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?

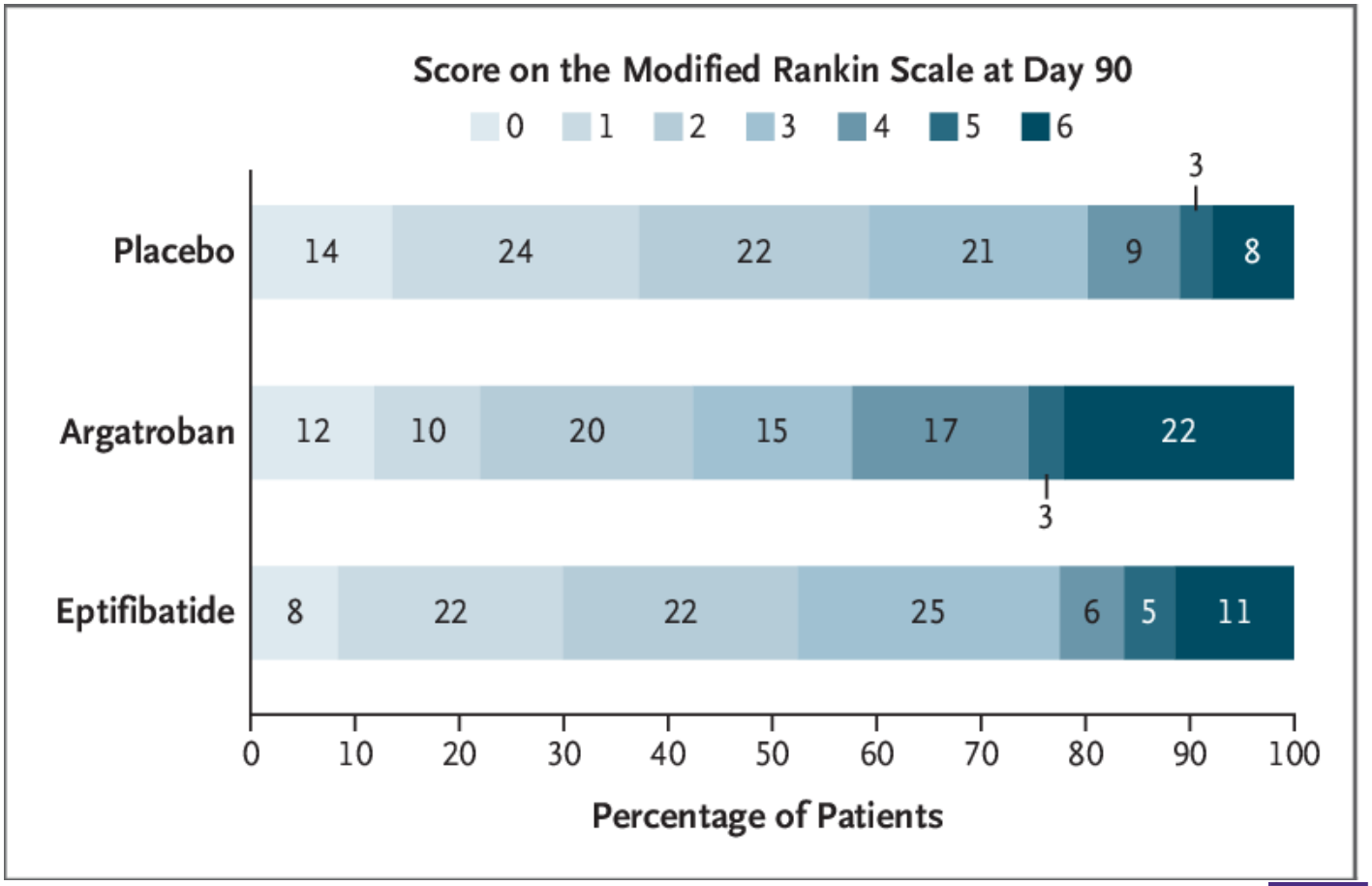
0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

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)

