the outcome under a precisely specified hypothetical scenario.
Principal stratum	What is the effect in a subgroup defined by potential event status?	Define the target population using potential occurrence of the event.

The strategy is part of the estimand—not a missing-data afterthought. A treatment-policy strategy cannot supply an outcome that no longer exists.

Sources: Kahan et al., Am J Epidemiol 2023; Kahan et al., BMJ 2026.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Extending the five attributes to cluster-randomized trials

Consensus extension of ICH E9(R1): **three attributes were modified; no new attributes were added.**

ICH E9(R1) attribute	CRT specification	What must now be explicit
Population of patients	**Population of individuals and clusters**	Target both levels; address post-randomization recruitment or selection when relevant.
Treatment conditions	Treatment conditions	Cluster implementation, individual exposure, and the exposure duration of interest.
Endpoint / outcome variable	Endpoint / outcome variable	Whether the endpoint is defined at the individual or cluster level, plus its scale and time frame.
Strategies for intercurrent events	**Strategies for individual-level and cluster-level intercurrent events**	Use the relevant baseline for each level; include cluster nonimplementation and movement between clusters.
Population-level summary measure	**Summary measure, including weighting and marginal versus cluster-specific interpretation**	Who counts equally—individuals, clusters, or cluster-periods—and, for non-collapsible measures, how the effect is interpreted.

Applies to parallel-group, cluster-crossover, and stepped-wedge trials.

Source: Kahan et al., BMJ 2026, CRT-Estimands Framework.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

What changes in a cluster-randomized trial?

Dimension	Individually randomized trial	Cluster-randomized trial
Randomization and baseline	Individuals are randomized; one natural baseline.	Clusters are randomized. Cluster baseline is randomization; individual baseline is enrollment or eligibility.
Target population	The treatment effect is defined for individuals.	The effect is defined for individuals within target clusters; assignment may affect who is recruited after randomization.
Treatment delivery	Treatment is usually assigned and delivered at the individual level.	Cluster implementation and individual uptake can both matter; intended exposure duration may vary.
Intercurrent events	Usually participant-level events such as discontinuation, switching, rescue therapy, or death.	Events can occur at either level: cluster nonimplementation, people leaving or switching clusters, or clusters merging or splitting.
Population summary	Individuals typically receive equal weight; a population-average interpretation is common.	Specify participant-average versus cluster-average (or other) weighting and, for non-collapsible measures, marginal versus cluster-specific interpretation.

These choices can change the answer: participant-average = 1.4 versus **cluster-average = 3.0** in the BMJ illustrative example with informative cluster size.

Source: Kahan et al., BMJ 2026, Boxes 2, 6, and 7.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Start with the question, then choose the analysis

1

Specify the estimand

For each outcome, write down all five attributes.

2

Choose an analysis that targets it

Align the design, trial conduct, data collection and estimator.

3

Check the assumptions

State assumptions and assess robustness with sensitivity analyses.

Before analysis, confirm the estimator targets the prespecified estimand.

NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®

Some references

American Journal of Epidemiology
Vol. 192, No. 6
https://doi.org/10.1093/aje/kwad036
Advance Access publication: February 14, 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Practice of Epidemiology

Eliminating Ambiguous Treatment Effects Using Estimands

Brennan C. Kahan*, Suzie Cro, Fan Li, and Michael O. Harhay

* Correspondence to Dr. Brennan C. Kahan, MRC Clinical Trials Unit at UCL, University College London, 90 High Holborn, London WC1V 6LJ, United Kingdom (e-mail: b.kahan@ucl.ac.uk).

Initially submitted August 25, 2022; accepted for publication February 13, 2023.

Most reported treatment effects in medical research studies are ambiguously defined, which can lead to misinterpretation of study results. This is because most authors do not attempt to describe what the treatment effect represents, and instead require readers to deduce this based on the reported statistical methods. However, this approach is challenging, because many methods provide counterintuitive results. For example, some methods include data from all patients, yet the resulting treatment effect applies only to a subset of patients, whereas other methods will exclude certain patients while results will apply to everyone. Additionally, some analyses provide estimates pertaining to hypothetical settings in which patients never die or discontinue treatment. Herein we introduce *estimands* as a solution to the aforementioned problem. An estimand is a clear description of what the treatment effect represents, thus saving readers the necessity of trying to infer this from study methods and potentially getting it wrong. We provide examples of how estimands can remove ambiguity from reported treatment effects and describe their current use in practice. The crux of our argument is that readers should not have to infer what investigators are estimating; they should be told explicitly.

estimands; estimates; estimators; randomized trials; treatment effects

Abbreviation: ICH-E9(R1), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use Efficacy Guideline 9, Revision 1.

