Working With Multiple Outcomes

Multicomponent Endpoints and Beyond

Moderator:
Emily O'Brien, PhD
Duke University



NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®

Objectives

- Examine the options for multicomponent endpoints and motivations for their use.
- Discuss strategies for handling multiple endpoints in pain trials.
- Review hierarchical composite endpoints and considerations for using them.

Our Panelists



Julie Fritz
University of Utah



Patrick Heagerty
University of Washington



Eric Roseen
Boston Medical Center



Pamela Shaw
Kaiser Permanente Washington Health Research Institute



Qilu Yu
NCCIH

Session Agenda

- Introduction (Emily O'Brien) – 5 min
- Presentations:
  - Overview of Hierarchical Composite Endpoints (Patrick Heagerty) – 10 min
  - Multicomponent Endpoint Approaches in Pain Trials (Julie Fritz) – 10 min
- Discussion and Audience Q&A (All Panelists) – 35 min
  - TAICHIKNEE Case Study (Eric Roseen)

Choosing Endpoints in Pain Clinical Trials

Julie Fritz
Distinguished Professor
Department of Physical Therapy & Athletic Training



What is the Challenge of Pain Assessments in Clinical Trials?

"The fact is, pain is defined introspectively by every man. The difficulty comes in verbalizing this well known experience… in saying what it is."

(H. Beecher, The Measurement of Pain, *Pharmacologic Rev*, 1957)

"Two of the most difficult problems in evaluating treatment efficacy in pain research involve what constitutes a 'successful' outcome and how best to measure it."

(D. Turk, et al, Core outcome domains for chronic pain clinical trials: IMMPACT recommendations, *PAIN*, 2003)

The Evolution of Pain Measurement in Clinical Trials

Key Shifts: From experimental measurement of nociception → quantitative measurement of the subjective pain experience → multidimensional measures

Historical Era & Conceptual Shift	Measurement Methods	Conceptual Impacts
1940s **Dolorimeter & Experimental Pain** Pain measured using instruments (dolorimeters) that apply controlled noxious stimuli to study thresholds and tolerance.	• Dolorimeters (thermal, electric, etc.) • Pain thresholds, tolerance time	Pain viewed as a physiologic response to a noxious stimuli
1950s–1960s **Beecher & the Gate Control Theory** Established that subjective pain can be quantified and studied scientifically; pain is actively modulated in the nervous system.	• Pain rating scales and categorical ratings	Pain is subjective, but measurable Pain is influenced and modulated by central processes
1970s **MPQ & Multidimensional Tools** Pain has sensory, affective, and evaluative dimensions that can be quantified.	• McGill Pain Questionnaire • Pain location and temporal patterns • Other multi-dimensional questionnaires	Pain as a multi-sensory phenomenon Instruments explicitly designed for clinical trials and practice

The Evolution of Pain Measurement in Clinical Trials

Key Shifts: From multidimensional measures → measuring multiple domains

Historical Era & Conceptual Shift	Measurement Methods	Conceptual Impacts
1980s–1990s **Expanding to Pain Impacts & Consequences** Recognition that pain intensity ≠ impacts of pain on function and interference with activities Focus on measuring multiple domains impacted by pain	• Multiple measures used to assess pain intensity and its impacts • New domains considered relevant such as mood, emotional distress, kinesiophobia, pain catastrophizing, etc. • Interest in single-item measures to capture patient perspectives (global ratings of change, etc.)	Established that pain, function, mood, quality of life, etc. are distinct, yet related constructs Growing emphasis on patient-centeredness, understanding the patient's individual pain experience Clinical trials increasingly use different measures and assess different domains – creates concerns about standardization
The Biopsychosocial Model Expansion of the domains considered relevant to understanding chronic pain and impact of treatments		
The Fear Avoidance Model Along with the biopsychosocial model, further helps explain discrepancies between pain pathology and chronic pain and disability.		

The Evolution of Pain Measurement in Clinical Trials

Key Shifts: From multiple pain measures and domains → standardization and core measure sets

Historical Era & Conceptual Shift	Measurement Methods	Conceptual Impacts
2000s–2010s **Advancing Measurement Science & Patient-Centered Outcomes** Proliferation of measures in pain clinical trials leads to calls for greater standardization. Greater focus on psychometrics Interest in real-world data and embedding PROs into routine practice increases **The IMMPACT Initiative** Develops core set of recommended outcome domains and measures for pain clinical trials (6 domains) Furthers the focus on clinical meaningfulness as a separate concept from statistical significance **CONSORT** Recommendations for clinical trial methods and reporting **PROMIS** Psychometrically-sound measurement tools for constructs relevant to pain able to be compared across conditions.	• Item Banks and Computer-Adaptive Measures for Multiple PROMIS domains • Recommended Core Measures (IMMPACT, 2005) **Pain Intensity** (NRS, categories, use of rescue analgesics) **Physical Function** (BPI, pain interference) **Emotional Function** (Beck Depression Inventory, Profiles of Mood States) **Participant Global Rating of Improvement and Satisfaction with Treatment** (PGIC) **Symptoms and Adverse Events** (passive capture with open-ended prompts) **Participant Disposition** (recruitment and progress through trial as specified in CONSORT)	Standardized outcome domains intended to improve comparability between trials Increased focus on defining meaningful change from perspective of patients or other stakeholders Clear pre-specification of methods and time-points for outcome assessment Distinction between primary and secondary trial outcomes Technological advances permit innovative strategies for gathering outcomes data and integrating into routine clinical care

Persistent Tensions in Pain Measurement for Clinical Trials

Pain is Multidimensional

Pain is shaped by interacting sensory, emotional, cognitive, behavioral, and contextual factors.

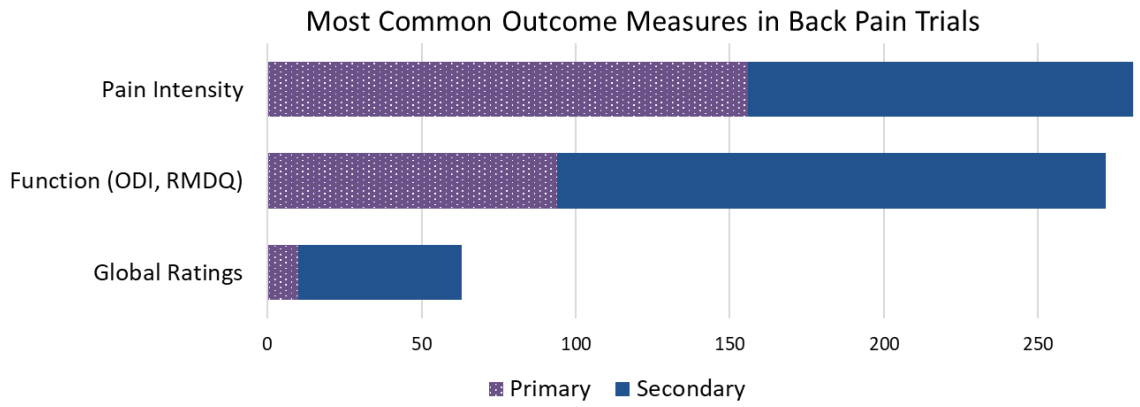
- Sensory (intensity, quality, location)
- Affective/ Emotional (distress, fear, sadness)
- Cognitive (attention, beliefs, catastrophizing)
- Behavioral/ Functional (activity, interference, participation)
- Contextual (social, cultural, environmental)