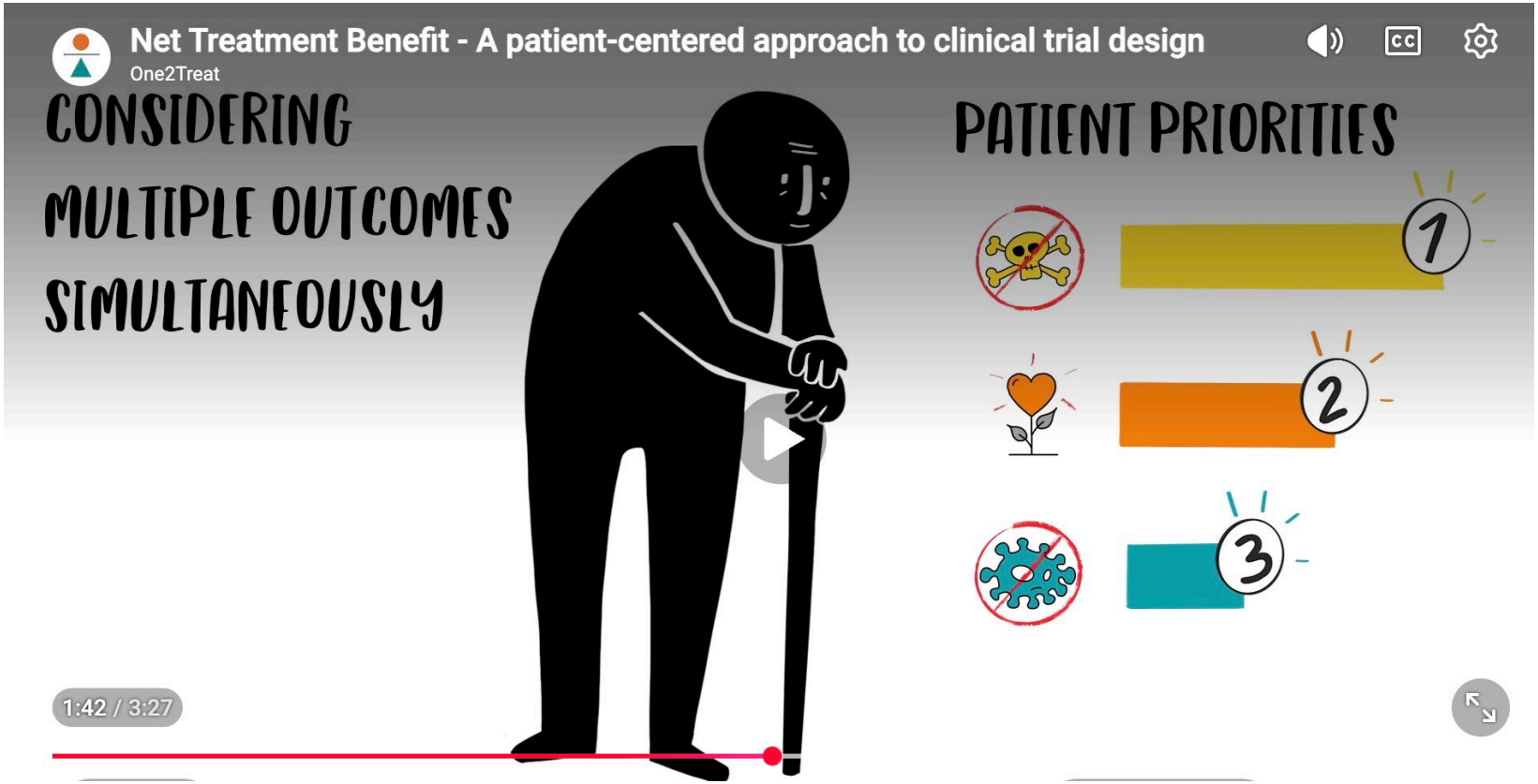
1. Survival; adequate clinical response; no adverse events
2. Survival; adequate clinical response; mild adverse event(s)
3. Survival; adequate clinical response; moderate adverse event(s)
4. Survival; adequate clinical response; severe adverse event(s)
5. Survival; inadequate clinical response without 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.

