St. Jude Children's Research Hospital
Finding cures. Saving children.

Western Tennessee ASA Chapter 2025 Annual Symposium

October 24, 2025 | 11:00 AM – 6:00 PM
St. Jude Children's Research Hospital
Marlo Thomas GEC Board Room

The Win Ratio and Beyond

Evolution, Applications, and Open Questions in Composite Endpoint Analysis

Lu Mao

lmao@biostat.wisc.edu

Department of Biostatistics & Medical Informatics

University of Wisconsin-Madison

Outline

- Introduction
- Properties and Use Cases
- Extensions and Variations
- Challenges and Open Questions
- Conclusion

Introduction

Composite Endpoint

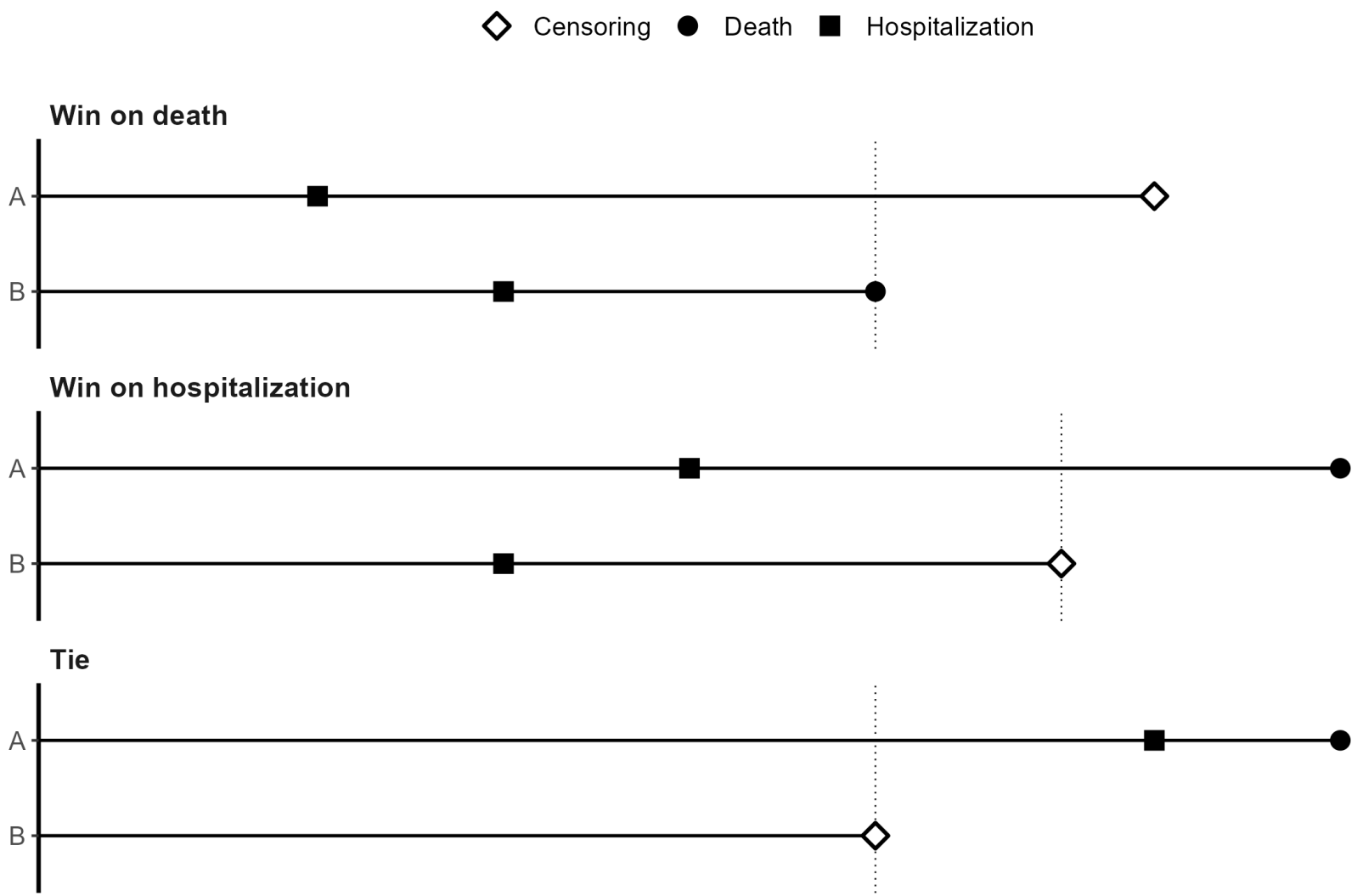
- **Traditional composite endpoint**
  - **Time to first event**
    - Major adverse cardiac event (MACE): death, heart failure, myocardio-infarction, stroke, etc.
  - **Limitations**:
    - Lack of clinical priority
    - Statistical inefficiency (waste of data)
- **Hierarchical composite endpoint (HCE)**
  - **Example**: Death > nonfatal MACE > six-minute walk test (6MWT)/ NYHA class