OPEN ACCESS
Check for updates

CRT-Estimands Framework: consensus based extension of the ICH E9(R1) addendum for cluster randomised trials

Brennan C Kahan,[1] Frank Bretz,[2,3] Andrew Copas,[1] Michael O Harhay,[1,4] Gary S Collins,[5] Melanie Bahti,[6] Dongquan Bi,[1] Marion K Campbell,[7] Katherine R Courtright,[4,6] Timothy Feeney,[8,9] Andrew B Forbes,[10] Amit X Garg,[11] Scott D Halpern,[4,6] Patrick J Heagerty,[12] Karla Hemming,[13] Sally Hopewell,[14] James P Hughes,[12] David M Murray,[15] Joseph S Ross,[16,17] Derek Stewart,[18] Monica Taljaard,[19,20] Obioha C Ukoumunne,[21] Catherine L Auriemma,[6] Fan Li[22]

For numbered affiliations see end of the article
Correspondence to: B C Kahan b.kahan@ucl.ac.uk; (ORCID 0000-0001-9957-0844)
Additional material is published online only. To view please visit the journal online.
Cite this as: *BMJ* 2026;393:e089050
http://dx.doi.org/10.1136/bmj-2025-089050
Accepted: 06 March 2026

In order to assess the relevance of trial results or the appropriate trials methods, interest holders (eg, clinicians, patients, and policy makers) need a clear description of the trial's research question. To improve clarity and consistency in clinical trials, the International Conference on Harmonisation (ICH) released the ICH E9(R1) addendum, which set out a framework for defining estimands—a precise description of the treatment effect to be estimated. While the ICH E9(R1) addendum has been widely adopted, it primarily focused on individually randomised trials. In contrast, cluster randomised trials, where groups of individuals are randomised, present additional challenges for defining estimands. Therefore, the CRT-Estimands Framework was developed as a consensus based extension of the ICH in cluster randomised trials, with the objective of improving the clarity of the estimands, and consequently, the research questions, in these trials. This article presents the CRT-Estimands Framework with explanations and examples of how it can be implemented. Adopting the framework will improve the clarity of estimands in cluster randomised trials and facilitate interest holders to make informed decisions from these trials.

A clearly defined research question is essential to ensure the appropriate interpretation of a randomised trial. Firstly, clarity around the research question helps trial investigators to design their trial using appropriate methodology. Secondly, it helps interest holders (including patients, clinicians, funders, regulators, health technology assessment agencies, and journal editors) to appraise the trial, by assessing the question's relevance and by evaluating whether the trial's methods, including its design and analysis, are appropriate for the chosen question. However, despite its critical importance, historically many trials do not sufficiently describe their research question.[1 2]

Statistics in Medicine

WILEY
Statistics in Medicine

COMMENTARY

What Makes an Estimand Useful? Guidance on the Choice of Intercurrent Event Strategies

Brennan C. Kahan[1] | Fan Li[2,3] | Michael O. Harhay[1,4] | Suzie Cro[5]

[1]MRC Clinical Trials Unit at UCL, London, UK | [2]Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut, USA | [3]Center for Methods in Implementation and Prevention Science, Yale University School of Public Health, New Haven, Connecticut, USA | [4]Center for Clinical Trials Innovation, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA | [5]Imperial Clinical Trials Unit, School of Public Health, Imperial College London, London, UK

Correspondence: Brennan C. Kahan (b.kahan@ucl.ac.uk)

Received: 25 September 2025 | **Revised:** 6 December 2025 | **Accepted:** 6 February 2026

Keywords: choice of intercurrent event strategy | composite | estimand | hypothetical | principal stratum | treatment policy | while-alive

ABSTRACT

While the use of estimands in randomized trials is increasing, there is little guidance on which intercurrent event strategies should be used. The article by Fleming et al. seeks to address this gap. They argue that strategies such as hypothetical, principal stratum, and while-alive generally cannot be used to reliably inform decision making, and that treatment policy (and composite for mortality) strategies should be used instead. In this Commentary we argue that there are a variety of settings where strategies such as hypothetical, principal stratum, and while-alive can reliably inform decision-making and are preferable to a treatment policy strategy. We provide an alternative approach for selecting intercurrent event strategies, which systematically considers the trade-off between relevance (whether it addresses a useful question) and reliability (the ability to be estimated such that stakeholders can have confidence in the results) of each strategy in order to identify those that can be used to robustly inform decision-making. Our overall conclusion is that there is no single intercurrent event strategy that is appropriate in all settings; all strategies can be beneficial when used in appropriate settings, but harmful when used in inappropriate settings.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Estimands Framework: A Brief Introduction Supplemental Slides

Fan Li, Yale School of Public Health

NIH Collaboratory Workshop, September 8, 2026

NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®

Supplemental Slides



NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®

Illustrative hospital cluster-crossover trial

Context, population, treatment conditions, and endpoint

Randomized clusters
Hospitals

Design
Alternating standard and enhanced cleaning periods.