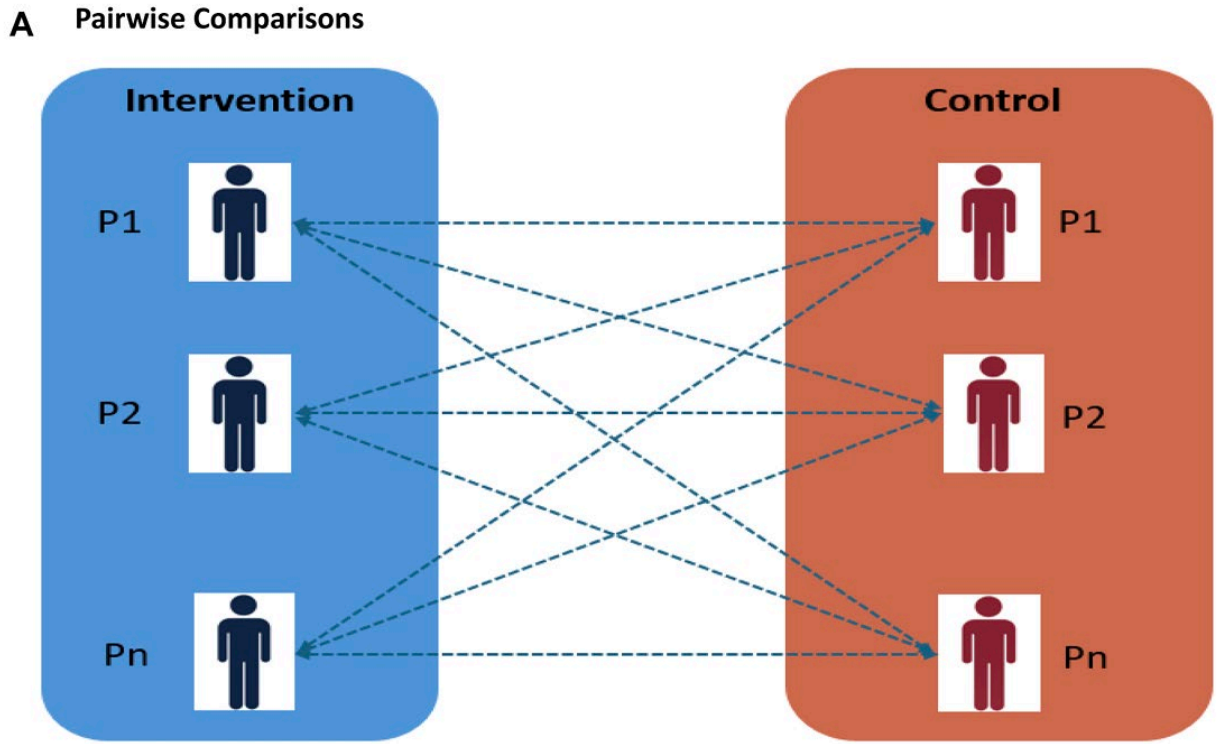
Evans et al. (2015) *Clinical Infectious Diseases*

