Unpacking Uncertainty, Meaningfulness, and Magnitude

Moderator:
Greg Simon, MD, MPH
Kaiser Permanente Washington
Health Research Institute

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Our Mission

- Review experiences of Collaboratory investigators in:
  - Choosing “meaningful” differences for pragmatic trials
  - Communicating findings to different audiences
- Discuss these issues from the perspective of different “customers” for pragmatic trial findings:
  - Patients and clinicians
  - Health system leaders / implementers
  - Regulators

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Let's Hear From…



Lynn DeBar
Kaiser Permanente Center for Health Research



Natalia Morone
Boston University & Boston Medical Center



Juan Marcos Gonzalez
Duke University



Devon Check
University of Michigan



Laura Lee Johnson
US Food & Drug Administration

How Meaningful is Clinically Meaningful Difference in Pragmatic Studies?

OPTIMUM Case Study

Natalia E. Morone, MD, MS
Associate Professor
Boston Medical Center/Boston University



Optimum

NIH NCCIH

OPTIMUM Case Study: Optimizing Pain Treatment in Medical Settings Using Mindfulness

JAMA Internal Medicine

RCT: Mindfulness-Based Group Medical Visits for Persons With Chronic Low Back Pain

POPULATION

133 Male, 318 Female



Adults with chronic low back pain

Mean age, 52.1 y

INTERVENTION

451 Individuals randomized

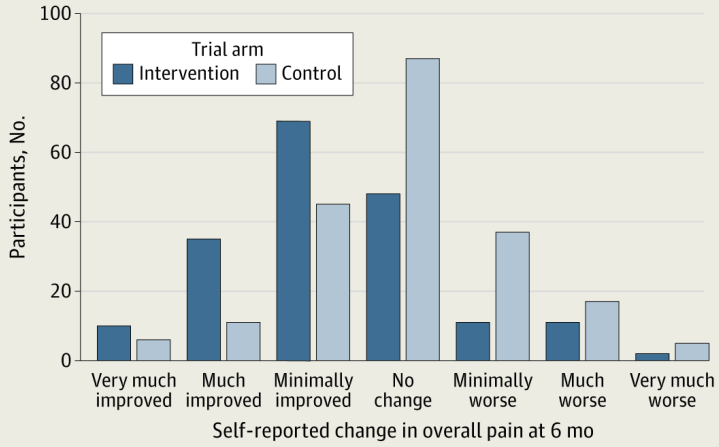


224 Mindfulness-based stress reduction (MBSR)
8-wk Telehealth MBSR led by mindfulness instructor focused on treating low back pain

227 Usual care
Routine treatment of chronic low back pain

FINDINGS

MBSR participants had significant improvement in PEG score at 6 mo compared with usual care



Mean change in PEG score at 6 mo:
MBSR: -1.21 (95% CI, -1.50 to 0.92)
Usual care: -0.59 (95% CI, -0.86 to -0.31)
Between-group difference: -0.62 (95% CI, -1.02 to -0.23); *P* <.002

SETTINGS / LOCATIONS

3 Primary care sites in the US

PRIMARY OUTCOME

Change in pain intensity and interference at 6 mo, measured using the Pain, Enjoyment of Life and General Activity (PEG) scale. Score range 0-10, with higher scores indicating worse pain.

Morone NE, Faurot KR, Weinberg J, et al. Mindfulness-based group medical visits for persons with chronic low back pain: a randomized clinical trial. *JAMA Intern Med*. Published online June 29, 2026. doi:10.1001/jamainternmed.2026.2186

© AMA

Designing for Clinical Meaning



Pragmatic design: broad entry criteria, participants followed for 12 months



Intervention translatable to clinical setting: Group Medical Visit with established E & M codes allows for billing by primary care clinicians



Telehealth design decreases burden on clinic for space and sometimes more convenient for patients

Assessing Meaningful Outcomes Over a Meaningful Timeframe

Measures have to be: Validated; Recommended by national organizations; Used in clinical settings; Meaningful to customers

Outcome Measures Over 12 Months	Customers
Pain, Enjoyment of Life, General Activity	Patients, clinicians, payors, health care systems
PROMIS-29	Some patients, clinicians, payors, health care systems
Global Impression of Change	Patients, clinicians, payors, health care systems
Qualitative data	Patients, clinicians, payors, health care systems

Clinically Meaningful Difference: PEG

1 point change is commonly accepted

- Range from 0.60 to 2.05
- 1-point change usually within and not between groups

Recommended by the CDC, CMS, VA, and adopted widely

OPTIMUM powered for a 1-point change ***between*** groups

*Pain, Enjoyment of Life, General Activity

Designing for the customer

Customers	Patients, clinicians, payors, healthcare system leaders
Key concepts, measures, and endpoints	Designed with implementation in mind • Widely recommend and widely used primary outcome with research and clinical utility (PEG) • Group medical visits are established in clinical care with workflows and billable codes • Patient perception of the utility of the intervention • Clinician perception of the utility of the intervention and ease of implementation • Safety

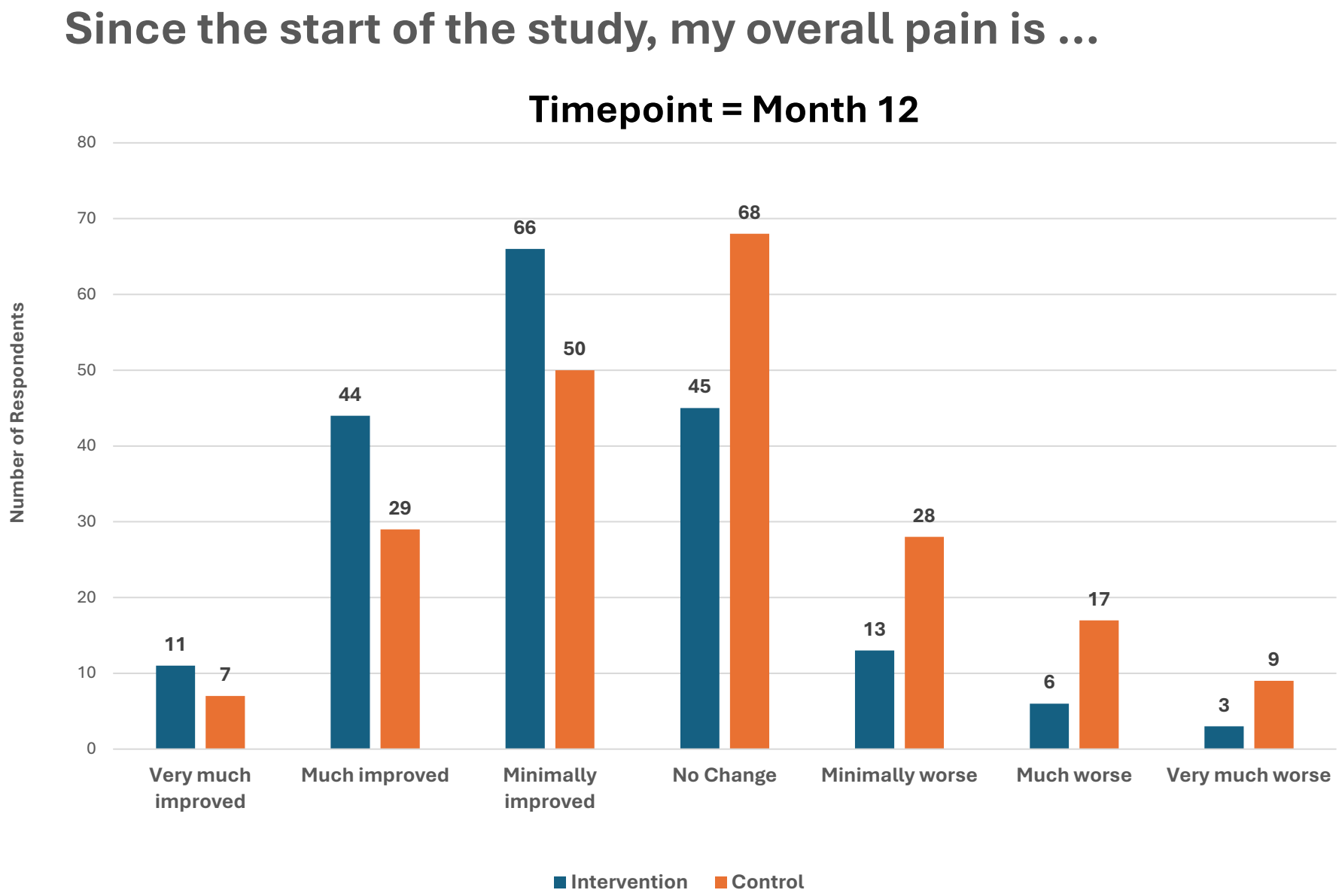
Main Outcomes

Group Comparisons for the Primary Outcome, PEG*. Intention To Treat Approach

	Model Estimated Means (95% CI*)		Mean Change From Baseline (95% CI)		Between Group Difference (95% CI)	P-value
PEG	**Intervention** N=224	**Control** N=227	**Intervention**	**Control**		
Baseline	6.37 (5.95, 6.79)	5.87 (5.45, 6.29)	--	--	--	
Week 8	5.21 (4.76, 5.65)	5.60 (5.16, 6.04)	-1.16 (-1.44, -0.88)	-0.27 (-0.53, -0.003)	-0.89 (-1.27, -0.51)	<0.001
Month 6	5.16 (4.69, 5.62)	5.28 (4.83, 5.74)	-1.21 (-1.50, -0.92)	-0.59 (-0.86, -0.31)	-0.62 (-1.02, -0.23)	0.002
Month 12	4.85 (4.39, 5.30)	5.09 (4.65, 5.54)	-1.52 (-1.81, -1.23)	-0.78 (-1.05, -0.50)	-0.74 (-1.14, -0.34)	<0.001

*PEG: Pain, Enjoyment of Life, General Activity; CI: Confidence Interval
The p-value for the group by time interaction in the PEG from baseline to Month 12 is <0.001
The model accounts for possible correlation within all site/cohort combinations. An unstructured covariance assumption was used for repeated measures over time with
comparisons between groups were performed using statistical contrasts. All means and confidence intervals are based on the model.

Results asking participants to rate their change



Safe with few related adverse events	
Increased pain from chair mindful movement and stretching	Resolved within a week
Anxiety during body scan	Resolved within a few minutes
Nausea during meditaion	Improved with time, continued in program
Sciatic pain during meditation	Rapid improvement

Adherence

Session Attendance	Mean (Std Dev)
All sites	4.3 (2.9)

Designing for the customer

Customers	Patients, clinicians, payors, healthcare system leaders
Key concepts, measures, and endpoints	Designed with implementation in mind • Widely recommend and widely used primary outcome with research and clinical utility (PEG) • **Group medical visits are established in clinical care with workflows and billable codes** • Patient perception of the utility of the intervention • **Clinician perception of the utility of the intervention and ease of implementation** • Safety

Considerations: PEG 1-point Threshold predefined vs Perceptions of Meaningfulness of the Intervention

PEG 1-point threshold difference within but not between groups achieved

The participants' subjective impression of change was highly relevant and reflected personal experience

The intervention was safe, with few adverse events

Despite low adherence, participants significantly improved up to 12 months

Considerations: PEG 1-point Threshold predefined vs Perceptions of Meaningfulness of the Intervention

A broader definition of meaningfulness would consider the unique perspectives of the “customers” beyond one measure’s clinical meaningfulness

Broader meaningfulness could include a variety of factors unique to the intervention such as patient perception of improvement, ease of incorporation into clinical care, availability of clinicians and healthcare staff to deliver the intervention

Patient reported outcome measures could consider a different threshold for clinically meaningful difference in a pragmatic setting in which there might be lower adherence and increased variability

Are Positive Main Outcomes and Cost Findings Enough?: Learnings from Three Recent Pragmatic Trials

Lynn DeBar, PhD, MPH
Kaiser Permanente Center for Health Research, Portland, OR, USA

Supported by NIH: UH2/UH3NS088731 (NINDS-PPACT), UG3/UH3AT010739 (HEAL / NCCIH - BackInAction), UG3/UH3AG067493 (HEAL / NIA – RESOLVE), UG3NS145203 (HEAL / NINDS – CARNATION)

PPACT Overview

DeBar et al, *Translational Behavioral Medicine*, 2012;
DeBar et al, *Contemporary Clinical Trials*, 2018;
Owen-Smith et al, *J Gen Internal Medicine*, 2018
Mayhew et al, *J of Pain*, 2019
Gruss et al, *Int J of Drug Policy*, 2019
Mayhew et al, *Contemp Clinical Trials Comm*, 2020
Gruss et al, *J Gen Internal Medicine*, 2020
DeBar et al, *Annals of Internal Medicine*, 2021;
Smith et al, *Medical Care*, 2022
Benes et al, *Pain Manag Nurs*, 2022