Outcome unit
Patients admitted to previously contaminated single-patient rooms.

	Attribute	Illustrative specification
1	**Population of individuals and clusters**	Patients admitted to previously contaminated single-patient rooms in tertiary, community, or Veterans Affairs hospitals.
2	**Treatment conditions**	Standard terminal disinfection: bleach Enhanced terminal disinfection: bleach and ultraviolet light
3	**Endpoint / outcome variable**	*Clostridioides difficile* infection during hospital admission.

Teaching example adapted from the BETR example in Table 3 of Kahan et al. (BMJ 2026); some details changed.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Intercurrent events in the example

Intercurrent event: an event after the relevant baseline that affects the interpretation or existence of the endpoint. **Here, the events are room-level implementation deviations in ultraviolet-light use.**

4 Strategies to handle intercurrent events

	Enhanced period	Standard period
Intended disinfection	**Bleach and ultraviolet light**	**Bleach only**
Intercurrent event	**Ultraviolet light not used**	**Ultraviolet light used**

Why this is intercurrent: the deviation changes which treatment version the patient's infection endpoint represents, so the estimand must say how that endpoint is handled.

What outcome should represent the patient if either deviation occurs?

Source: Kahan et al., BMJ 2026, Table 3.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Hypothetical strategy: correct implementation

4 Strategies to handle intercurrent events

Scientific question: What would the difference in infection proportions have been if both disinfection strategies had been implemented exactly as intended?

Enhanced period
UV not used → **Use the outcome that would have been observed if ultraviolet light had been used.**

Standard period
UV used → **Use the outcome that would have been observed if ultraviolet light had not been used.**

Target comparison: enhanced versus standard terminal disinfection, if both are implemented correctly.

The hypothetical strategy is part of the estimand: it defines the treatment versions and outcomes being compared.

Population-level summary: each hospital counts equally

5 Population-level summary measure

Hospital 1	Hospital 2	…	Hospital J
Equal weight	Equal weight		Equal weight

Construction: within each hospital, compute the *Clostridioides difficile* infection proportion under each correctly implemented strategy; take the difference, then average across hospitals.

Weighting	Each hospital contributes equally, regardless of its number of eligible patients.
Effect measure	Cluster-average difference in infection proportions.
Interpretation	A difference in proportions is collapsible, so no separate marginal versus cluster-specific choice is required.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Putting the five attributes together

Attribute	Illustrative specification
1 **Population**	Patients admitted to previously contaminated single-patient rooms in tertiary, community, or Veterans Affairs hospitals.
2 **Treatment conditions**	Standard terminal disinfection (bleach) versus enhanced terminal disinfection (bleach + ultraviolet light), if implemented correctly.
3 **Endpoint**	*Clostridioides difficile* infection during hospital admission.
4 **Intercurrent-event strategy**	Hypothetical strategy for ultraviolet light omitted in an enhanced period and ultraviolet light used in a standard period.
5 **Population-level summary**	Cluster-average difference in infection proportions; hospitals weighted equally.

Estimand in one sentence

The estimand is the cluster-average difference in proportions of *Clostridioides difficile* infection between standard terminal disinfection (bleach) and enhanced terminal disinfection (bleach and ultraviolet light), if implemented correctly (ultraviolet light always used for enhanced and never for standard disinfection), in patients admitted to previously contaminated single-patient rooms in tertiary, community, or Veterans Affairs hospitals.