Hierarchical Composite: Pairwise, sequentially



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

Hierarchical Composite: Pairwise, sequentially

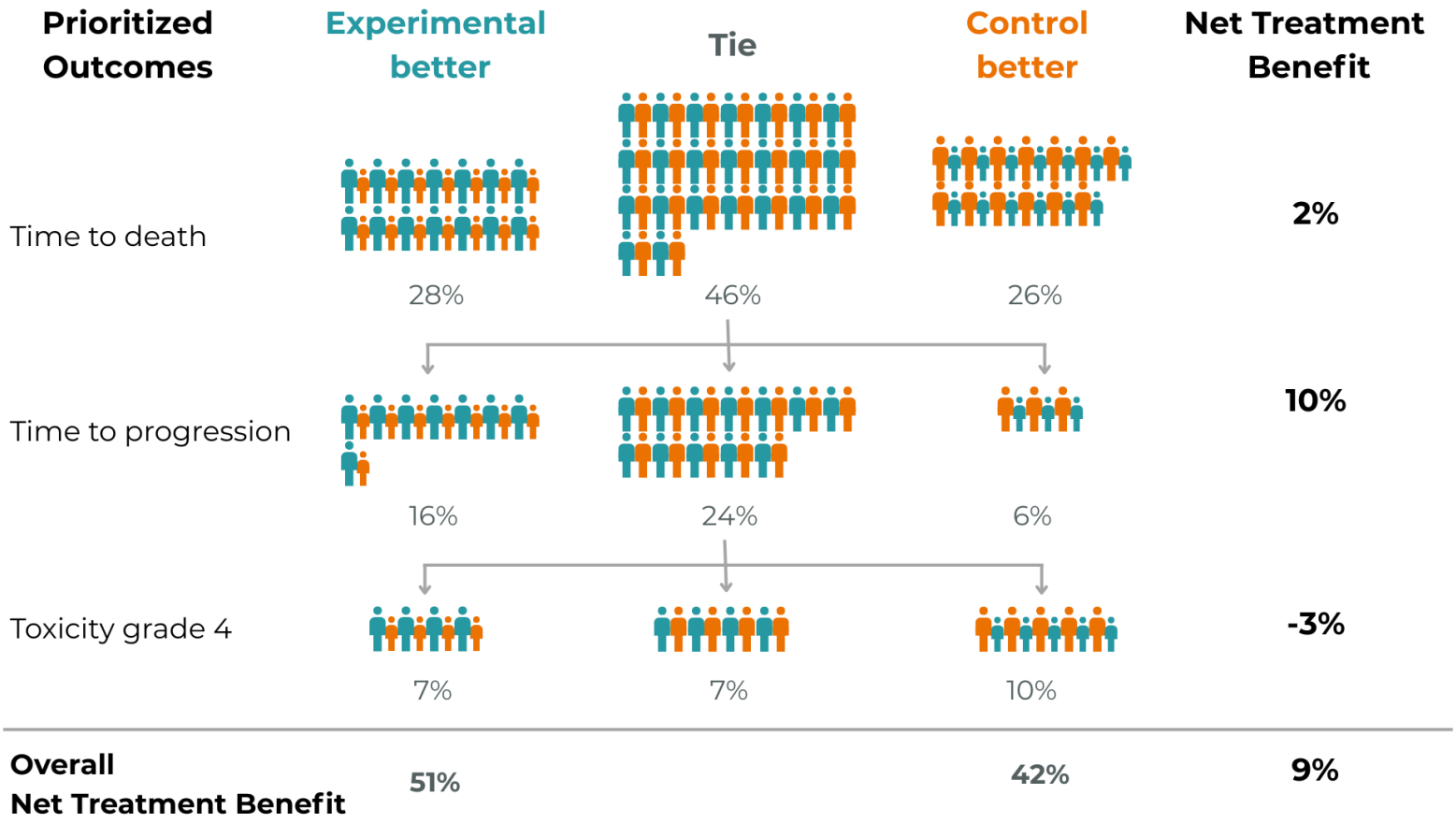


Outcomes:

- 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

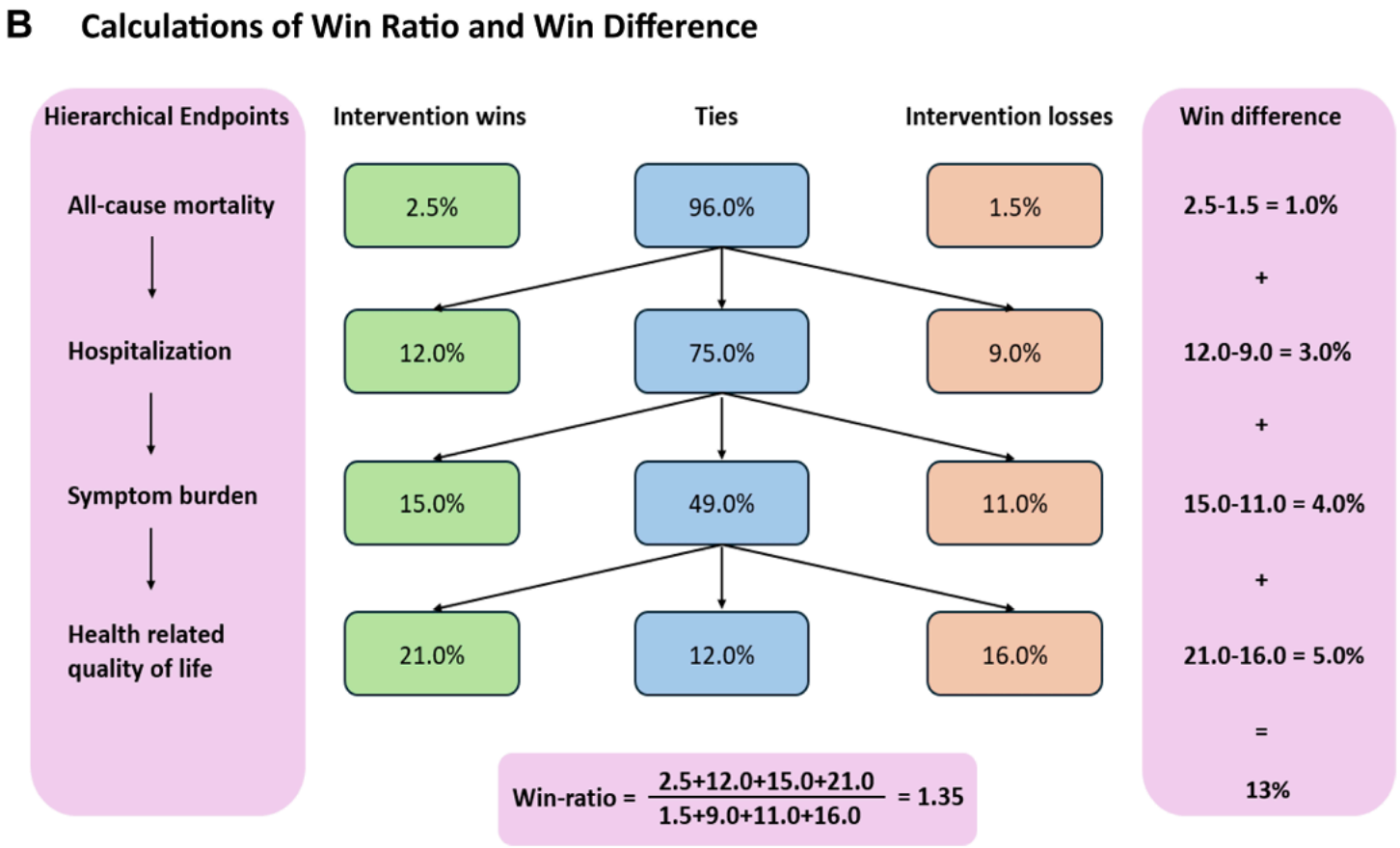


Fig. 1: Illustrative example of pairwise comparisons (A) and calculations of win ratio and win difference (B) in a hierarchical endpoint analysis.