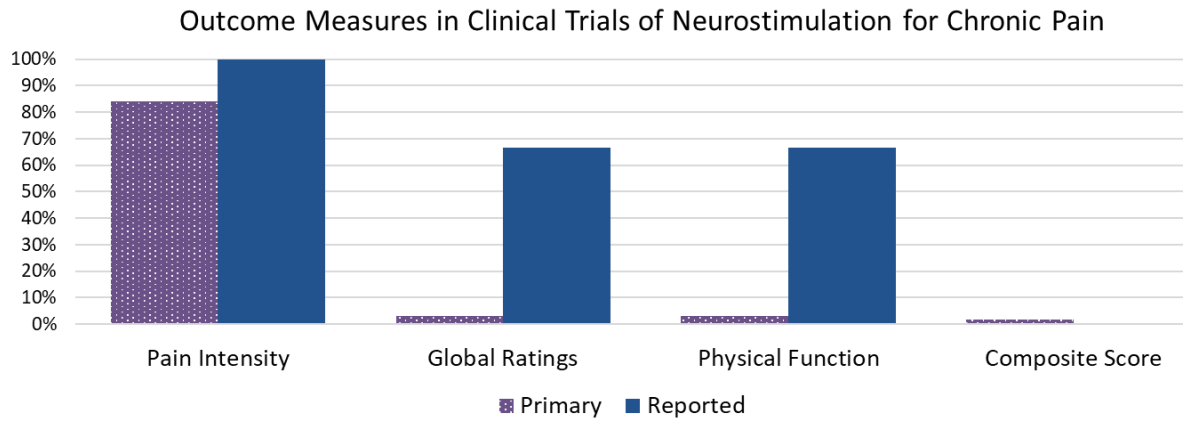
THE TENSION

Trials Rely on Single 1^0 Endpoints – Often Pain Intensity

Clinical trials require measures that are practical, efficient, responsive, and interpretable.

- Low burden for patients and staff
- Responsive to change
- Easy to analyze and interpret
- Standardized across studies and populations
- Regulatory acceptability and precedent





FROM: Froud R, et al. A Systematic Review of Outcome Measures Use, Analytical Approaches, Reporting Methods, and Publication Volume by Year in Low Back Pain Trials Published between 1980 and 2012. PLoS One. 2016;11(10):e0164573.

FROM: D'Souza RS, et al. Randomized Clinical Trials on Neurostimulation for Chronic Pain: Do They Comply With the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) Criteria? Neuromodulation. 2026 Jan;29(1):66-73.

Composite Outcomes: A Potential Next Step in Pain Trial Design

Moving from multiple outcomes to a clinically meaningful measure of overall treatment benefit

CURRENT PARADIGM

PRIMARY OUTCOME

Pain intensity

SECONDARY OUTCOMES

- Physical function
- Pain interference
- Sleep/emotional health
- Global improvement

COMPOSITE PRIMARY OUTCOME

- Pain Intensity +
- Physical Function +
- Pain Interference +
- Sleep/emotional health +
- Global Improvement +

ONE PRIMARY ENDPOINT

Define a clinically meaningful pattern of treatment success

POTENTIAL BENEFITS

More Patient-Centered

Captures improvement beyond pain intensity alone.

Preserves Multidimensionality

Makes pain, function, and other important domains part of the primary estimand.

Identifies a Single Primary Outcome

Provides one prespecified primary endpoint

Can Incorporate Benefits and Harms

Can create a risk/benefit profile for participants instead of treating harm and benefit separately

Composite Outcomes

Validation of a holistic composite outcome measure for the evaluation of chronic pain interventions

Taylor ,et al. *PAIN Reports* 9(6):p e1202, December 2024

The ***holistic minimal clinically important difference (MCID)*** is a composite outcome developed by an expert panel based on IMMPACT recommended outcomes.

Domain	Measure	Threshold for Baseline Impairment	Threshold for MCID Improvement
Pain Intensity	100mm VAS	≥ 60mm	≥ 30% decrease in VAS
Health-Related Quality of Life	EQ-5D-5L	EQ-5D ≤ 0.83	≥ 0.74 point increase in EQ-5D
Sleep Quality	Pittsburgh Sleep Quality Index	PSQI ≥ 6.3	≥3 point decrease in PSQI
Physical Function	Oswestry Disability Index	ODI ≥ 10.2	≥ 10 point decrease in ODI
Emotional Function	Profile of Mood State Brief Tool	POMS ≥ 17.7	≥ 10 point decrease in POMS mood

Cumulative Responder Score:

$$\frac{\text{Sum of total MCIDs achieved}}{\text{Number of domains impaired}}$$

MCID score = 1.0

indicates a clinically meaningful change was obtained on average across all domains impaired at baseline

Composite Outcomes

PAIN

Development and preliminary validation of an integrated efficacy–tolerability composite measure for the evaluation of analgesics

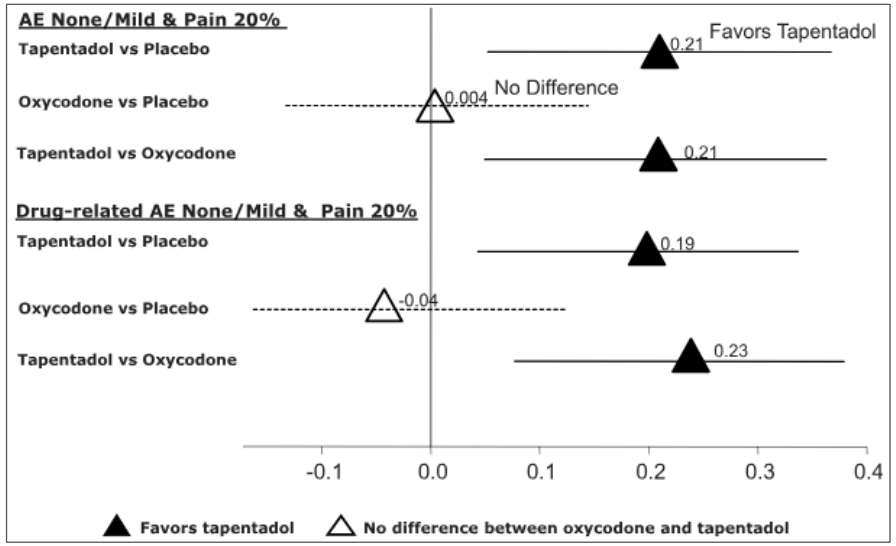
Katz, et al. *PAIN 156*(7):1357-1365, July, 2015

Goal was to develop, for an analgesic medication, an ***Efficacy-Tolerability Composite*** measure that integrates pain reduction (efficacy) and AEs (tolerability) for future evaluation and use.