AIM: Integrate interdisciplinary services into primary care to help patients adopt self-management skills to:

- Manage chronic pain (decrease pain severity / improve functioning)
  - Limit use of opioid medication
  - Identify exacerbating factors amenable to treatment

Focus on feasibility and sustainability

DESIGN: Cluster (PCP)-randomized PCT *(106 clusters, 273 PCPs, 851 patients)*

ELIGIBILITY: Chronic pain, long-term opioid tx (prioritizing ≥ 120 MED, benzodiazepine co-use, high utilizers [≥ 12 visits in 3 months])

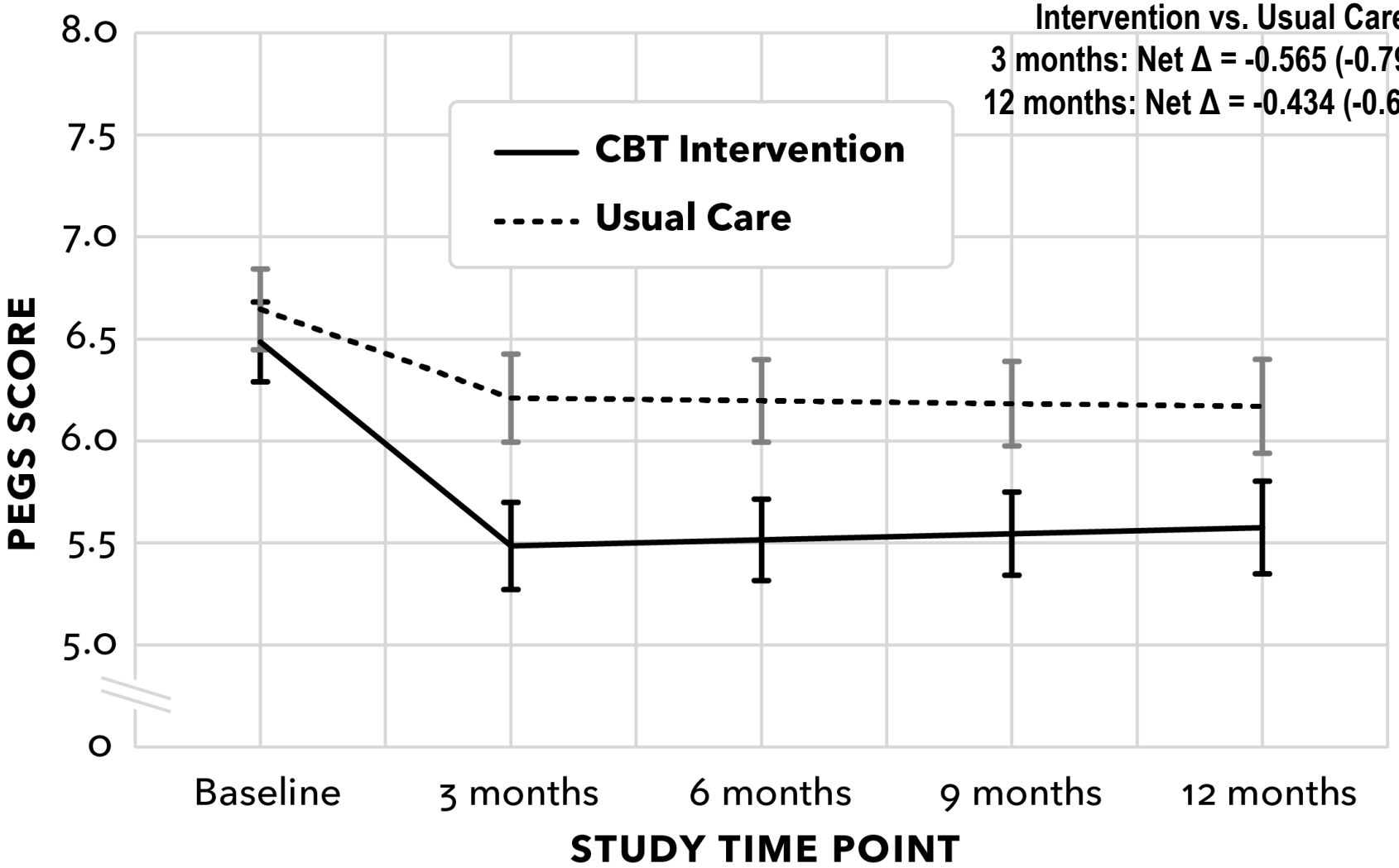
INTERVENTION: Behavioral specialist, nurse case manager, PT, and pharmacist team; 12-week core CBT + adapted movement groups

OUTCOMES: **Pain (3(4)-item PEG[S] [pain intensity, interference with enjoyment of life, general activity, and sleep]),** patient satisfaction, opioid dose, pain-related health services, cost, and PCP burden/burnout

CUSTOMER: Healthcare system leaders/payers initiated on behalf of PCPs

Primary Outcome (PEGS) by Treatment Group

Intervention vs. Usual Care from Baseline to…
3 months: Net Δ = -0.565 (-0.796 to -0.333); ES = -0.28
12 months: Net Δ = -0.434 (-0.690 to -0.178); ES = -0.21

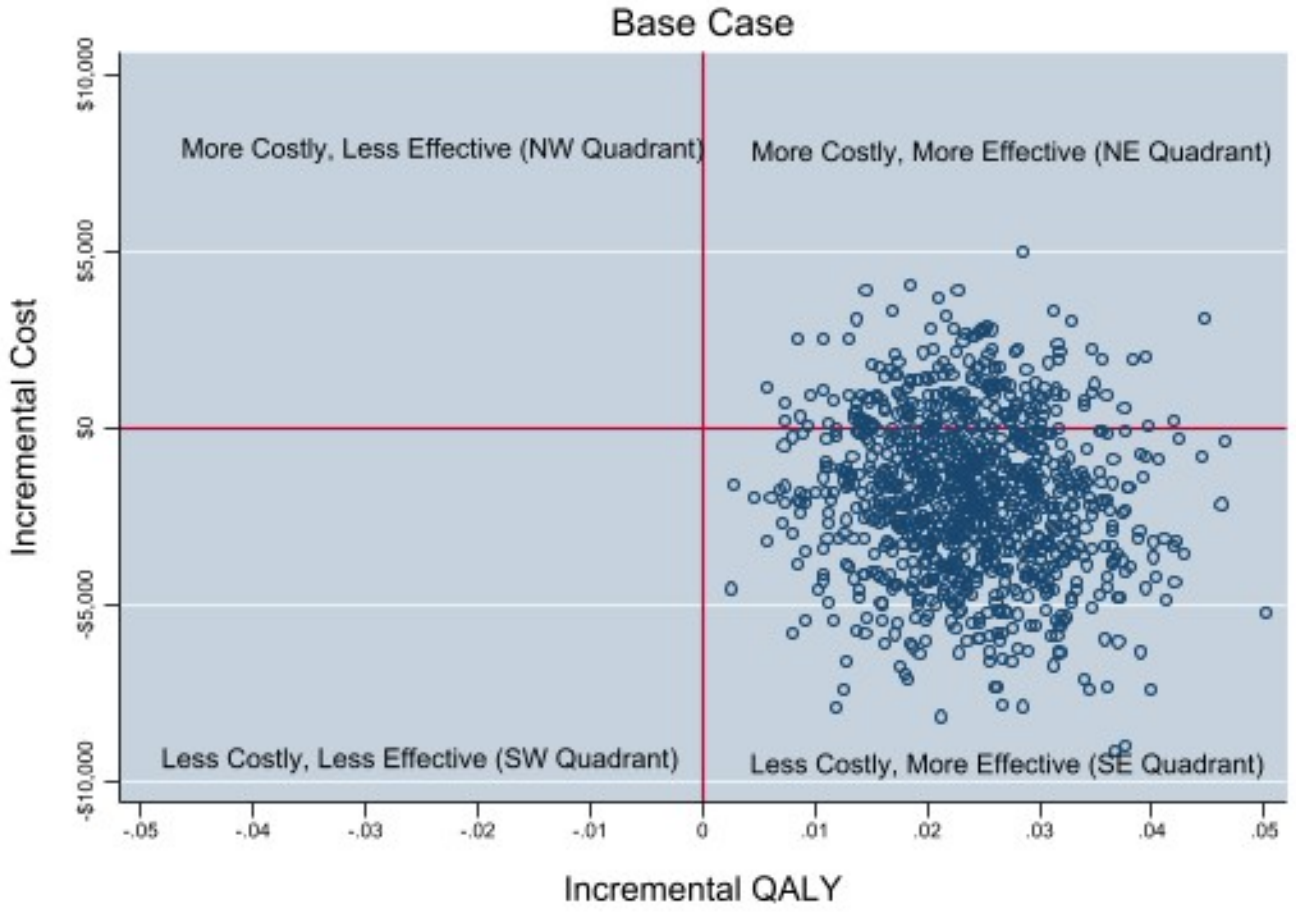


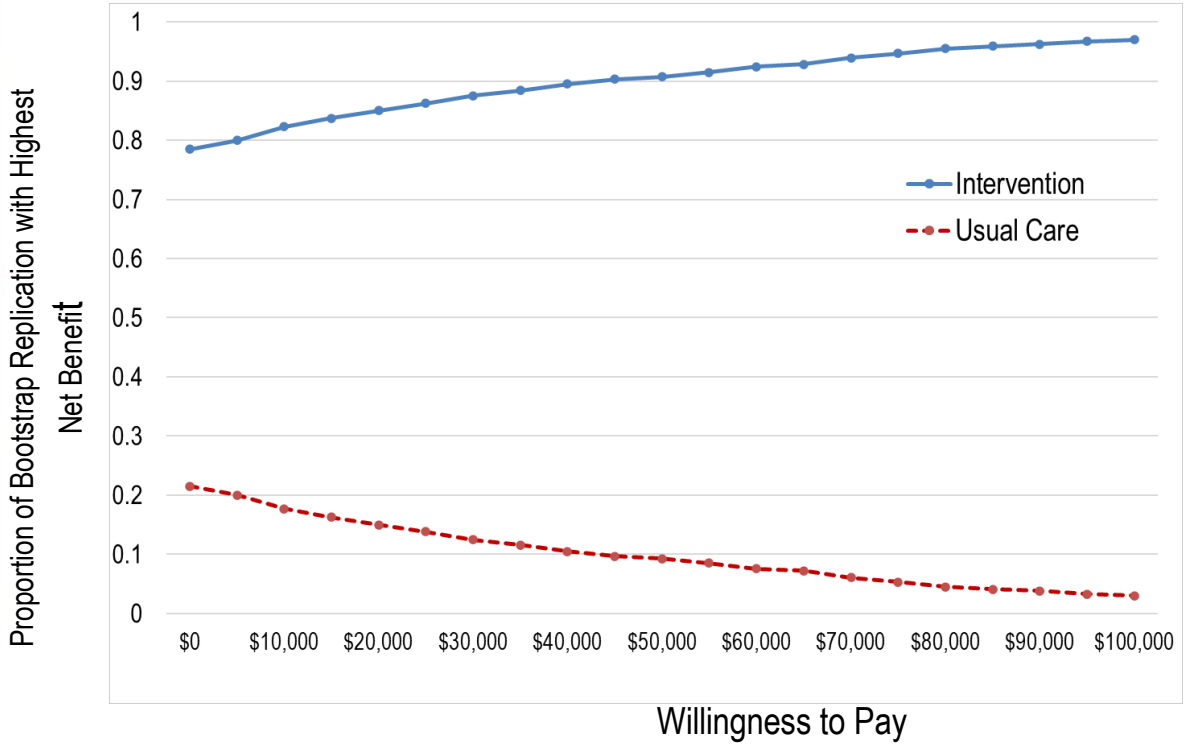
PEGS (4 items) –
- **P**ain intensity

Interference with
- **E**njoyment of life
- **G**eneral Activity
- **S**leep

DeBar L, et al. *Ann Intern Med.* 2021.