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

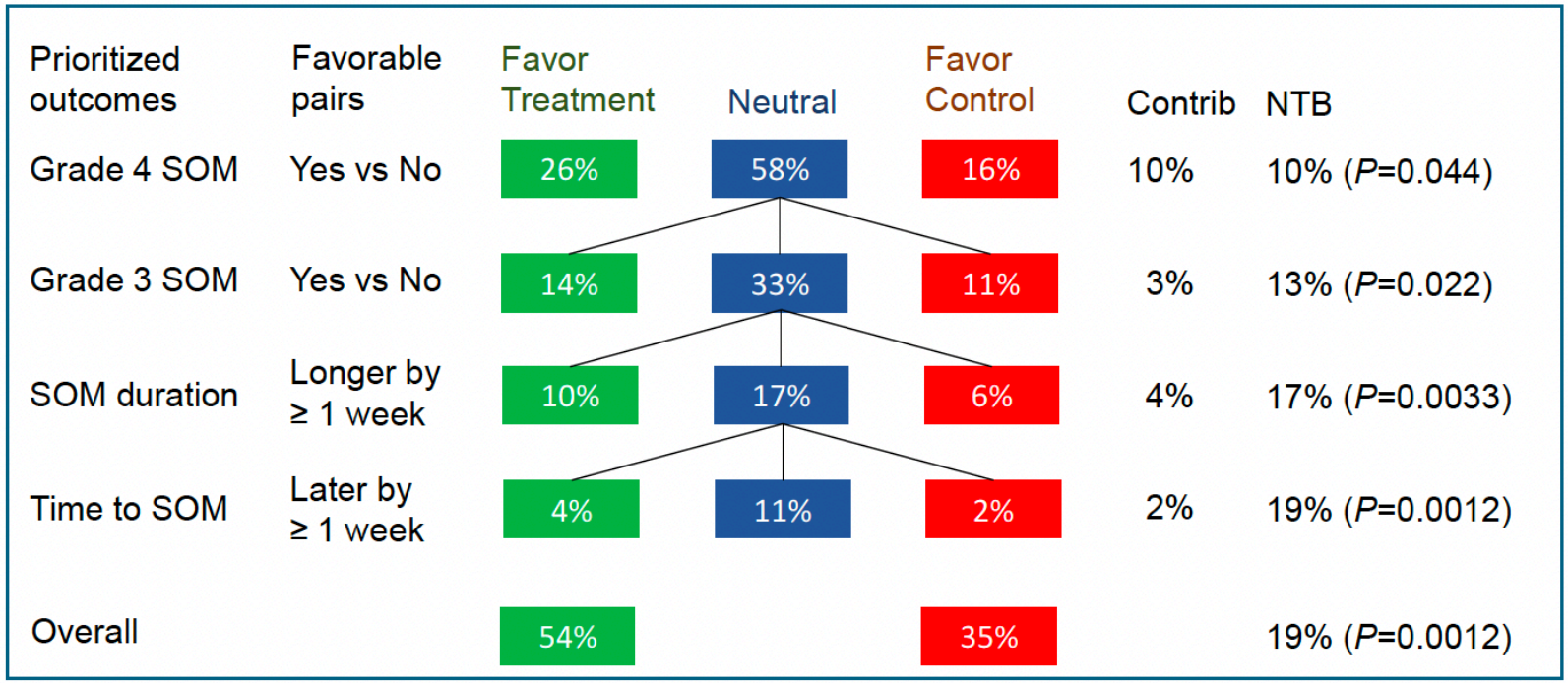
Figure 1. Schematic View of the Multivariate Generalized Pairwise Comparison Analysis of Scenario 1

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

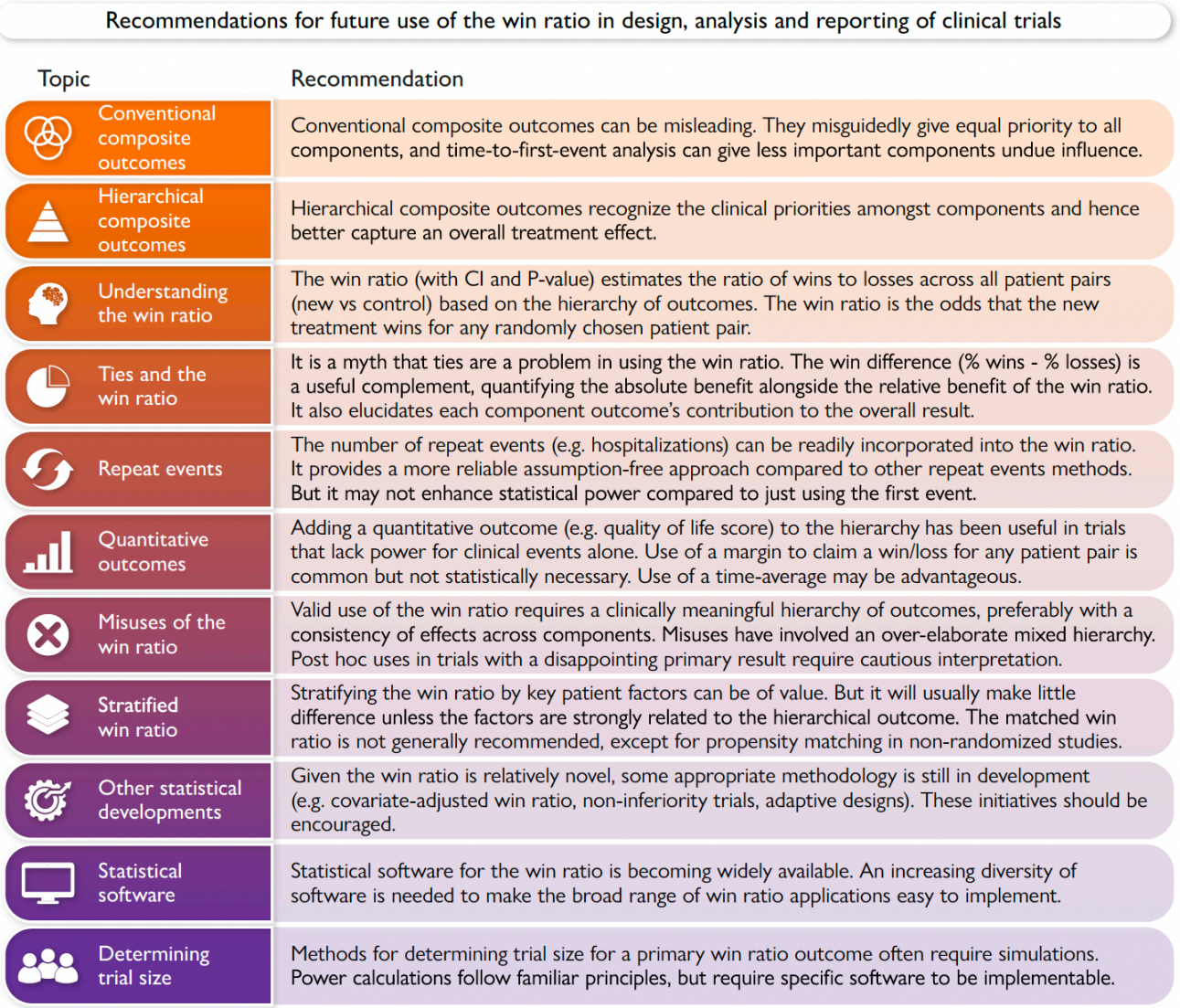
Figure 1: Illustration of GPC analysis of four prioritized outcomes