Nine possible ETC criteria to determine whether a patient was an efficacy-tolerability responder:

efficacy \ tolerability	No AE	No/mild AE	No/mild drug-related AE
≥20% pain reduction	Y/N	Y/N	Y/N
≥30% pain reduction	Y/N	Y/N	Y/N
≥50% pain reduction	Y/N	Y/N	Y/N

Exploratory analyses of a Phase III trial comparing tapentadol, oxycodone, and placebo for chronic LBP



Composite Outcomes

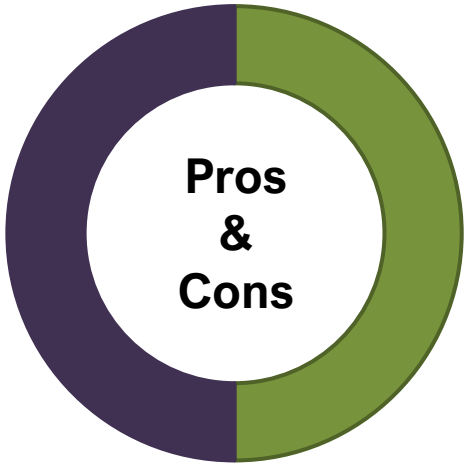
PROS

Clinical Relevance:
Can be more comprehensive by including multiple domains

Net Benefit Analysis:
Can incorporate benefits and harms within a single endpoint

Achieves Single Endpoint:
Meets criterion for a pre-specified single primary endpoint

Statistical Efficiency:
Can increase event rates



CONS

Misleading Interpretation:
Choice of components may not reflect priorities of patients

Assumption of Uniform Direction:
Interventions may not impact components in same direction

Equal Component Weighting:
Changes in less important outcomes may obscure more important changes

Statistical Inefficiency:
Including unresponsive components can reduce effects

Hierarchical Composite Outcomes

A Hierarchical Composite Outcome contains multiple components along with a ranking of order of clinical importance.

- ✓ Overcomes problem of equal weighting of composite outcomes
- ✓ Reflects real-world decision-making that often requires trade-offs

Hypothetical Chronic Pain Example – Desirability of Outcome Ranking (DOOR)

Rank	Outcome at 12 weeks	Desirability
1	≥50% pain reduction **and** meaningful improvement in function, with no important adverse event	**Most desirable**
2	≥30% pain reduction **and** meaningful functional improvement, with no important adverse event	Very desirable
3	≥50% pain reduction without meaningful functional improvement	Desirable
4	Meaningful functional improvement without ≥30% pain reduction	Moderately desirable
5	≥30% pain reduction or functional improvement, but with an important treatment-related AE	Less desirable
6	No meaningful improvement	Poor
7	Worsening pain/function and/or serious treatment-related harm	**Least desirable**

What is the probability that a patient receiving the intervention has a more desirable overall outcome than a patient receiving usual care?

How Often Are Composite Outcomes Used in Clinical Trials

Non-Pain Trials

Cardiovascular disease	**45.8%** of RCTs published from 2011-2017 used a composite outcome, about 80% as the primary outcome. (Shaikh et al, *Am Heart J*, 2020)
Nephrology	**17.4%** of nephrology trials from 2019–2023 used a composite primary outcome. (Fishbane et.al, *Kidney Med*, 2025)
Critical care	**39.1%** of critical-care trials from 2019-2023 use composite outcomes. (Fishbane et.al, *Kidney Med*, 2025)
COVID-19	**39.1%** of 417 COVID trials used a composite primary outcome, particularly trials of more severe disease. **1%** used a hierarchical composite.
Oncology	Composite/time-to-event endpoints such as progression-free survival, disease-free survival, and event-free survival are extensively used.

JAMA Network | Open

Published Online: July 28, 2026

Through 2024, 92 RCTs (1.8%) used hierarchical composite endpoints:

- ✓ 43.5% were in cardiology, 10.9% were infectious disease
- ✓ 43.5% were drug trials

Pain Trials

Nishtar et al, *Reg Anesth Pain Med, 2024*	Examined 73 RCTs in major pain journals from 2016-2021 and 2 (2.7%) had a composite outcome incorporating pain and harms.

How Often Are Composite Outcomes Used in Clinical Trials

Reasons Why Pain May be Well-Suited for Composite Clinical Trial Outcomes

- ✓ Pain is a multi-dimensional experience and outcomes are not easily captured by unidimensional instruments
- ✓ Treatments may improve some components but not others
- ✓ Patients' treatment priorities are heterogeneous
- ✓ Component outcomes are of similar ultimate gravity
- ✓ Treatments typically produce modest effects across multiple domains rather than dramatic improvement in one outcome.

Selecting Outcomes for Pain Pragmatic Clinical Trials



Key Consideration

"as the primary outcome becomes less recognizably important to patients, or is assessed on criteria seldom used in usual care, the trial becomes more explanatory [and less pragmatic]“. (PRECIS-2 Criteria, Loudon et al, 2015)

CHALLENGES

- ✓ Composite outcomes can be difficult to interpret
- ✓ Multiple measures may not be routinely available in EHR or administratively
- ✓ Degree of missingness may vary across components of a composite outcome

Summary

Conceptual Understanding of Pain Assessment



* Despite this evolution, pain intensity remains the most common primary endpoint
* The nature of the pain experience is in tension with regulatory guidance and practices for clinical trials that prioritize singular endpoints
* Composite outcomes seem well-suited to pain trials but have been rarely implemented
* Composite outcomes may introduce particular challenges for PCTs

Composite Endpoints

Patrick J. Heagerty PhD
Professor, Department of Biostatistics
Director, Center for Biomedical Statistics
University of Washington

UNIVERSITY of WASHINGTON



Overview

- **Composite endpoints are common**
  - e.g. Major adverse cardiovascular events (MACE)
  - e.g. Progression-free survival (PFS)
- **Combining elements of benefit and harm**
- **Hierarchical composites (win statistics)**
- **Q: Opportunities for pragmatic trials?**

Example: Pain Outcome (intensity + interference)