Incremental Cost-Effectiveness Plane*and Cost-Effectiveness Acceptability Curve





* Cost per QALY Gained Based on Intervention Replication Costs ($2,145 and Total Medical Care Costs)

Smith et al, *Medical Care*, 2022

RESOLVE

Tailored Non-Pharmacotherapy Services for Chronic Pain: Testing Scalable and Pragmatic Approaches

Funded as part of NIH HEAL Pain Effectiveness Research Network Sub-initiative: UG3/UH3AG067493

AIM: To evaluate the effectiveness of two cognitive-behavioral therapy (CBT)-based chronic pain treatments delivered via different telehealth modalities (online and live virtual coach-led) to:

- Achieve clinically meaningful improvements in pain severity (pain intensity + pain related interference) and other secondary pain-related and quality of life outcomes (post-treatment and at 12-month follow-up)
- Determine which telehealth intervention best for whom (subgroup outcome and cost analyses)
- Evaluate cost-utility and cost effectiveness of the telehealth interventions relative to one another and usual care
- Qualitative analyses to understand patient, clinician, and healthcare system experience

Mayhew, et al. Trials, 2023.
DeBar et al, JAMA, 2025
Cook A et al, PAIN, 2026
Dickerson et al, *Value in Health*, 2026
Von Korff et al, *J of Pain*, 2026
Balderson et al, *Pain Medicine*, 2025
Schneider et al, *Permanente Journal*, 2026
Cook S et al, J Clin Transl Sci, 2026

DESIGN: 3-arm parallel group (online CBT/coach-led CBT/ enhanced usual care) pragmatic comparative effectiveness trial *(2,333 participants)*

ELIGIBILITY: ≥ 18 years of age with EHR musculoskeletal pain-related diagnosis; meet threshold for high impact chronic pain (≥ 3 months of pain that limit life or work activities [most or every day])

CUSTOMER: Healthcare system leaders/payers, clinicians, and patients (co-equal)

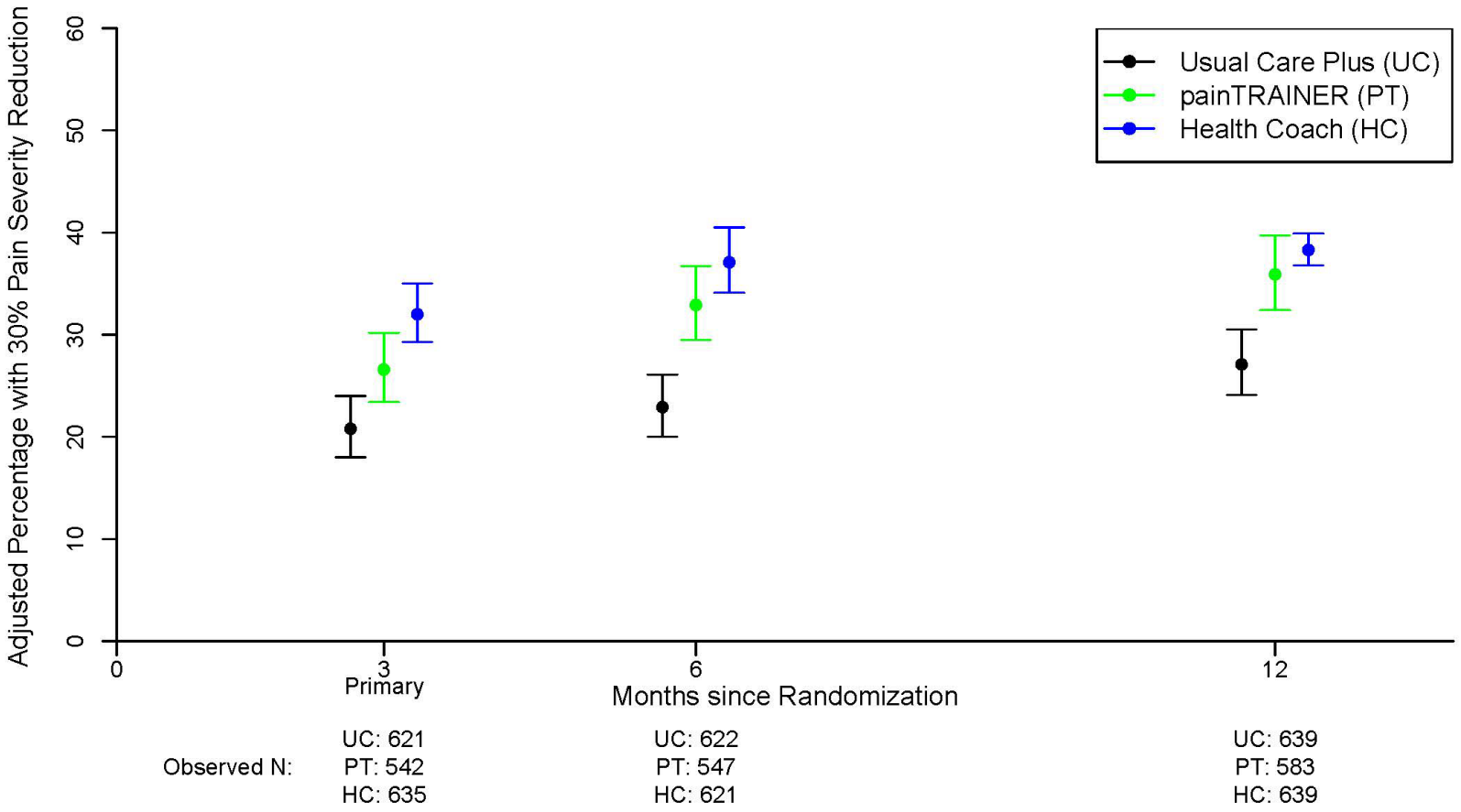
SETTINGS:

- Kaiser Permanente Northwest
- Kaiser Permanente Washington
- Kaiser Permanente Georgia
- Essentia Health (Upper Midwest)

INTERVENTIONS:

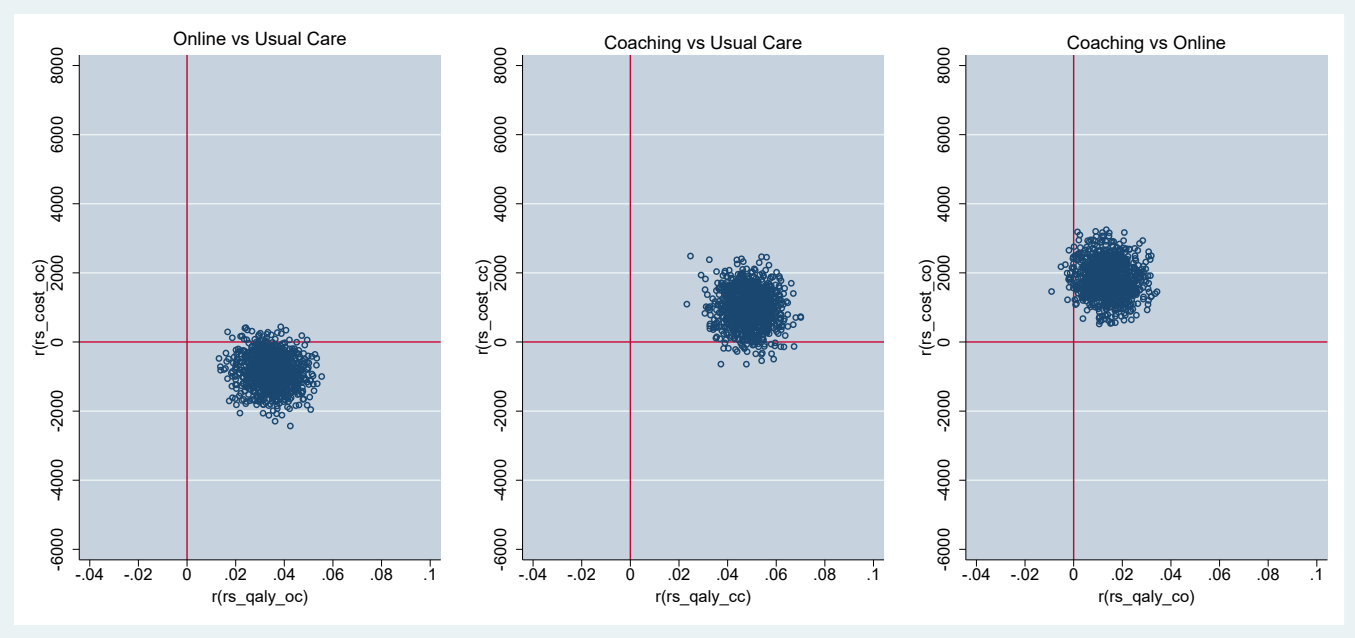
- **Arm 1:** Online CBT for chronic pain (PainTRAINER™) with eight approximately weekly sessions over 12 weeks (tech support & monitoring included
- **Arm 2:** Virtual, live coach-led CBT for chronic pain (telephone or video) with eight approximately weekly sessions over 12 weeks

Main Outcome: Adjusted Percentage with 30% or Greater Reduction in Pain Severity (Minimal Clinically Important Difference)

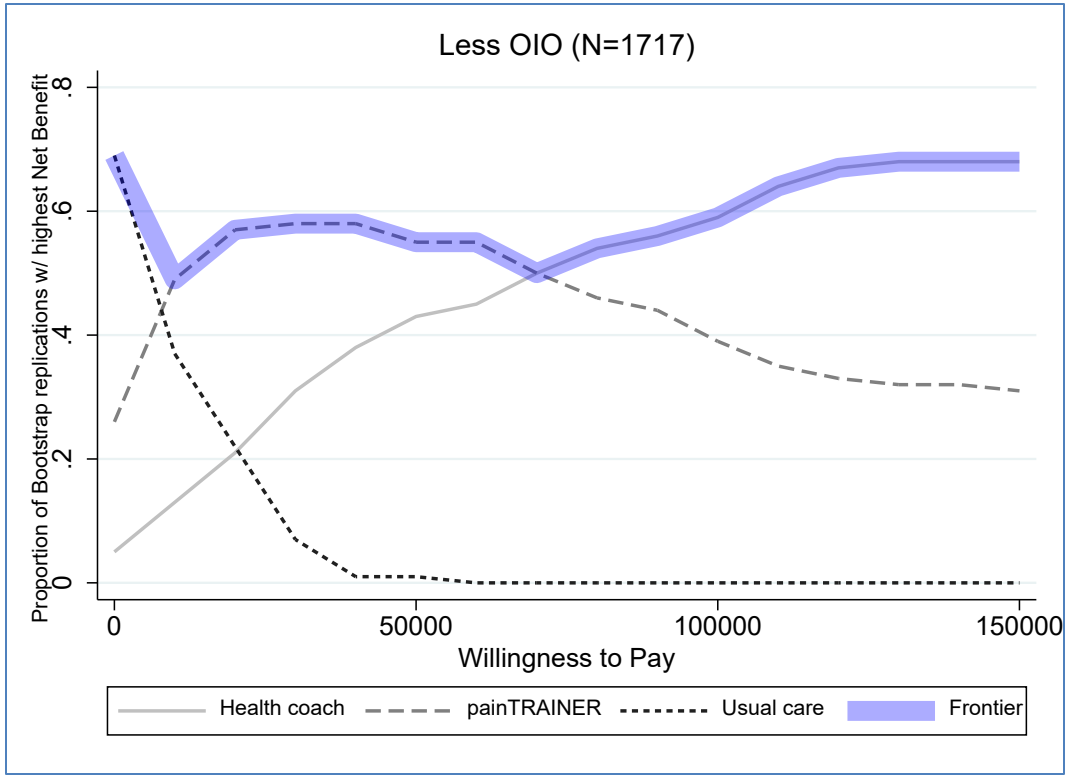


DeBar, Mayhew, Wellman et al. *JAMA*, 2025

Cost Effectiveness Analysis: Pain Related Costs and Quality Adjusted Life Years (QALYs)



Dickerson, O'Keefe-Rosetti, Mayhew et al, *Value in Health*, 2026



Pragmatic Trial of Acupuncture for Chronic Low Back Pain in Older Adults

Funded as part of NIH HEAL PRISM (Collaboratory) Sub-initiative: UG3/UH3AT010739

AIM: To test the effectiveness of acupuncture needling among older adults with chronic low back pain to:

- Improve back pain-related disability
- Evaluate acupuncture needling dose-dependence and safety for older adults
- Inform CMS Medicare Coverage Decision/Impact (outcomes + acupuncturist / key informant interviews)
- Evaluate cost-utility and cost effectiveness of acupuncture intervention

semi-flexible acupuncture protocol and community acupuncturist care provision focus

Nielsen A, et al. *Glob Adv Health Med*. 2021
DeBar et al, *Contemporary Clinical Trials*. 2023.
Teets et al, *Global Adv Health Med*, 2025.
DeBar et al, *JAMA Network Open*, 2025
Justice et al, *Contemporary Clinical Trials*, 2025
Herman et al, *Spine*, 2026
Nielsen et al, *Integrative Medicine Research*, 2026

DESIGN: 3-arm (standard acupuncture / enhanced acupuncture / usual medical care only) pragmatic trial *(800 participants)*

SETTINGS:

- Kaiser Permanente Washington
- Kaiser Permanente Northern California
- Sutter Health Northern California
- Institute of Family Health NYC

ELIGIBILITY: ≥ 65 years of age with EHR diagnosis of uncomplicated chronic low back pain meeting threshold of pain-related general activity interference (≥ 3 on PE**G**)