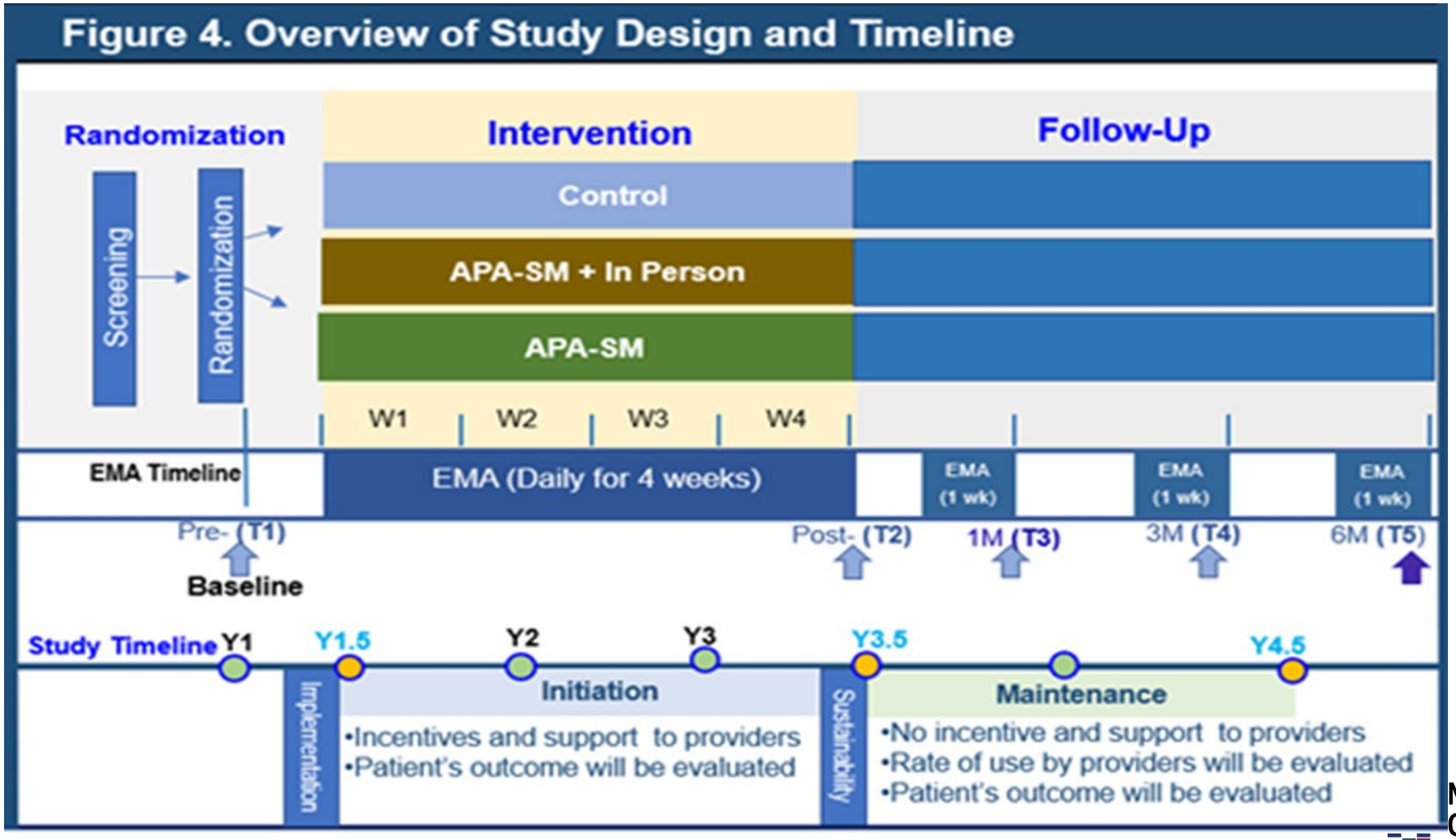
Adapted from the BETR example and summary estimand in Table 3 of Kahan et al., BMJ 2026.

The Estimand Framework: APA-SM Study

Hulin Wu, PhD, Betty Wheless Trotter Professor
Department of Biostatistics and Data Science
School of Public Health
University of Texas Health Science Center at Houston

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Personalized Auricular Point Acupressure for Chronic Pain Self-Management in Rural Populations



NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Strategies for Addressing Intercurrent Events

Potential intercurrent events:

- New alternative pain medications and pain treatments
- New analgesic use
- Discontinuation or interruption of protocol treatments before the protocol intervention ends
- Unrelated death before the protocol intervention ends

Strategies:

- Treatment policy strategy:
  - Discontinuation or interruption of protocol treatments
- While on treatment strategies:
  - New alternative pain medications and pain treatments
  - New analgesic use
  - unrelated death
- Principal stratum strategies: Sensitivity analyses

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Estimand Attributes

- Treatments: APA treatment and education control—allow to add-on to the existing pain treatments or care
- Population: 18 years or older in rural area with chronic CMP≥4 on a scale of 0-10—mITT population for primary analysis
- Variable (primary endpoint): Pain impact (3-item PEG) change from baseline to post-intervention (4 weeks).
- Intercurrent events: see above
- Population-level summary: Compare the mean PEG score change from baseline to post-intervention between APA treatment and education control using MMRM