1. What number best describes your pain on average in the past week:

0	1	2	3	4	5	6	7	8	9	10
No pain										Pain as bad as you can imagine

2. What number best describes how, during the past week, pain has interfered with your enjoyment of life?

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

3. What number best describes how, during the past week, pain has interfered with your general activity?

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

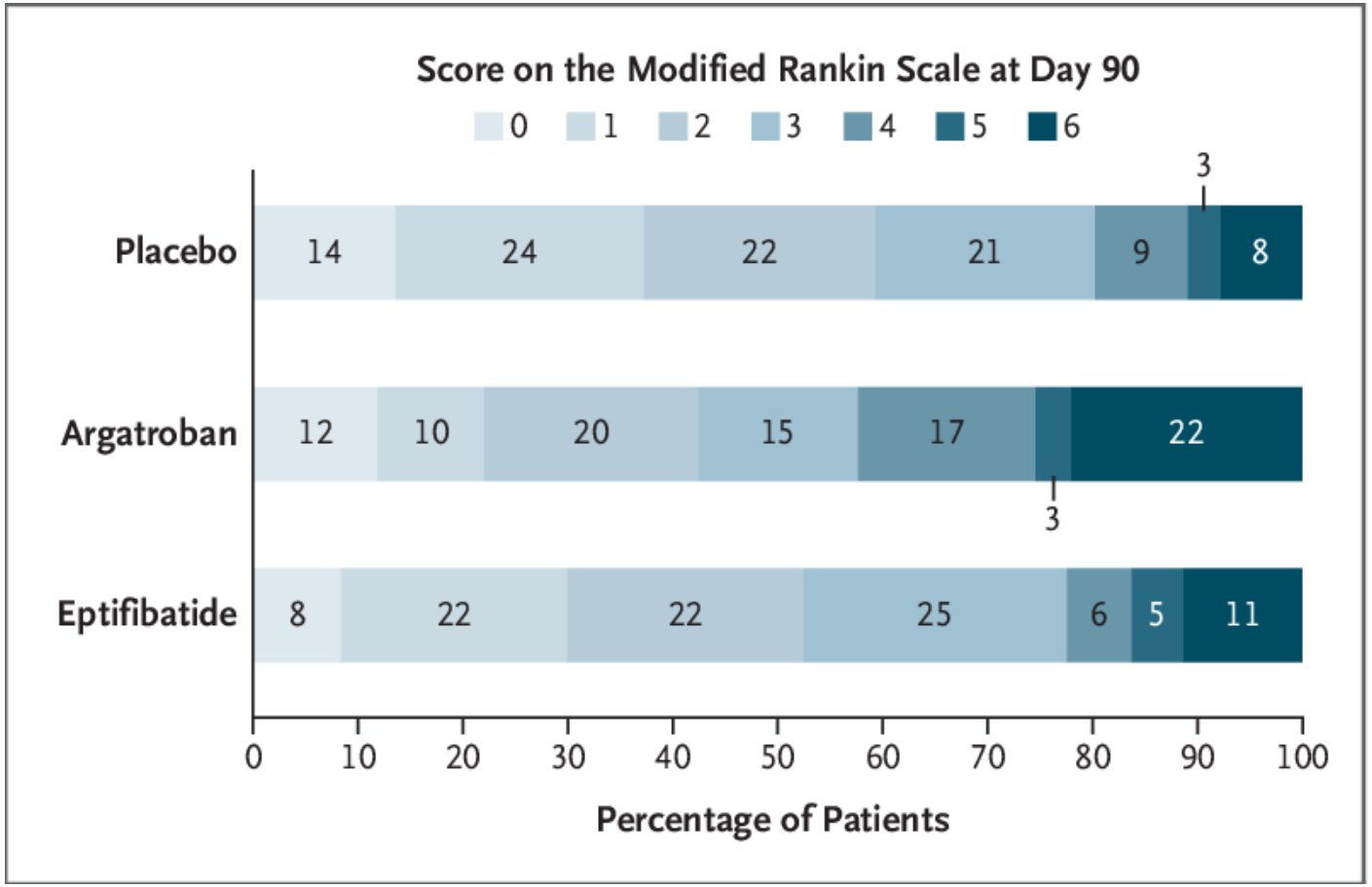
Average of these three items = PEG score

Example: Modified Rankin Scale (death, disability)

Definition	mRS	Measured construct
No symptoms	0	No complaints
No significant disability despite symptoms. Able to carry out all usual duties and activities.	1	Symptoms
Slight disability. Unable to carry out all previous activities, but able to look after own affairs without assistance.	2	Impairments in social roles
Moderate disability. Requiring some help, but able to walk without assistance.	3	Impairments in iADL tasks
Moderately severe disability. Unable to walk without assistance and unable to attend to own bodily needs without assistance.	4	Impairments in ADL tasks
Severe disability. Bedridden, incontinent, and requiring constant nursing care and attention.	5	Impairments in mobility
Death	6	Death



Example: Modified Rankin Scale (death, disability)



Adeoye et al. NEJM (2024)

Example: DOOR (benefit, harm)

Desirability of Outcome Ranking (DOOR)

- Evans et al. (2015)
- Composite of outcomes
  - Rank combinations of benefit and harm
  - Ordinal outcome
- Idea: look at joint distribution rather than both margins…

Table 3.

Overall Clinical Outcome for the SCOUT-CAP[a] Trial (From Most to Least Desirable)

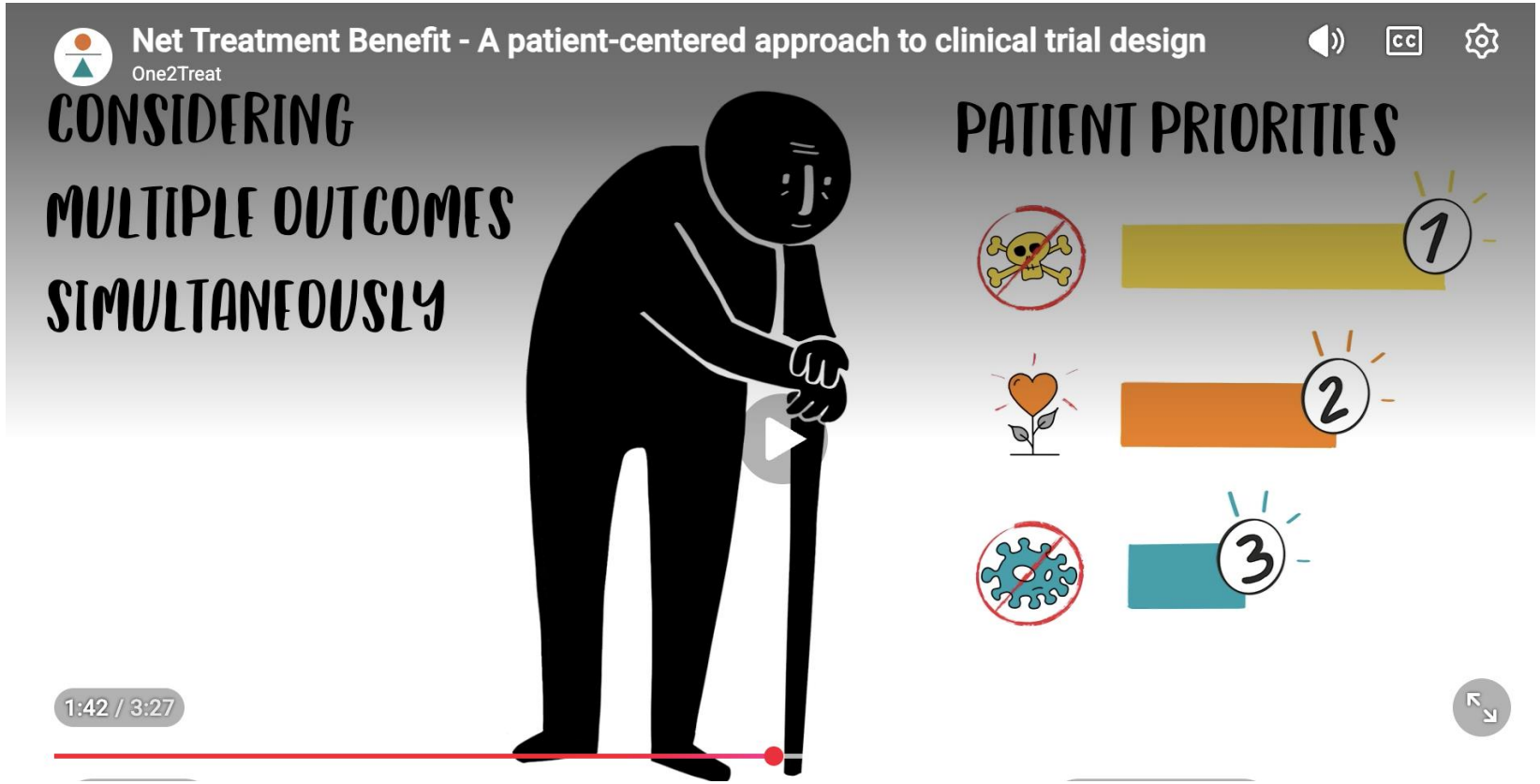
1. Survival; adequate clinical response; no adverse events
2. Survival; adequate clinical response; mild adverse event(s)
3. Survival; adequate clinical response; moderate adverse event(s)
4. Survival; adequate clinical response; severe adverse event(s)
5. Survival; inadequate clinical response without additional emergency department or clinic visit or hospitalization
6. Survival; inadequate clinical response with additional emergency department or clinic visit but without hospitalization; any grade of adverse event
7. Survival; inadequate clinical response with hospitalization; any grade of adverse event
8. Death