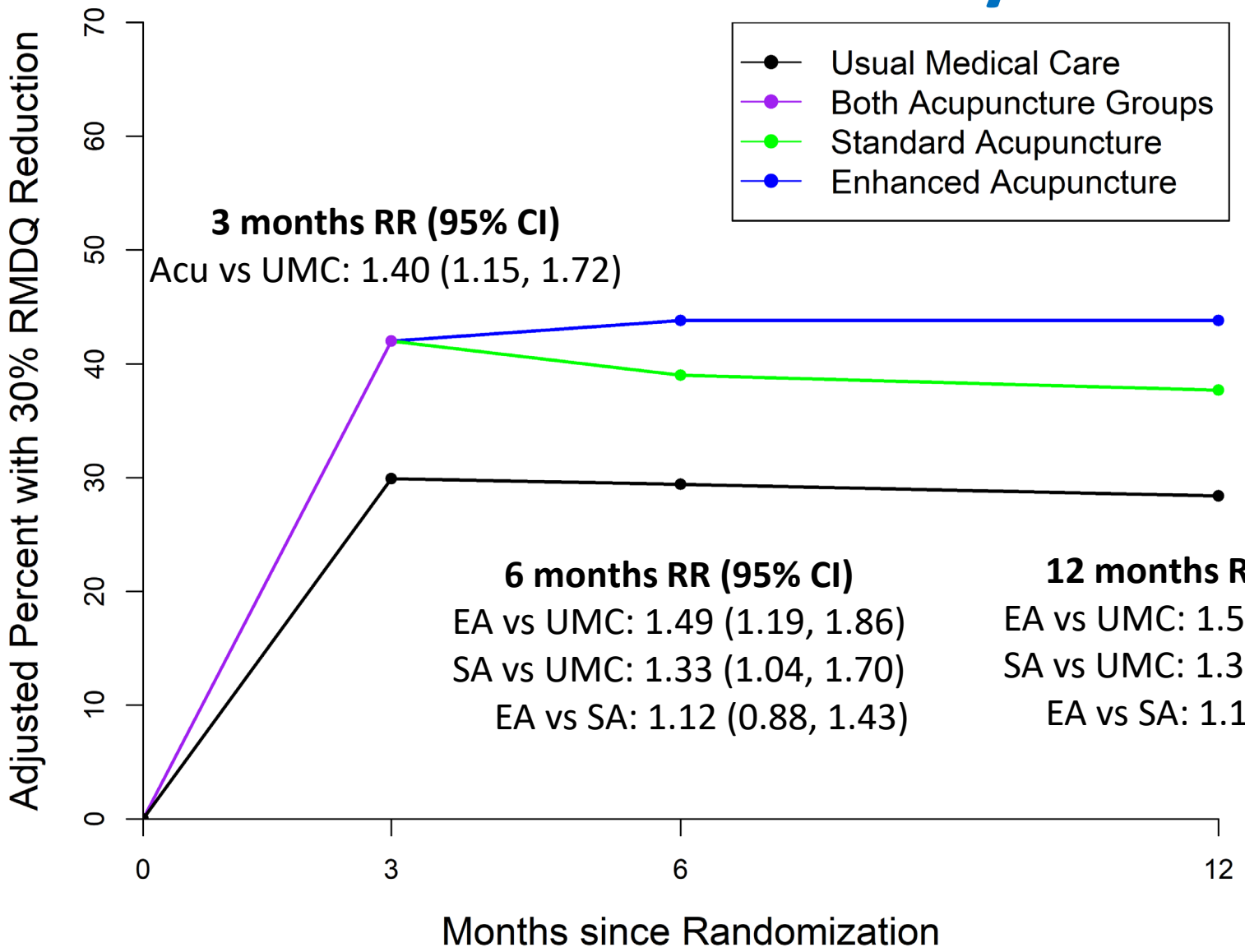
INTERVENTIONS:

- **Arm 1:** Standard (15 treatment sessions over 12 weeks)
- **Arm 2:** Enhanced (6 additional maintenance sessions over following 12 weeks)
- **Arm 3:** Usual medical care only

50+ community- (KPNC, KPWA, SH) or primary clinic-based (IFH) acupuncturists delivering intervention

CUSTOMER: Payers/regulators > patients > acupuncturists (& their professional organizations)

Minimal Clinically Important Differences in Functional Disability



3 months RR (95% CI)
Acu vs UMC: 1.40 (1.15, 1.72)

6 months RR (95% CI)
EA vs UMC: 1.49 (1.19, 1.86)
SA vs UMC: 1.33 (1.04, 1.70)
EA vs SA: 1.12 (0.88, 1.43)

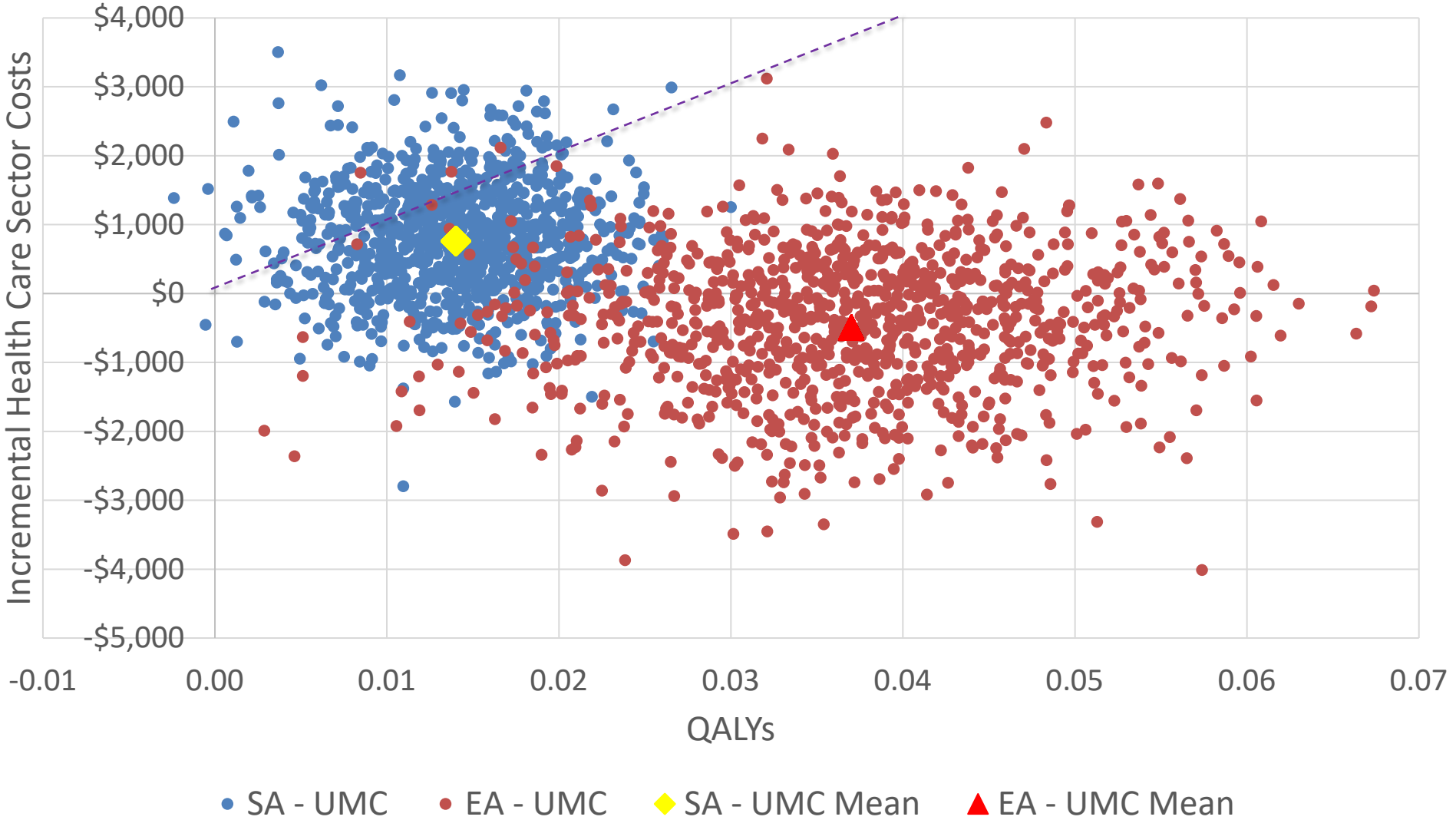
12 months RR (95% CI)
EA vs UMC: 1.54 (1.25, 1.91)
SA vs UMC: 1.33 (1.06, 1.66)
EA vs SA: 1.16 (0.98, 1.37)

DeBar et al, *JAMA Network Open*, 2025

All statistically significant (P Value<0.05) except for the comparison of EA vs SA at 6 and 12 months

RR- Relative Risk
CI – Confidence Interval

Health Care Sector Cost / QALY Cost-Effectiveness Plane



Herman et al, *Spine*, 2026

	PPACT	RESOLVE	BackInAction
Customer-related learnings	• Healthcare system leadership dynamic w/shifting priorities that may not stay in alignment with study priorities • Cost effectiveness analysis findings less impactful than hoped despite cost offset – "savings" in one place can be "costs" elsewhere in the system	Qualitative data can enrich understanding substantively: • Healthcare system leaders / clinicians feedback balanced respective costs & benefits for online versus virtual health coach led programs • Patient reflections provided insights into the pattern of pain outcomes for each active intervention over time	Attenuating factors re outcome significance: • (In)Ability for older adults to access Medicare reimbursed acupuncture services • Maintenance acupuncture cadence likely misaligned with real world care practices • Few 2^0 outcome improvements counter to acupuncturist expectations for holistic care
How big of effect / outcome relevant to customer(s)?	Magnitude of effect relevant?: • Stories / specific PCP-patient experiences mattered a lot • Lack of tx-related opioid morphine equivalent dose reduction may have discouraged more robust program uptake in 2010's era	Most important: • Effect size relative to other pain-related txs(medications and other nonpharmacotherapy txs) • Sustained improvements after end of active treatment • Aligned 1^0 and 2^0 outcomes	Most important: • Effect size relative to other pain-related interventions (meds / other NPT) • Sustained improvements after end of active treatment • Impact on both pain severity and pain intensity in short run

Tensions and Trade Offs Takeaways

- MCID important but "magnitude" relative to other treatments and also sustained impact past treatment window critical
- Qualitative findings (Fact based stories) can be as (more?) compelling than strong quantitative data – how do we use such findings more robustly?
- Cost effectiveness findings often less impactful than we'd hoped (be aware of complex systems and contextual factors that can impede uptake)
- Alignment of outcomes may be as/more important than single primary outcome

When Is Effective Meaningful?

Interpreting magnitude through evidence, context, uncertainty, and patient preferences

Juan Marcos Gonzalez, PhD

NIH Pragmatic Trials Collaboratory

Clinical evidence does not interpret itself. It becomes meaningful only after passing judgments about what matters, for whom, and under what conditions.

From perceptibility to meaning to value

1

PERCEPTIBILITY

Can the clinical change be noticed or detected?

Bounded by the measuring instrument and its ability to detect change in the context of use

2

MEANING

Does the change make a difference?

The difference may have personal or clinical significance, but significance alone does not establish worth

3

VALUE

Is the change worth pursuing?

The benefit is judged relative to risks, burden, cost, and enabled activities

Shifts the question from “can we do it?” to “should we do it?”

All evidence passes through judgment

The meaning of an effect emerges from a series of value-laden interpretations.

Statistical significance	Is the magnitude of an effect relative to uncertainty sufficient?
Context	Are the comparators, outcomes, timeframes, and settings sufficient?
Expectations	Are average effects sufficient?
Tolerance	Are the tradeoffs implied by the evidence worthwhile?

The same evidence can support different judgments

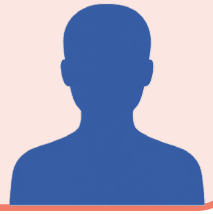
Each perspective asks a different question and uses a different value function.

System administrators — Can the change be achieved efficiently, in a scalable and equitable manner?

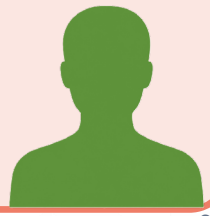
Clinicians — Does it change health or management?

Patient — What does the change enable, and is it worth what achieving it requires?