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

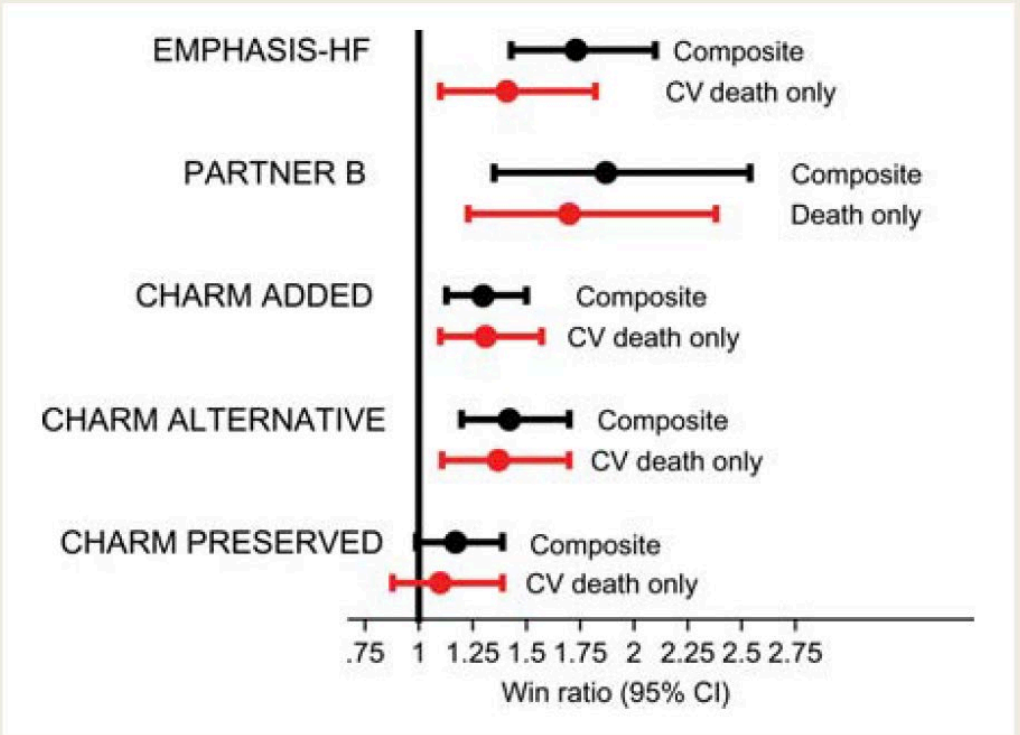


Figure 2 A graphical display of win ratio estimates (95% CI) for the three trials.

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*