Adequate clinical response was defined based on absence of all of the following as assessed on day 11–14 after initiation of therapy: (1) fever unless related to a new process that is unrelated to the prior diagnosis of pneumonia, (2) tachypnea, (3) increased work of breathing (retractions, nasal flaring, grunting), and (4) a medically attended visit to an emergency department/clinic or hospitalization for persistent or worsening pneumonia at any time after randomization. This is a draft version from a developing clinical trial.

[a] Short-Course Outpatient Therapy for Community-Acquired Pneumonia in Children.

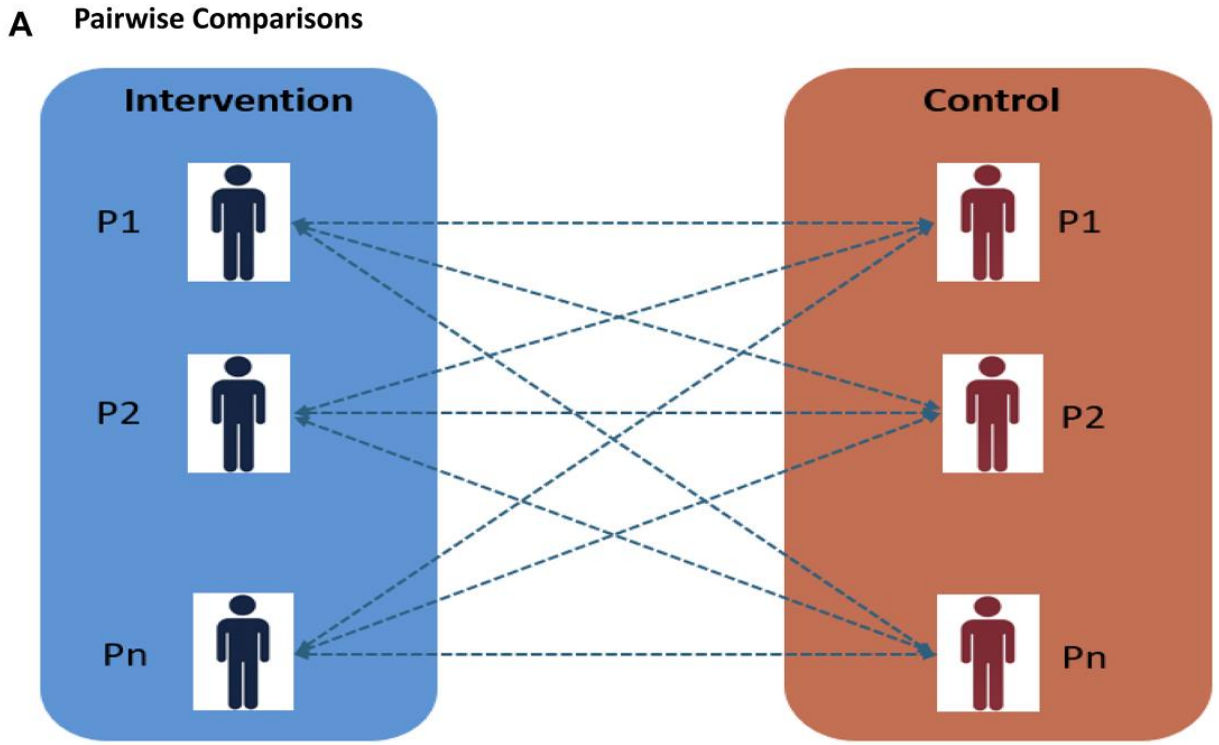
Evans et al. (2015) *Clinical Infectious Diseases*