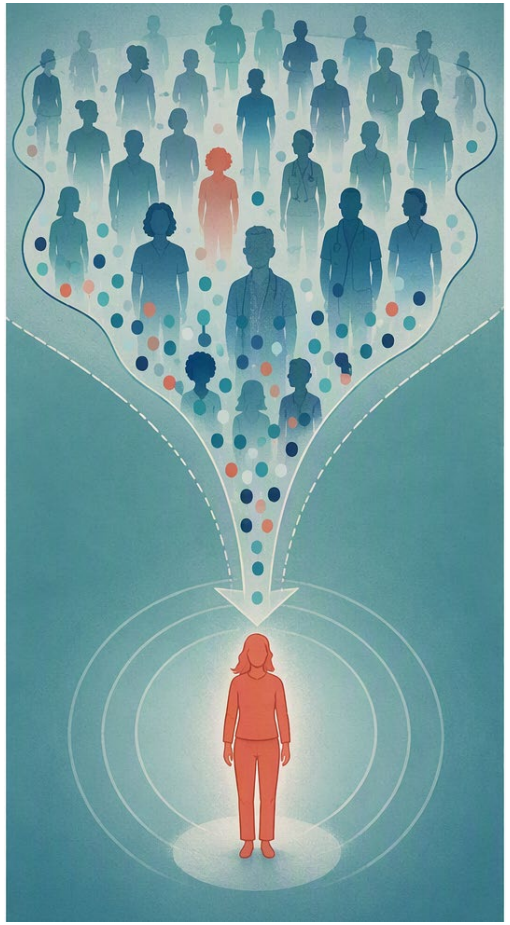
Patient 1 — Enjoys hiking and sports. May accept greater risks for a benefit that improves physical activity



Patient 2 — Has a sedentary life. May accept less risk for the same clinical benefit related to improvements in physical activity



The clinical problem of N = 1



Evidence question

What happened on average, and how uncertain is that estimate?

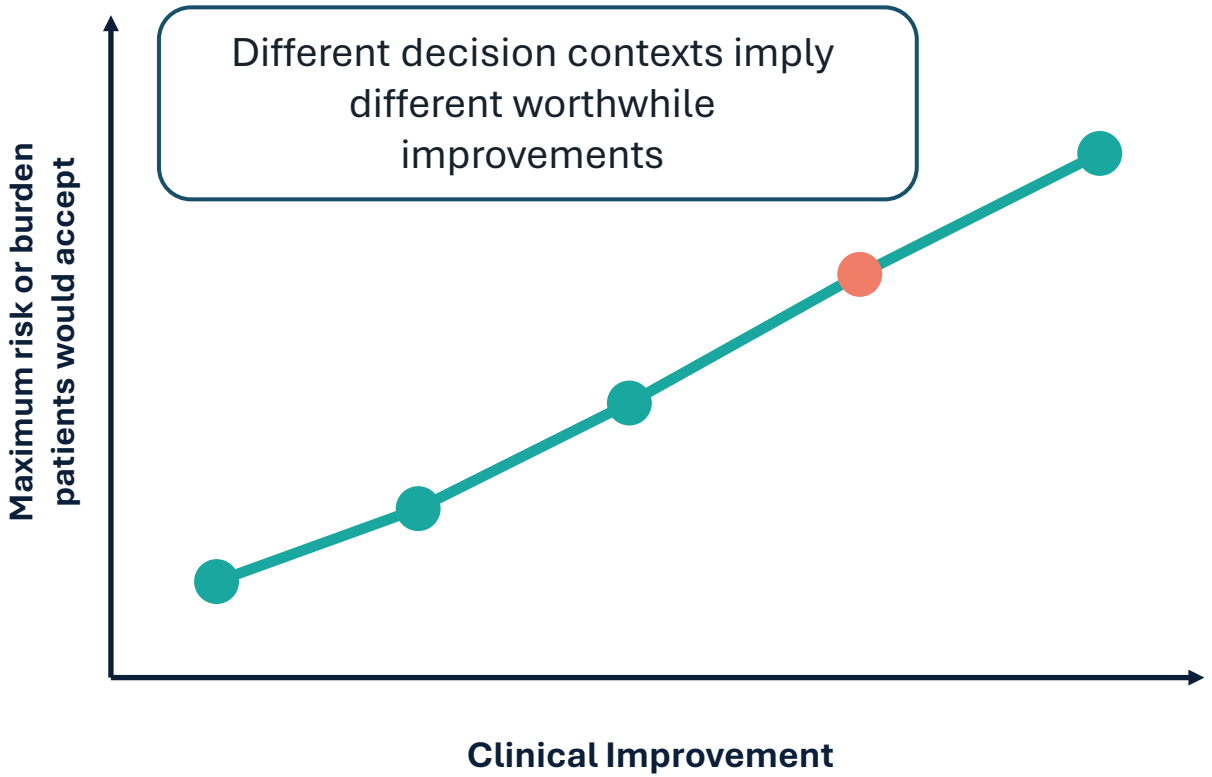
Applicability question

How much should this patient expect the evidence to describe them?

Decision question

Given the patient goals and risk tolerance, is the tradeoff acceptable?

Value sets context-specific thresholds



JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2025, VOL. 35, NO. 5, 981–1000
https://doi.org/10.1080/10543406.2023.2170400

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Use of Bayesian decision analysis to maximize value in patient-centered randomized clinical trials in Parkinson's disease

Shomesh E. Chaudhuri[a], Zied Ben Chaouch[a,b], Brett Hauber[c,d], Brennan Mange[c], Mo Zhou[e], Stephanie Christopher[f], Dawn Bardot[f], Margaret Sheehan[g], Anne Donnelly[g], Lauren McLaughlin[h], Brittany Caldwell[e], Heather L. Benz[e], Martin Ho[e], Anindita Saha[e], Katrina Gwinn[e], Murray Sheldon[e], and Andrew W. Lo[a,i,j]

[a]Laboratory for Financial Engineering, MIT Sloan School of Management, Cambridge, MA, USA; [b]Electrical Engineering and Computer Science Department, MIT, Cambridge, MA, USA; [c]RTI Health Solutions, Research Triangle Park, NC, USA; [d]CHOICE Institute, University of Washington School of Pharmacy, Seattle, WA, USA; [e]Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA; [f]Medical Device Innovation Consortium, Arlington, VA, USA; [g]Patient Council, The Michael J. Fox Foundation for Parkinson's Research, New York, NY, USA; [h]Strategy and Planning, The Michael J. Fox Foundation for Parkinson's Research, New York, NY, USA; [i]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA; [j]Santa Fe Institute, Santa Fe, NM, USA

ABSTRACT

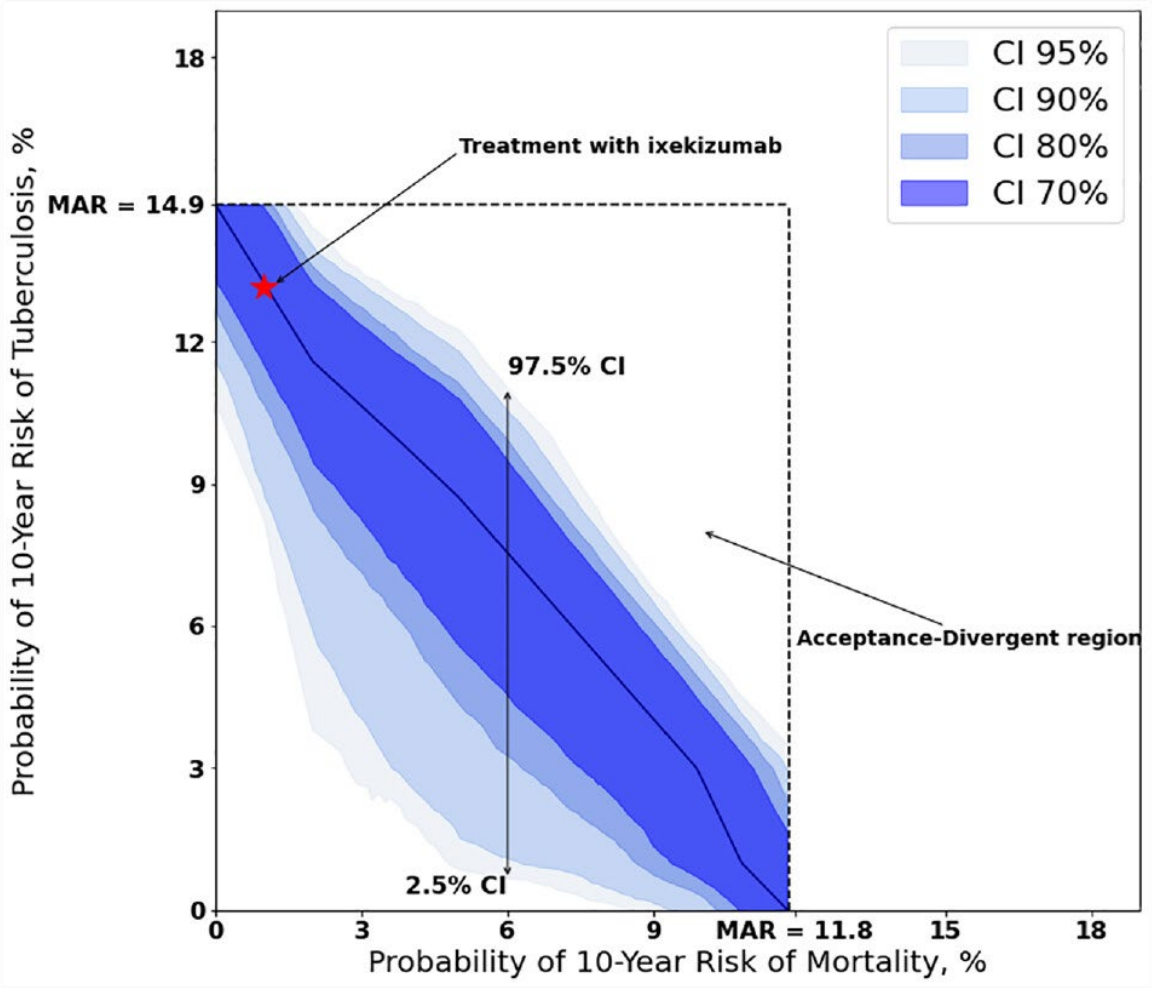
A fixed one-sided significance level of 5% is commonly used to interpret the statistical significance of randomized clinical trial (RCT) outcomes. While it is necessary to reduce the false positive rate, the threshold used could be chosen quantitatively and transparently to specifically reflect patient preferences regarding benefit–risk tradeoffs as well as other considerations. How can patient preferences be explicitly incorporated into RCTs in Parkinson's disease (PD), and what is the impact on statistical thresholds for device approval? In this analysis, we apply Bayesian decision analysis (BDA) to PD patient preference scores elicited from survey data. BDA allows us to choose a sample size (n) and significance level (α) that maximizes the overall expected value to patients of a balanced two-arm fixed-sample RCT, where the expected value is computed under both null and alternative hypotheses. For PD patients who had previously received deep brain stimulation (DBS) treatment, the BDA-optimal significance levels fell between 4.0% and 10.0%, similar to or greater than the traditional value of 5%. Conversely, for patients who had never received DBS, the optimal significance level ranged from 0.2% to 4.4%. In both of these populations, the optimal significance level increased with th

ARTICLE HISTORY
Received 27 May 2022
Accepted 15 January 2023

KEYWORDS
Bayesian decision analysis; patient preferences; Parkinson's disease; medical devices; benefit–risk

Extending this idea to uncertainty can imply accepting different significance thresholds

Tradeoffs: A tool to establish thresholds



Maximum acceptable risks can be derived from discrete-choice experiments

By observing the rates at which specific tradeoffs are accepted we can quantify the exchange rate between outcomes/risks

This information can help quantify the degree to which tradeoffs are acceptable for a specific group

Montano-Campos, J. F., Gonzalez, J. M., Rickert, T., Fairchild, A. O., Levitan, B., & Reed, S. D. (2023). Use of patient preferences data regarding multiple risks to inform regulatory decisions. *MDM Policy & Practice*, *8*(1), 23814683221148715.

Uncertainty can motivate behavior

When evidence does not fully specify what actions will produce, patients and clinicians can fill the gaps with personal expectations.

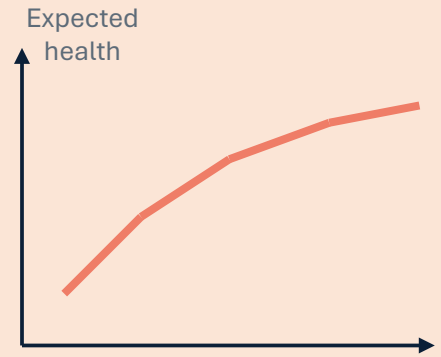
1 INCOMPLETE EVIDENCE

Evidence rarely defines the consequences of every possible action for one patient.

- Benefit probability
- Adverse-effect probability
- Partial adherence
- Patient-specific response

2 MENTAL MODEL

"If I take this action, what health will it produce, and at what cost?"



Evidence + experience + information from others

3 BEHAVIOR

Patient behavior

Initiate or decline

Adhere fully or strategically

Persist, switch, or discontinue

Clinician behavior

The same logic can frame recommendations, monitoring, and treatment adjustment under incomplete patient-specific evidence.