Key Challenges: Discussion

- Primary outcome measurements: 3-times-a-day EMA vs. 7-day recall
- Primary endpoint: Two time point change (from baseline to Day 28) vs. Average effect during the whole intervention period (28 days)
- Analysis: Modeling of all longitudinal data vs. two-time point change
- Sample size justification: Modeling of all longitudinal data vs. two-time point change

ARBOR-Telehealth Study

Richard Skolasky, MA, ScD
Professor
Departments of Orthopaedic Surgery and Physical Medicine and Rehabilitation
The Johns Hopkins University School of Medicine

Elizabeth Colantuoni, PhD
Research Professor
Department of Biostatistics
Bloomberg School of Public Health, Johns Hopkins University

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

TidalHealth · Johns Hopkins Medicine · MedStar Health · NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases

ARBOR TELEHEALTH

ARBOR-Telehealth

Advancing Rural Back pain Outcomes using Rehabilitation-Telehealth

434
Participants randomized

8 weeks
Physical therapy (PT) / Education

IRGT
Individually randomized, group treatment

RE-AIM
Implementation framework

Study design & timeline

Primary care visit for chronic LBP -or- Self-referral from rural community → Randomized (1:1, stratified) → Telehealth PT / Education control (10 modules) → 12 & 26-wk assessments → 52-wk assessment

Enrolled: Jan 2025 – Oct 2027

Both arms run over 8 weeks; educational control includes 10 self-management modules available throughout

Study aims

- 1. Reduce LBP-related disability (primary, ODI)
- 2. Reduce opioid use (secondary)
- 3. Examine heterogeneity of treatment effect
- 4. Implementation outcomes (RE-AIM)

Eligibility & setting

- Adults 18+ with chronic LBP (ODI ≥ 24%, NRS ≥ 4/10)
- Recruited via primary care referral or community outreach
- Rural communities across Maryland & Delaware
- Partnership: TidalHealth & Johns Hopkins

MPIs: Richard L. Skolasky, ScD & Kevin McLaughlin, DPT | NIAMS UH3-AR083838

Estimand Framework – Key Considerations

Individually Randomized Group Treatment Trial

- Participants are individually randomized
  - Education arm: self-administer intervention
  - Telehealth PT arm: participants nested within PTs
- Clustering can impact definitions of
  - Treatment conditions
  - Intercurrent events
  - And maybe others

Intercurrent events

- Participants
  - Discontinuation of protocol interventions, e.g. attending fewer than 8 PT sessions
  - Concomitant interventions, i.e. use of other treatments for LBP, e.g. acupuncture
- PTs
  - A PT can no longer participate in the trial. Participants receiving care from this PT may finish the remaining 8 PT sessions with another PT

Primary Estimand: Treatment Policy

Attribute	Definition
Population	Adults 18+ with chronic LBP living in rural communities with access to a health care system with primary care physicians and physical therapists
Treatment conditions	Completion of 8 telehealth PT sessions, delivered by the same PT, or completion of 10 self-administered on-line self-management modules
Endpoint	Difference in ODI comparing 12-week and baseline scores
Intercurrent-event strategy	Individuals are not restricted from seeking additional modalities of care for chronic LBP and may finish the telehealth PT sessions with a different PT if the original PT is no longer available
Population-level summary measure	Individual-average difference in 12-week change in ODI comparing telehealth PT to on-line education

Primary Estimand – In one sentence

The estimand is the individual-average difference in 12-week change in ODI scores comparing adults with chronic LBP living in rural communities with access to a health care system who are prescribed 8 telehealth PT sessions or a self-administered self-management on-line program, regardless of the degree to which they complete the program, whether they receive care from the same PT for all telehealth sessions and whether they seek additional modalities of care for their chronic LBP.

Secondary Estimands – Hypothetical Strategy Example

What if all participants received the intended intervention?

The estimand is the individual-average difference in 12-week change in ODI scores comparing adults with chronic LBP living in rural communities with access to a health care system who complete 8 telehealth PT sessions from the same PT or 10 self-administered self-management on-line modules, regardless of whether they seek additional modalities of care for their chronic LBP.

Question to the group: Hypothetical strategies are certainly interesting. But, how central/critical are they for pragmatic trials?

Crucial role of PI/study team/biostatistician interactions

- Key themes:
  - Teamwork makes the dreams work!
  - Estimand framework is not a “set it and forget it” framework

- Planning the trial: Elizabeth relied heavily on prior experience from Richard/Kevin

- Trial start up and implementation: Full team effort including regular meetings
  - Regular, scheduled communications between trial leadership, rural health care center partners, study coordinators, and biostats teams
  - Qualitative experience from the study coordinators is crucial to inform biostats team

Questions?

Please pose your questions in the Q&A

NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®