Hierarchical Composite: Pairwise, sequentially



https://one2treat.com/resources/net-treatment-benefit/

Hierarchical Composite: Pairwise, sequentially

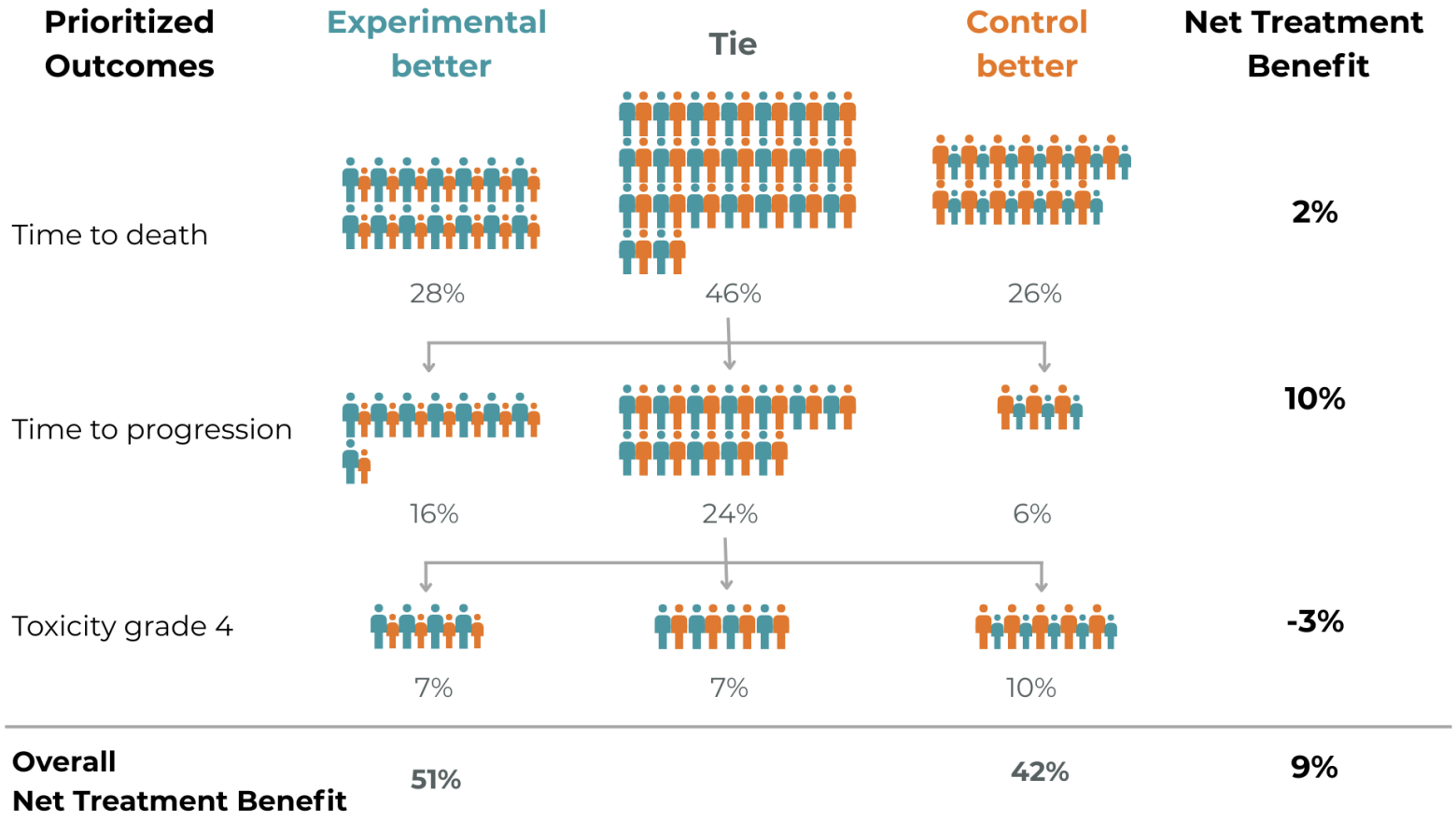


Outcomes:

- Time to death
- Time to progression
- Toxicity grade 4

Vart (2025) *eClinicalMedicine*

Hierarchical Composite: Pairwise, sequentially



https://one2treat.com/resources/net-treatment-benefit/

Hierarchical Composite: Pairwise, sequentially

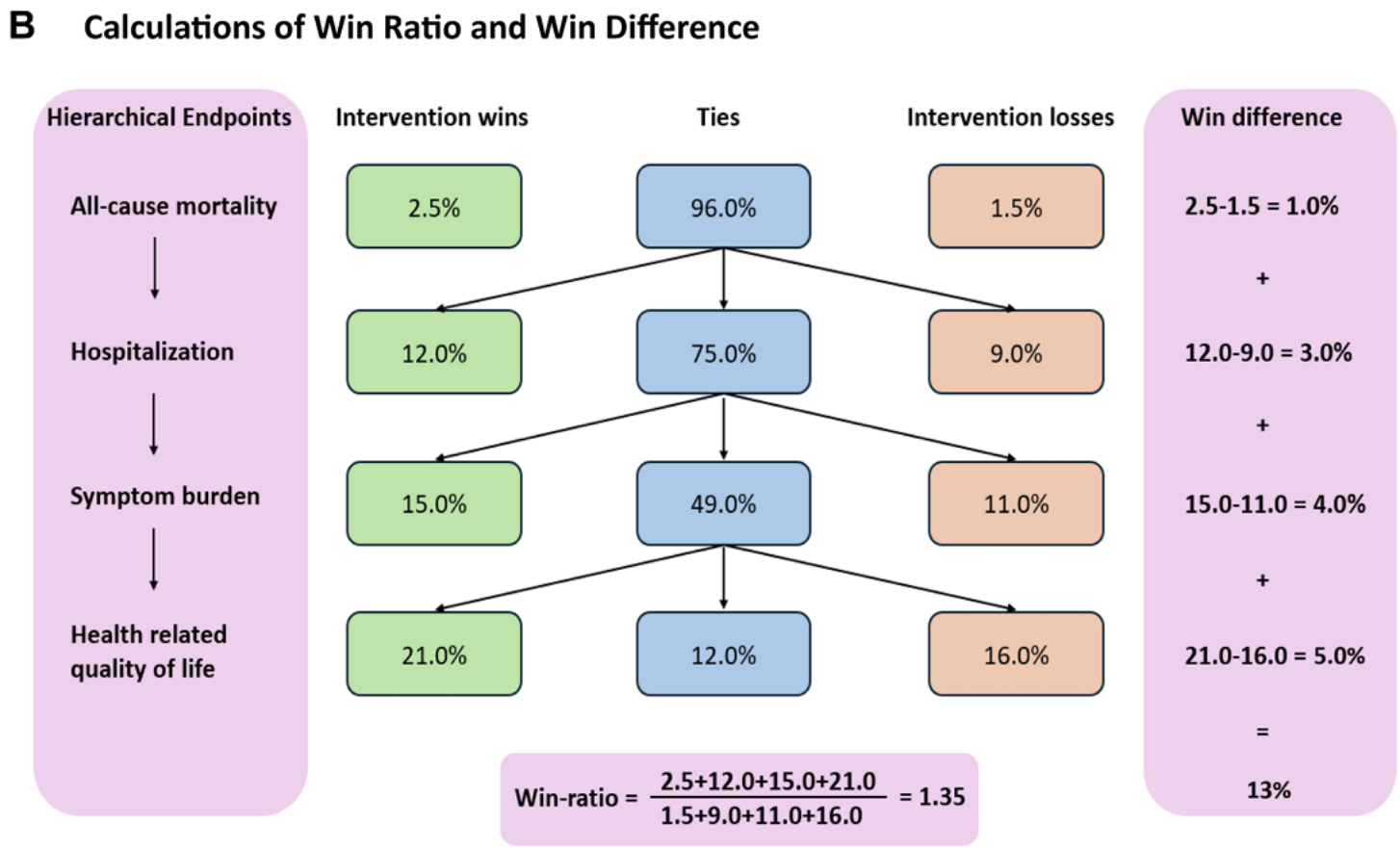


Fig. 1: Illustrative example of pairwise comparisons (A) and calculations of win ratio and win difference (B) in a hierarchical endpoint analysis.

Vart (2025) *eClinicalMedicine*

Hierarchical Composite: Pairwise, sequentially

JAMA Network Open

(2025)

Original Investigation | Surgery

Generalized Pairwise Comparisons to Support Shared Decision-Making in the CODA Trial

Samuel Salvaggio, PhD; Sarah E. Monsell, MSc; Patrick J. Heagerty, PhD; Mickael De Backer, PhD; Emilie Barré, MSc; Jean-Christophe Chiem, PhD; Everardo D. Saad, MD; Marc Buyse, ScD; David R. Flum, MD, MPH