Ambiguity is not just wider uncertainty. It can change the beliefs through which evidence becomes behavior.

Muiruri, C., van den Broek-Altenburg, E. M., Bosworth, H. B., Cené, C. W., & Gonzalez, J. M. (2023). A Quantitative Framework for Medication Non-Adherence: Integrating Patient Treatment Expectations and Preferences. *Patient Preference and Adherence*, *17*, 3135–3145. https://doi.org/10.2147/PPA.S434640

Evidence is only part of the decision

- **Evidence** provides the context and empirical basis under which decisions are made.
- **Judgments** about relevance, uncertainty and value give the evidence meaning.
- In clinic, the decision is whether the effect is worthwhile for this patient, in this context, under this uncertainty.

What Makes an Intervention Implementable? Evidence for Health System Decision Makers

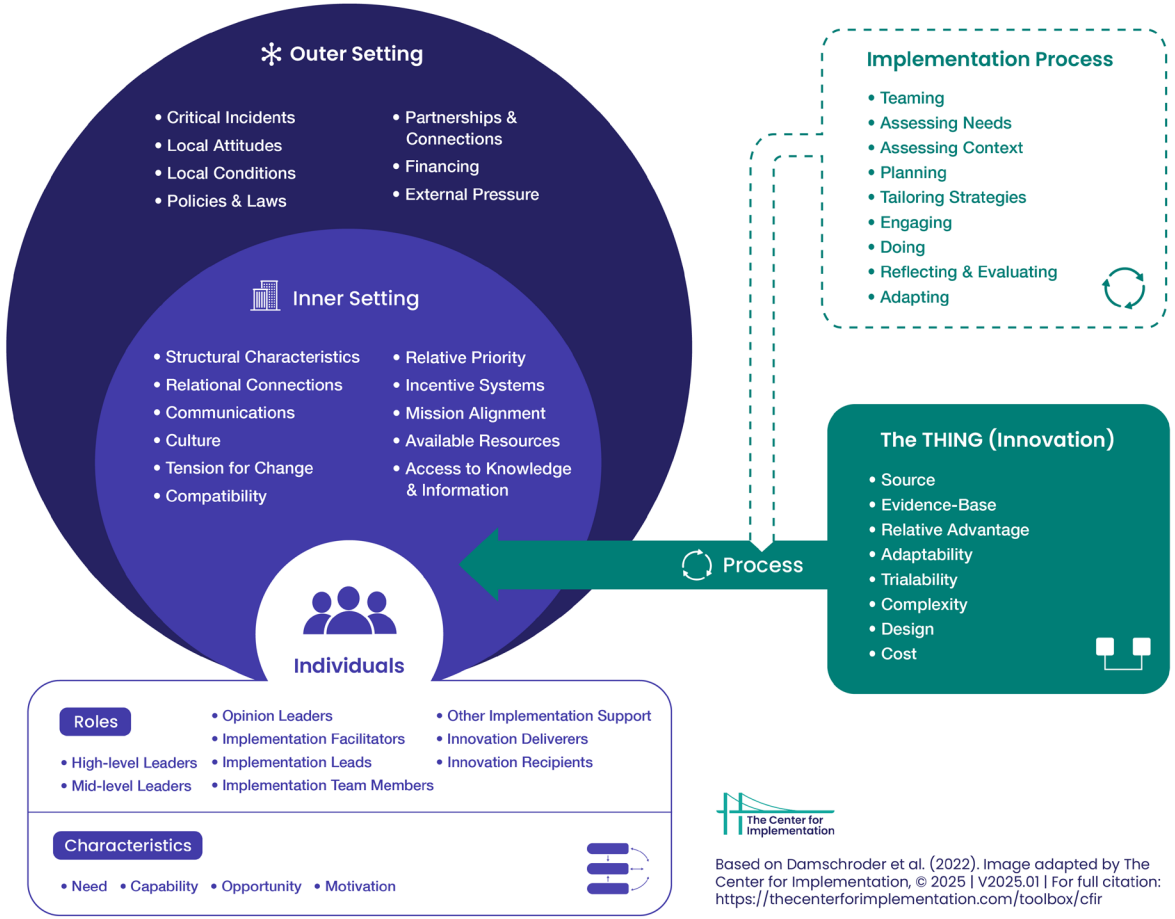
Devon Check, PhD

Co-lead, Implementation Science Core Working Group

NIH Pragmatic Trials Collaboratory

Implementation Determinants

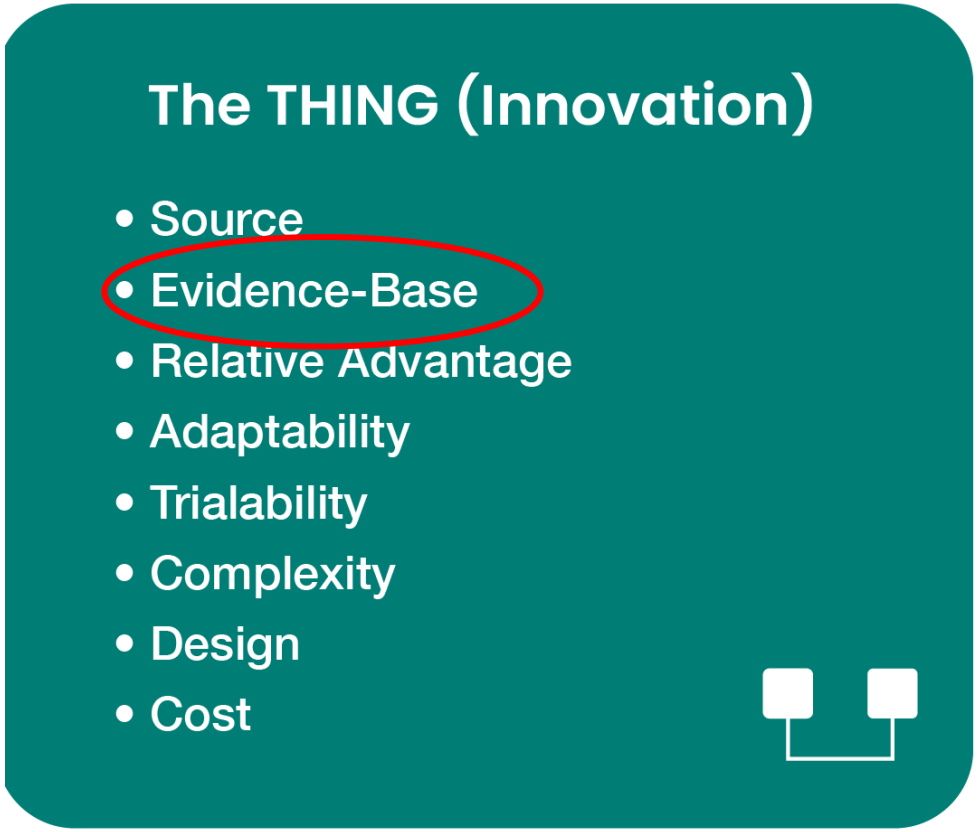
Consolidated Framework for Implementation Research (CFIR)



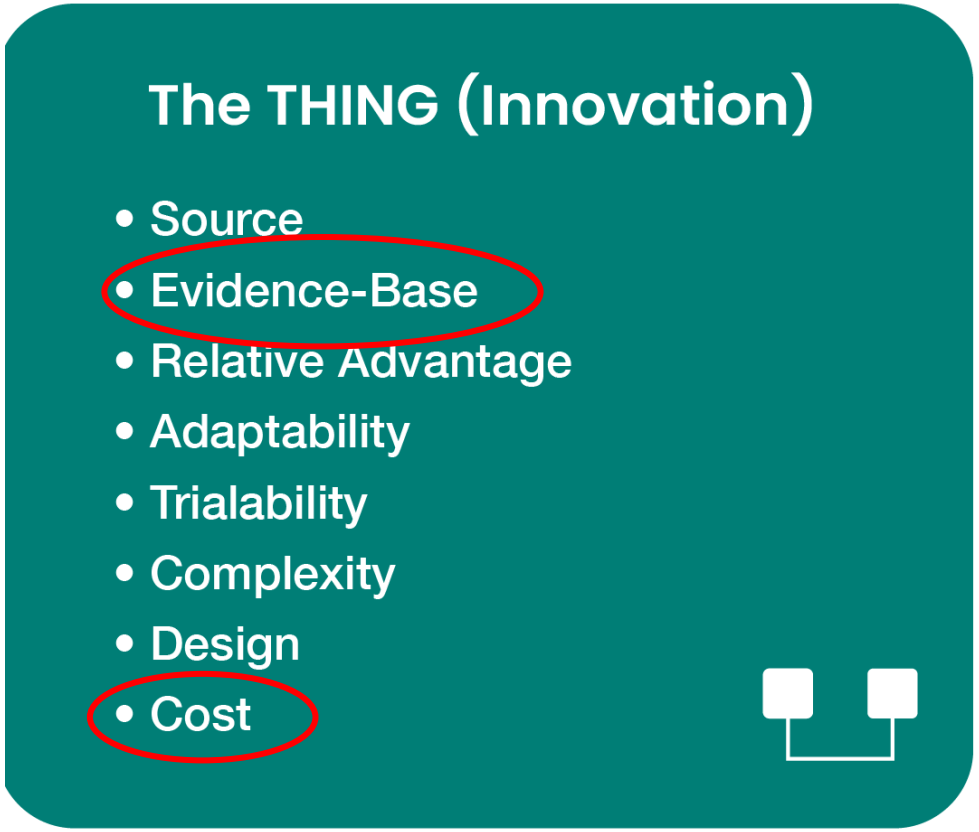
The Center for Implementation

Based on Damschroder et al. (2022). Image adapted by The Center for Implementation, © 2025 | V2025.01 | For full citation: https://thecenterforimplementation.com/toolbox/cfir

Implementation Determinants



Implementation Determinants

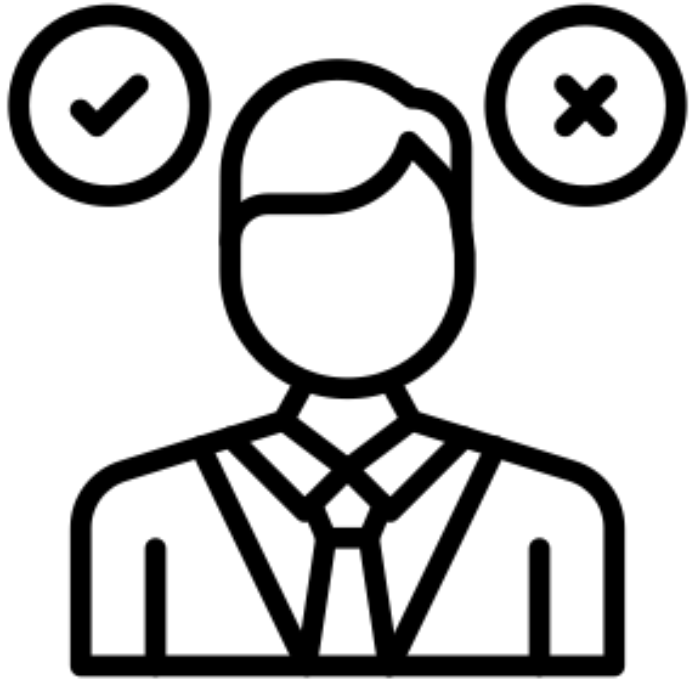


Economic Evaluations in ePCTs

Evaluate 2+ alternatives in terms of their relative costs and outcomes

Cost-effectiveness analysis: Measures program costs against natural health units, **e.g., cost per case prevented**

Cost-utility analysis: Measures program costs against a generic measure of health status that considers both morbidity and mortality, **e.g., cost per quality-adjusted life year (QALY) gained**



PACT
Pain
Program for Active
Coping & Training

ORIGINAL ARTICLE

Economic Evaluation

A Randomized Pragmatic Trial of a Primary Care-based Cognitive Behavioral Intervention for Adults Receiving Long-term Opioids for Chronic Pain

David H. Smith, PhD, RPh, Maureen O'Keeffe-Rosetti, MS,* Michael C. Leo, PhD,* Meghan Mayhew, MPH,* Lindsay Benes, PhD, RN, CNS,*† Allison Bonifay, MA, LPC,* Richard A. Deyo, MD, MPH,‡ Charles R. Elder, MD, MPH,* Francis J. Keefe, PhD,§ Carmit McMullen, PhD,* Ashli Owen-Smith, PhD, SM,‖¶ Connie M. Trinacty, PhD,# William M. Vollmer, PhD,* and Lynn DeBar, PhD, MPH***

Background: Chronic pain is prevalent and costly; cost-effective nonpharmacological approaches that reduce pain and improve patient functioning are needed.

Objective: Report the incremental cost-effectiveness ratio (ICER), compared with usual care, of cognitive behavioral therapy aimed at improving functioning and pain among patients with chronic pain on long-term opioid treatment.

Design: Economic evaluation conducted alongside a pragmatic cluster randomized trial.

Subjects: Adults with chronic pain on long-term opioid treatment (N = 814).

Intervention: A cognitive behavioral therapy intervention teaching pain self-management skills in 12 weekly, 90-minute groups delivered by an interdisciplinary team (behaviorists, nurses) with additional support from physical therapists, and pharmacists.

Outcome Measures: Cost per quality adjusted life year (QALY) gained, and cost per additional responder (≥30% improvement on standard scale assessment of Pain, Enjoyment, General Activity, and Sleep). Costs were estimated as-delivered, and replication.

Results: Per patient intervention replication costs were $2145 ($2574 as-delivered). Those costs were completely offset by lower medical care costs; inclusive of the intervention, total medical care over follow-up was $1841 lower for intervention patients. Intervention group patients also had greater QALY and responder gains than did controls. Supplemental analyses using pain-related medical care costs revealed ICERs of $35,000, and $53,000 per QALY (for replication, and as-delivered intervention costs, respectively); the ICER when excluding patients with outlier follow-up costs was $106,000.

Limitations: Limited to 1-year follow-up; identification of pain-related utilization potentially incomplete.

Conclusion: The intervention was the optimal choice at commonly accepted levels of willingness-to-pay for QALY gains; this finding was robust to sensitivity analyses.

Key Words: pain, opioid, economic evaluation, cost-effectiveness, pragmatic trial

(*Med Care* 2022;60: 423–431)

"Per-patient intervention replication costs…were completely offset by lower medical care costs."

Factors Affecting Post-trial Sustainment or De-implementation of Study Interventions: A Narrative Review

Terren Green, BA[1], Hayden B. Bosworth, PhD[1,2,3,4], Gloria D. Coronado, PhD[5], Lynn DeBar, PhD, MPH[5], Beverly B. Green, MD, MPH[6], Susan S. Huang, MD, MPH[7], Jeffrey G. Jarvik, MD, MPH[8], Vincent Mor, PhD[9], Douglas Zatzick, MD[10], Kevin P. Weinfurt, PhD[1,2], and Devon K. Check, PhD[2]

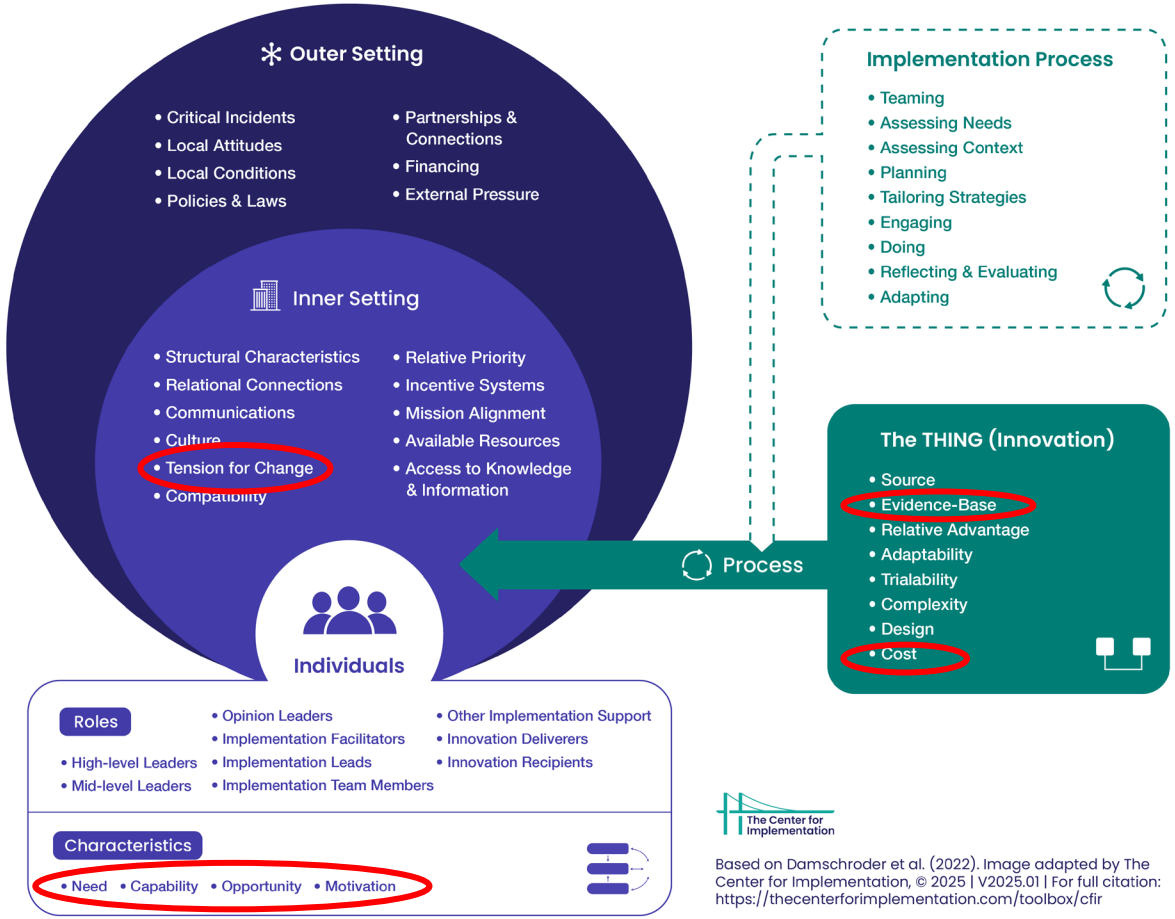
Table 1 Summary Details of NIH Pragmatic Trials Collaboratory Demonstration Projects

Project name	Project start/end date	Project focus	Results for primary outcome	Intervention status (sustained, spread, de-implemented)	Factors leading to sustainment or de-implementation
ABATE (Active Bathing to Eliminate Infection)	April 2014/February 2019	Evaluate chlorhexidine bathing plus mupirocin versus regular soap bathing to reduce bloodstream infection in hospitalized patients	Null	*Sustained* for those with medical devices in HCA *Spread* to other systems via partnership with AHRQ	*Benefits outside primary outcome*—Subgroup of patients with medical devices saw reduction of bloodstream infections *Policy*—Study team partners with AHRQ. Policy penalizes infections for urinary catheters and central lines
LIRE (Lumbar Imaging with Reporting of Epidemiology)	October 2013/December 2018	Evaluate impact of including standard epidemiologic information in lumbar spine imaging reports to reduce spine-related healthcare utilization after imaging	Null	*Sustained* in 2 of 4 health systems	*Benefits outside primary outcome*—Secondary analysis showed a reduction in subsequent opioid prescriptions *Resource intensive*—For health systems that sustained the intervention, there were no additional costs required to do so
PPACT (Collaborative Care for Chronic Pain in Primary Care)	April 2014/February 2018	Test the use of CBT interventions in primary care settings to improve chronic pain for patients in long-term opioid therapy	Positive	*Modified sustainment* of less resource-intense versions of the intervention	*Resource intensive*—Heavy staffing costs and system needs led to feasibility concerns. Cost analysis cannot be conducted until all claims are received resulting in a lack of cost-effectiveness data when considering intervention sustainment



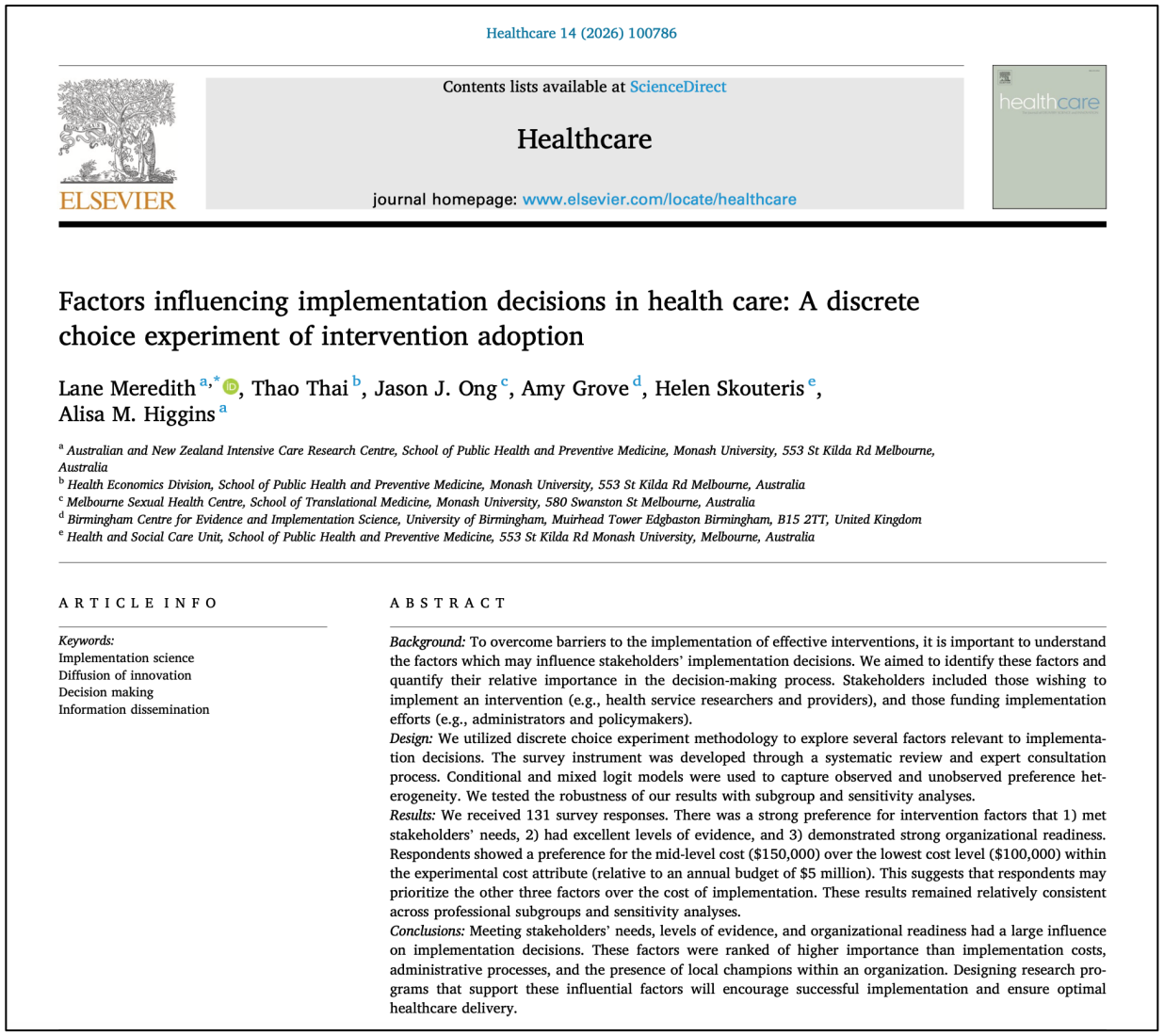
Implementation Determinants

Consolidated Framework for Implementation Research (CFIR)



The Center for Implementation

Based on Damschroder et al. (2022). Image adapted by The Center for Implementation, © 2025 | V2025.01 | For full citation: https://thecenterforimplementation.com/toolbox/cfir



Healthcare 14 (2026) 100786

Contents lists available at ScienceDirect

Healthcare

journal homepage: www.elsevier.com/locate/healthcare

Factors influencing implementation decisions in health care: A discrete choice experiment of intervention adoption

Lane Meredith [a,*], Thao Thai [b], Jason J. Ong [c], Amy Grove [d], Helen Skouteris [e], Alisa M. Higgins [a]

[a] Australian and New Zealand Intensive Care Research Centre, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Rd Melbourne, Australia
[b] Health Economics Division, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Rd Melbourne, Australia
[c] Melbourne Sexual Health Centre, School of Translational Medicine, Monash University, 580 Swanston St Melbourne, Australia
[d] Birmingham Centre for Evidence and Implementation Science, University of Birmingham, Muirhead Tower Edgbaston Birmingham, B15 2TT, United Kingdom
[e] Health and Social Care Unit, School of Public Health and Preventive Medicine, 553 St Kilda Rd Monash University, Melbourne, Australia

ARTICLE INFO

Keywords:
Implementation science
Diffusion of innovation
Decision making
Information dissemination

ABSTRACT

Background: To overcome barriers to the implementation of effective interventions, it is important to understand the factors which may influence stakeholders' implementation decisions. We aimed to identify these factors and quantify their relative importance in the decision-making process. Stakeholders included those wishing to implement an intervention (e.g., health service researchers and providers), and those funding implementation efforts (e.g., administrators and policymakers).