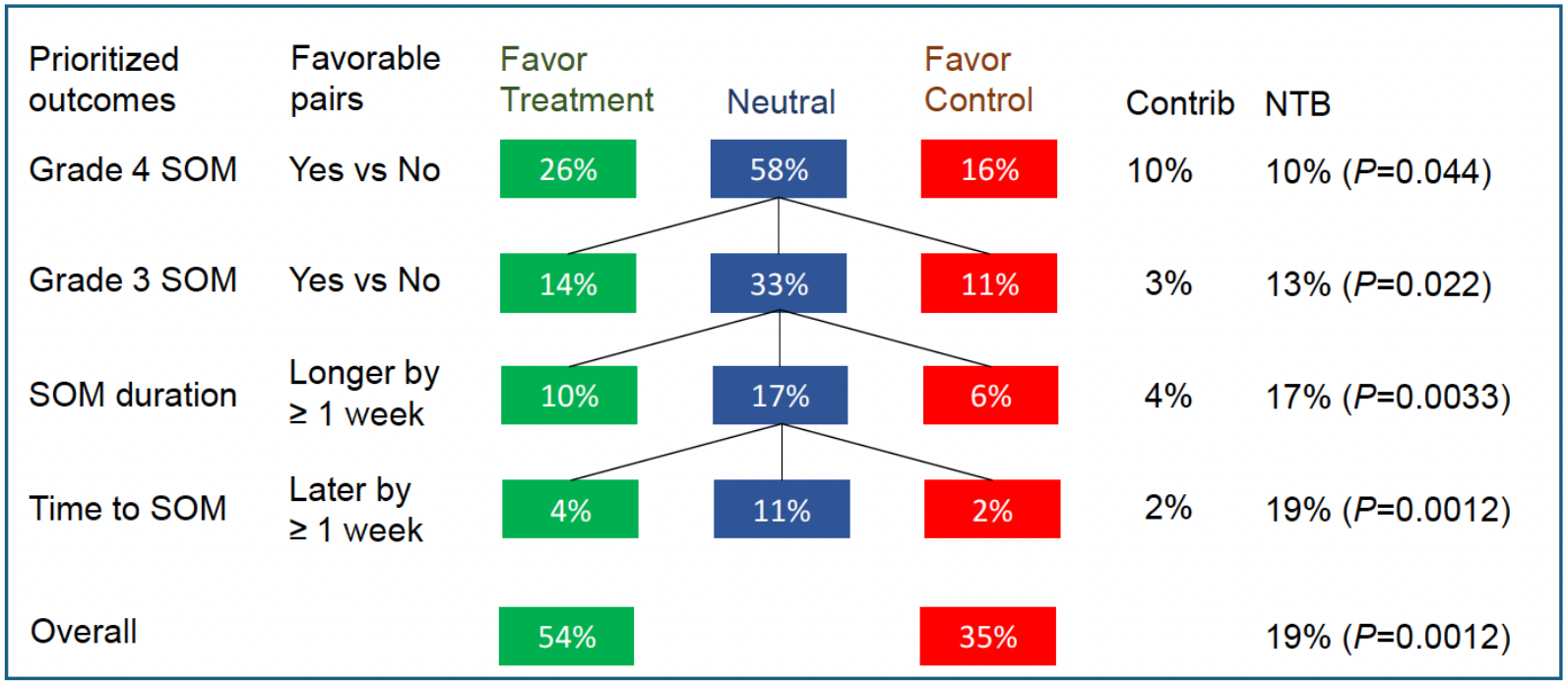
Figure 1. Schematic View of the Multivariate Generalized Pairwise Comparison Analysis of Scenario 1

Prioritized outcomes	Antibiotics better, %	Neutral, %	Surgery better, %	NTB, %
EQ-5D at 30 d	23.9	59.1	17.0	6.9
Symptom resolution	9.3	41.2	8.6	0.6
Any hospitalization	0.6	37.8	2.8	-2.2
Any drainage procedure	0.3	37.2	0.3	0
Days work missed by patient	5.9	28.1	3.2	2.7
Days work missed by caretaker	1.2	26.3	0.6	0.6
Length of hospital stay	15.3	0	11.1	4.1
NTB for all prioritized outcomes (P <.001)				12.8

Each pair is classified hierarchically through the list of outcomes. The pair is assessed on the first outcome, here the European Quality of Life–5 Dimensions (EQ-5D) score at 30 days. If the result favors or disfavors the treatment, the pair is classified accordingly. Here, 23.9% of all pairs favored antibiotics, while 17.0% of the pairs favored surgery. However, if the pair is neutral, the assessment is carried over to the subsequent outcome for further classification, here symptom resolution. Here 59.1% of all pairs were neutral for EQ-5D at 30 days and were therefore assessed by symptom resolution. The lines in the figure illustrate how neutral or tied matches are resolved at the next level of the hierarchy.

Hierarchical Composite: Pairwise, sequentially

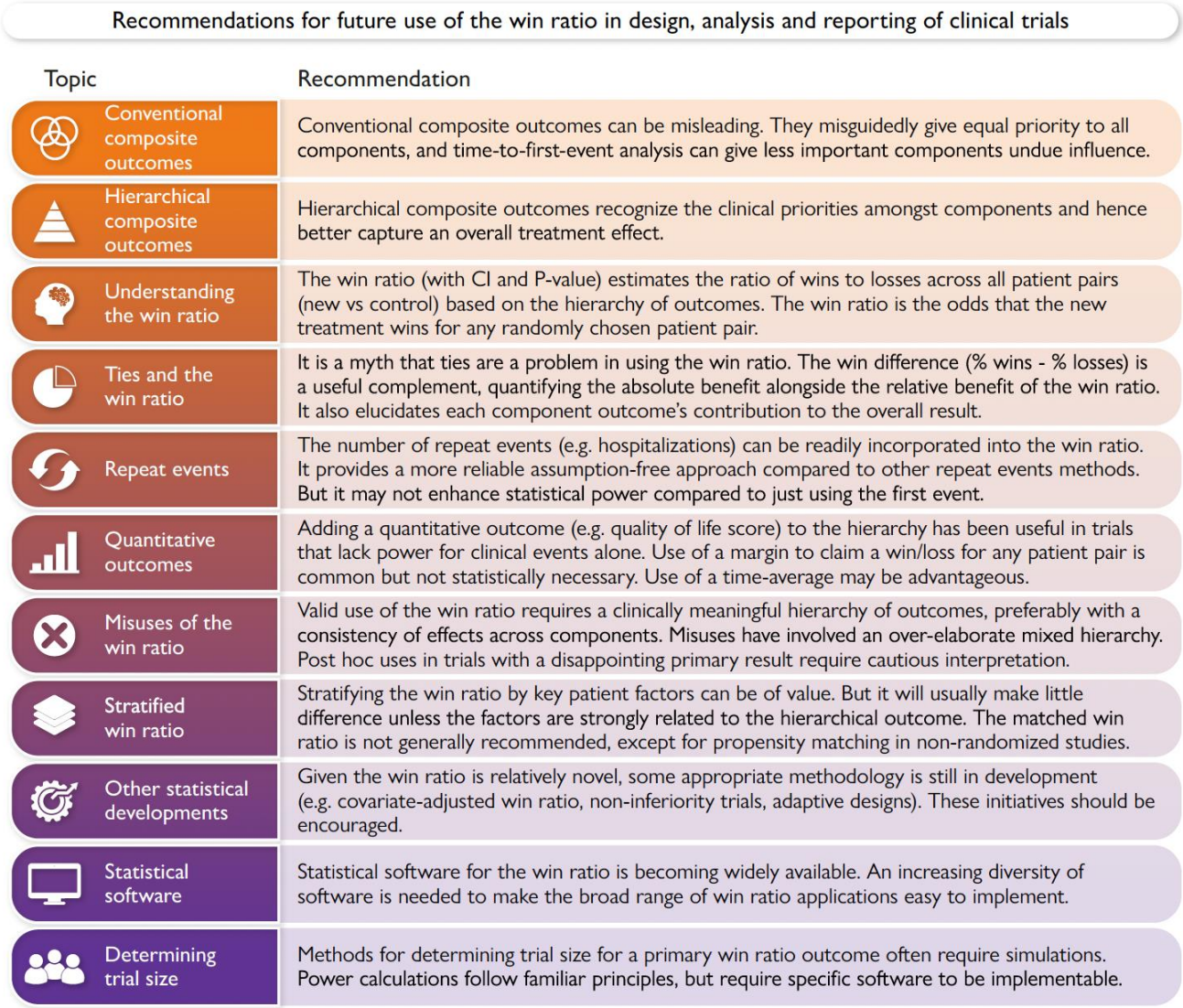
Figure 1: Illustration of GPC analysis of four prioritized outcomes