Design: We utilized discrete choice experiment methodology to explore several factors relevant to implementation decisions. The survey instrument was developed through a systematic review and expert consultation process. Conditional and mixed logit models were used to capture observed and unobserved preference heterogeneity. We tested the robustness of our results with subgroup and sensitivity analyses.

Results: We received 131 survey responses. There was a strong preference for intervention factors that 1) met stakeholders' needs, 2) had excellent levels of evidence, and 3) demonstrated strong organizational readiness. Respondents showed a preference for the mid-level cost ($150,000) over the lowest cost level ($100,000) within the experimental cost attribute (relative to an annual budget of $5 million). This suggests that respondents may prioritize the other three factors over the cost of implementation. These results remained relatively consistent across professional subgroups and sensitivity analyses.

Conclusions: Meeting stakeholders' needs, levels of evidence, and organizational readiness had a large influence on implementation decisions. These factors were ranked of higher importance than implementation costs, administrative processes, and the presence of local champions within an organization. Designing research programs that support these influential factors will encourage successful implementation and ensure optimal healthcare delivery.

Surveyed 131 healthcare professionals, researchers, managers, senior leaders, policymakers to explore factors related to implementation decisions:

- Total cost
- Administrative processes
- Meeting stakeholders' needs (patients and staff)
- Local champions
- Organizational readiness
- Level of evidence

	Intervention A	Intervention B
Total cost	$100,000	$200,000
Administrative processes	Expedited process (6 months) – simple and quick	Rigorous process (1 year) - unknown or arduous
Meeting stakeholders' needs	Not at all – Innovation does not address staff and patient issues	Somewhat – Innovation addresses the concerns of some staff members and patients
Local champion	Senior-level support	Collegiate-level support (peer-level)
Organizational readiness	Somewhat ready – Staff are open to change but do not feel entirely capable of adopting the changes	Not ready at all - Staff are not open to change. Staff do not feel capable of carrying out the necessary changes
Level of evidence	Poor – Innovation has limited, low-quality evidence	Good – Innovation has preliminary evidence

	Intervention A	Intervention B
Total cost	$100,000	$200,000
Administrative processes	Expedited process (6 months) – simple and quick	Rigorous process (1 year) - unknown or arduous
Meeting stakeholders' needs	Not at all – Innovation does not address staff and patient issues	Somewhat – Innovation addresses the concerns of some staff members and patients
Local champion	Senior-level support	Collegiate-level support (peer-level)
Organizational readiness	Somewhat ready – Staff are open to change but do not feel entirely capable of adopting the changes	Not ready at all - Staff are not open to change. Staff do not feel capable of carrying out the necessary changes
Level of evidence	Poor – Innovation has limited, low-quality evidence	Good – Innovation has preliminary evidence

Most important attributes:

1. Meeting stakeholders' needs
2. Level of evidence
3. Organizational readiness

Cost was important, but respondents often willing to pay more for interventions that demonstrated improvements in above attributes

Clinician/Staff Needs: Available measures



Weiner *et al. Implementation Science* (2017) 12:108
DOI 10.1186/s13012-017-0635-3

Implementation Science

RESEARCH Open Access

Psychometric assessment of three newly developed implementation outcome measures

Bryan J. Weiner[1*], Cara C. Lewis[2,3,4], Cameo Stanick[5], Byron J. Powell[6], Caitlin N. Dorsey[2], Alecia S. Clary[6], Marcella H. Boynton[7] and Heather Halko[8]

Abstract

Background: Implementation outcome measures are essential for monitoring and evaluating the success of implementation efforts. Yet, currently available measures lack conceptual clarity and have largely unknown reliability and validity. This study developed and psychometrically assessed three new measures: the Acceptability of Intervention Measure (AIM), Intervention Appropriateness Measure (IAM), and Feasibility of Intervention Measure (FIM).

Methods: Thirty-six implementation scientists and 27 mental health professionals assigned 31 items to the constructs and rated their confidence in their assignments. The Wilcoxon one-sample signed rank test was used to assess substantive and discriminant content validity. Exploratory and confirmatory factor analysis (EFA and CFA) and Cronbach alphas were used to assess the validity of the conceptual model. Three hundred twenty-six mental health counselors read one of six randomly assigned vignettes depicting a therapist contemplating adopting an evidence-based practice (EBP). Participants used 15 items to rate the therapist's perceptions of the acceptability, appropriateness, and feasibility of adopting the EBP. CFA and Cronbach alphas were used to refine the scales, assess structural validity, and assess reliability. Analysis of variance (ANOVA) was used to assess known-groups validity. Finally, half of the counselors were randomly assigned to receive the same vignette and the other half the opposite vignette; and all were asked to re-rate acceptability, appropriateness, and feasibility. Pearson correlation coefficients were used to assess test-retest reliability and linear regression to assess sensitivity to change.

Results: All but five items exhibited substantive and discriminant content validity. A trimmed CFA with five items per construct exhibited acceptable model fit (CFI = 0.98, RMSEA = 0.08) and high factor loadings (0.79 to 0.94). The alphas for 5-item scales were between 0.87 and 0.89. Scale refinement based on measure-specific CFAs and Cronbach alphas using vignette data produced 4-item scales (α's from 0.85 to 0.91). A three-factor CFA exhibited acceptable fit (CFI = 0.96, RMSEA = 0.08) and high factor loadings (0.75 to 0.89), indicating structural validity. ANOVA showed significant main effects, indicating known-groups validity. Test-retest reliability coefficients ranged from 0.73 to 0.88. Regression analysis indicated each measure was sensitive to change in both directions.

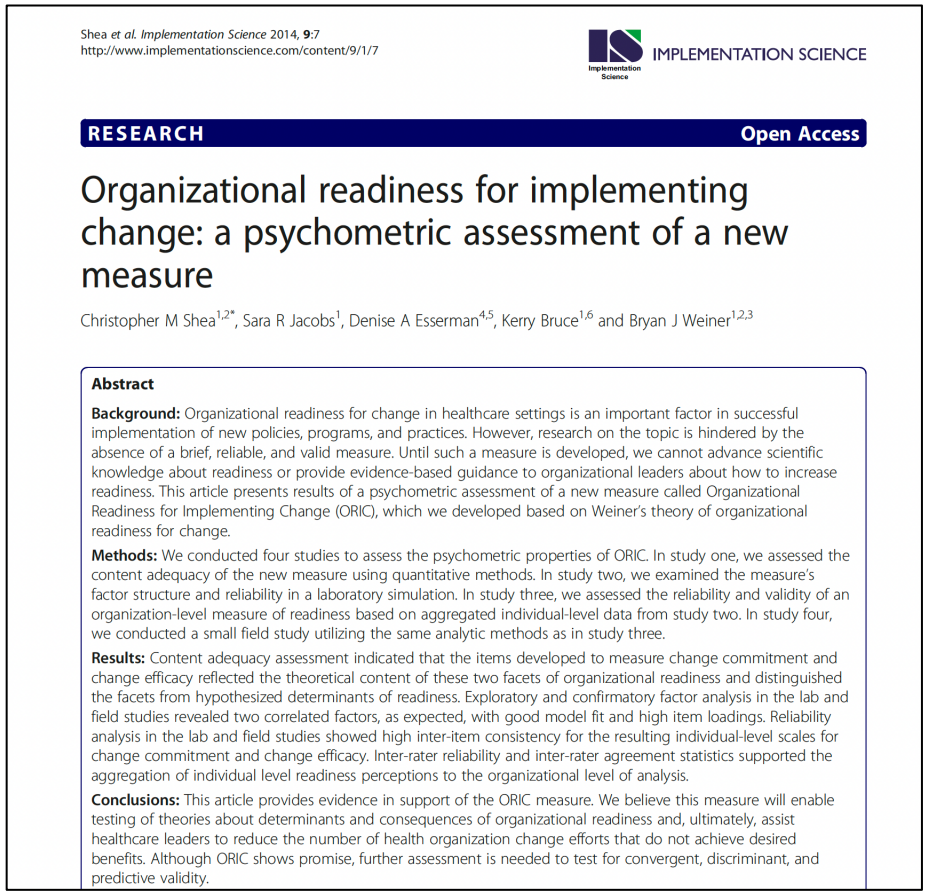
Conclusions: The AIM, IAM, and FIM demonstrate promising psychometric properties. Predictive validity assessment is planned.

Keywords: Acceptability of Intervention Measure (AIM), Intervention Appropriateness Measure (IAM), Feasibility of Intervention Measure (FIM), Implementation research, Acceptability, Feasibility, Appropriateness, Implementation outcomes, Measure, Structural validity, Test-retest, Known-groups

- **<u>Appropriateness</u>: Does it fit our clinical context and needs?**
- <u>Acceptability</u>: Do clinicians and staff like it?
- <u>Feasibility</u>: Can we realistically implement it?

Weiner BJ et al., Implementation Science, 2017

Organizational Readiness: Available measures



Shea *et al. Implementation Science* 2014, **9**:7
http://www.implementationscience.com/content/9/1/7

IMPLEMENTATION SCIENCE

RESEARCH Open Access

Organizational readiness for implementing change: a psychometric assessment of a new measure

Christopher M Shea[1,2*], Sara R Jacobs[1], Denise A Esserman[4,5], Kerry Bruce[1,6] and Bryan J Weiner[1,2,3]

Abstract

Background: Organizational readiness for change in healthcare settings is an important factor in successful implementation of new policies, programs, and practices. However, research on the topic is hindered by the absence of a brief, reliable, and valid measure. Until such a measure is developed, we cannot advance scientific knowledge about readiness or provide evidence-based guidance to organizational leaders about how to increase readiness. This article presents results of a psychometric assessment of a new measure called Organizational Readiness for Implementing Change (ORIC), which we developed based on Weiner's theory of organizational readiness for change.

Methods: We conducted four studies to assess the psychometric properties of ORIC. In study one, we assessed the content adequacy of the new measure using quantitative methods. In study two, we examined the measure's factor structure and reliability in a laboratory simulation. In study three, we assessed the reliability and validity of an organization-level measure of readiness based on aggregated individual-level data from study two. In study four, we conducted a small field study utilizing the same analytic methods as in study three.

Results: Content adequacy assessment indicated that the items developed to measure change commitment and change efficacy reflected the theoretical content of these two facets of organizational readiness and distinguished the facets from hypothesized determinants of readiness. Exploratory and confirmatory factor analysis in the lab and field studies revealed two correlated factors, as expected, with good model fit and high item loadings. Reliability analysis in the lab and field studies showed high inter-item consistency for the resulting individual-level scales for change commitment and change efficacy. Inter-rater reliability and inter-rater agreement statistics supported the aggregation of individual level readiness perceptions to the organizational level of analysis.

Conclusions: This article provides evidence in support of the ORIC measure. We believe this measure will enable testing of theories about determinants and consequences of organizational readiness and, ultimately, assist healthcare leaders to reduce the number of health organization change efforts that do not achieve desired benefits. Although ORIC shows promise, further assessment is needed to test for convergent, discriminant, and predictive validity.

Organizational Readiness for Implementing Change (ORIC)

(Adapt the questionnaire to your purposes by specifying [the change] you are investigating inside the brackets)

		1 disagree	2 somewhat disagree	3 Neither agree nor disagree	4 somewhat agree	5 agree
1	People who work here feel confident that the organization can get people invested in implementing *[this change]*.	○	○	○	○	○
2	People who work here are committed to implementing *[this change]*.	○	○	○	○	○
3	People who work here feel confident that they can keep track of progress in implementing *[this change]*.	○	○	○	○	○
4	People who work here will do whatever it takes to implement *[this change]* delivery.	○	○	○	○	○
5	People who work here feel confident that the organization can support people as they adjust to *[this change]*.	○	○	○	○	○
6	People who work here want to implement *[this change]*.	○	○	○	○	○
7	People who work here feel confident that they can keep the momentum going in implementing *[this change]*.	○	○	○	○	○
8	People who work here feel confident that they can handle the challenges that might arise in implementing *[this change]*.	○	○	○	○	○
9	People who work here are determined to implement *[this change]*.	○	○	○	○	○
10	People who work here feel confident that they can coordinate tasks so that implementation goes smoothly.	○	○	○	○	○
11	People who work here are motivated to implement *[this change]*.	○	○	○	○	○
12	People who work here feel confident that they can manage the politics of implementing *[this change]*.	○	○	○	○	○

Shea CM et al., Implementation Science, 2014

Takeaways

- Intervention effects (and their magnitude) are considered in the context of myriad other factors that impact implementation
- Cost-effectiveness analyses can help decision-makers interpret an intervention's effects in the context of its cost and resource requirements
- ePCT teams should consider measuring other constructs that – based on available data – may be important to decision-makers
  - Existing measures of appropriateness, organizational readiness
  - Qualitative data collection

Questions?

Please pose your questions in the Q&A

NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®