Salvaggio et al. (2024) *JAMA Network Open* (supplement)
Anderson et al. (2024) *Int J Rad Oncol, Biol, Phys*

Hierarchical Composite: Pairwise, sequentially

(2024)



Recommendations for future use of the win ratio in design, analysis and reporting of clinical trials

Topic	Recommendation
Conventional composite outcomes	Conventional composite outcomes can be misleading. They misguidedly give equal priority to all components, and time-to-first-event analysis can give less important components undue influence.
Hierarchical composite outcomes	Hierarchical composite outcomes recognize the clinical priorities amongst components and hence better capture an overall treatment effect.
Understanding the win ratio	The win ratio (with CI and P-value) estimates the ratio of wins to losses across all patient pairs (new vs control) based on the hierarchy of outcomes. The win ratio is the odds that the new treatment wins for any randomly chosen patient pair.
Ties and the win ratio	It is a myth that ties are a problem in using the win ratio. The win difference (% wins - % losses) is a useful complement, quantifying the absolute benefit alongside the relative benefit of the win ratio. It also elucidates each component outcome's contribution to the overall result.
Repeat events	The number of repeat events (e.g. hospitalizations) can be readily incorporated into the win ratio. It provides a more reliable assumption-free approach compared to other repeat events methods. But it may not enhance statistical power compared to just using the first event.
Quantitative outcomes	Adding a quantitative outcome (e.g. quality of life score) to the hierarchy has been useful in trials that lack power for clinical events alone. Use of a margin to claim a win/loss for any patient pair is common but not statistically necessary. Use of a time-average may be advantageous.
Misuses of the win ratio	Valid use of the win ratio requires a clinically meaningful hierarchy of outcomes, preferably with a consistency of effects across components. Misuses have involved an over-elaborate mixed hierarchy. Post hoc uses in trials with a disappointing primary result require cautious interpretation.
Stratified win ratio	Stratifying the win ratio by key patient factors can be of value. But it will usually make little difference unless the factors are strongly related to the hierarchical outcome. The matched win ratio is not generally recommended, except for propensity matching in non-randomized studies.
Other statistical developments	Given the win ratio is relatively novel, some appropriate methodology is still in development (e.g. covariate-adjusted win ratio, non-inferiority trials, adaptive designs). These initiatives should be encouraged.
Statistical software	Statistical software for the win ratio is becoming widely available. An increasing diversity of software is needed to make the broad range of win ratio applications easy to implement.
Determining trial size	Methods for determining trial size for a primary win ratio outcome often require simulations. Power calculations follow familiar principles, but require specific software to be implementable.

Pocock et al. (2024) *European Heart Journal*

Hierarchical Composite: Pairwise, sequentially

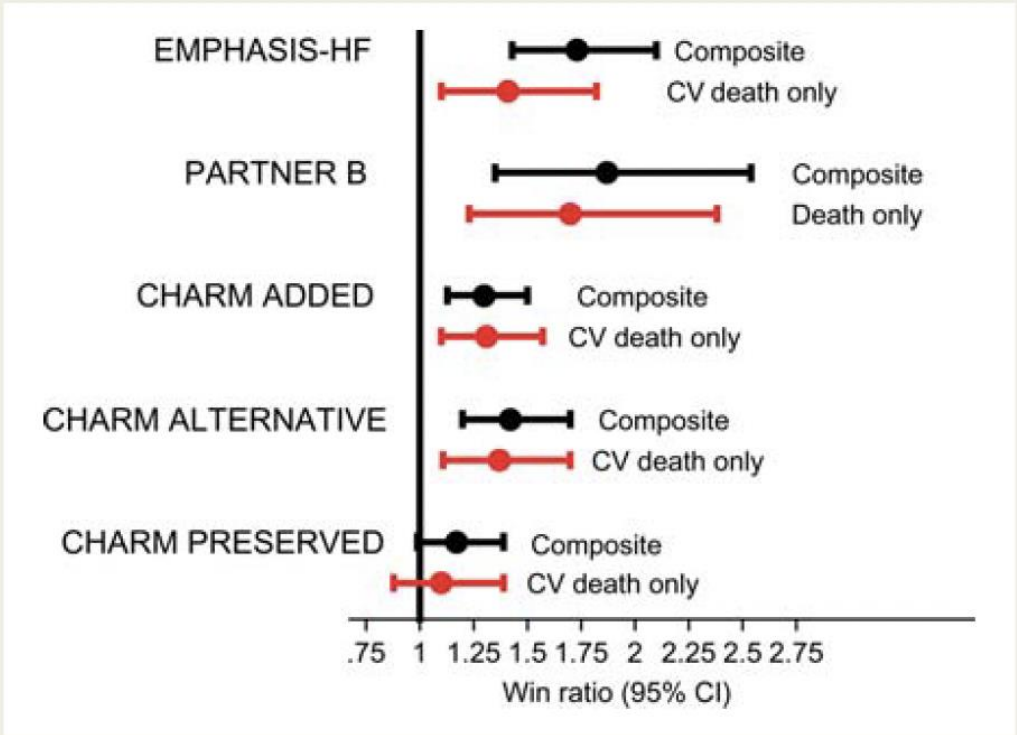


Figure 2 A graphical display of win ratio estimates (95% CI) for the three trials.

Pocock et al. (2012) *European Heart Journal*

Conclusions

- **Composite outcomes**
  - Patient reported outcome scales (quantitative)
  - Event-time outcomes (death, hospitalization)
  - Event-time and functional scales
  - Benefit and harm (DOOR)

- **Hierarchical composites**
  - Win Ratio
  - Net Treatment Benefit
  - R01 HL178513-01 “New win methods for addressing multiple and composite outcomes in cluster-randomized trials” (PI: Fan Li)

Key References:

- Pocock et al. (2012) *European Heart Journal*
- Pocock et al. (2024) *European Heart Journal*
- Evans et al. (2015) *Clinical Infectious Disease*
- Salvaggio et al. (2019) *JAMA Network Open*
- Buyse (2010) *Statistics in Medicine*

Questions?

Please pose your questions in the Q&A

NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®