Win Ratio: Basics

- **A common HCE**
  - **Proposed and popularized** by Pocock et al. (2012)
  - **Treatment vs control**: generalized pairwise comparisons (GPC) (Buyse et al., 2025)
    - **Win-loss**: Longer survival $\rightarrow$ fewer/later nonfatal MACE $\rightarrow$ better 6MWT/NYHA
  - **Effect size**: WR $=$ wins / losses
    - Fold increase in probability of winning over lossing
- **Alternative metrics**
  - **Net benefit** (Proportion in favor): NB $=$ wins $-$ losses (Buyse, 2010)
  - **Win odds**: WO $=$ (wins $+ 2^{-1}$ties) / (losses $+ 2^{-1}$ties) (Brunner et al., 2021)

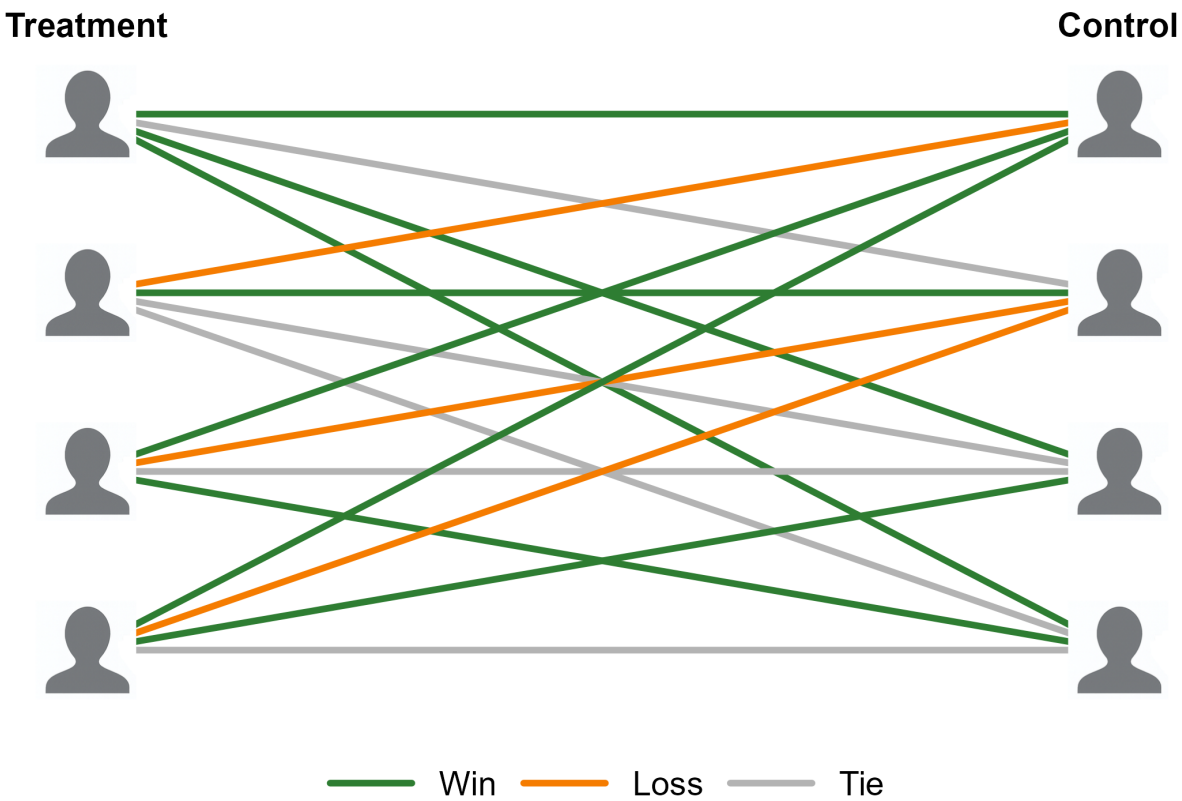
Prioritized Comparison

Cut off at earlier end of follow-up



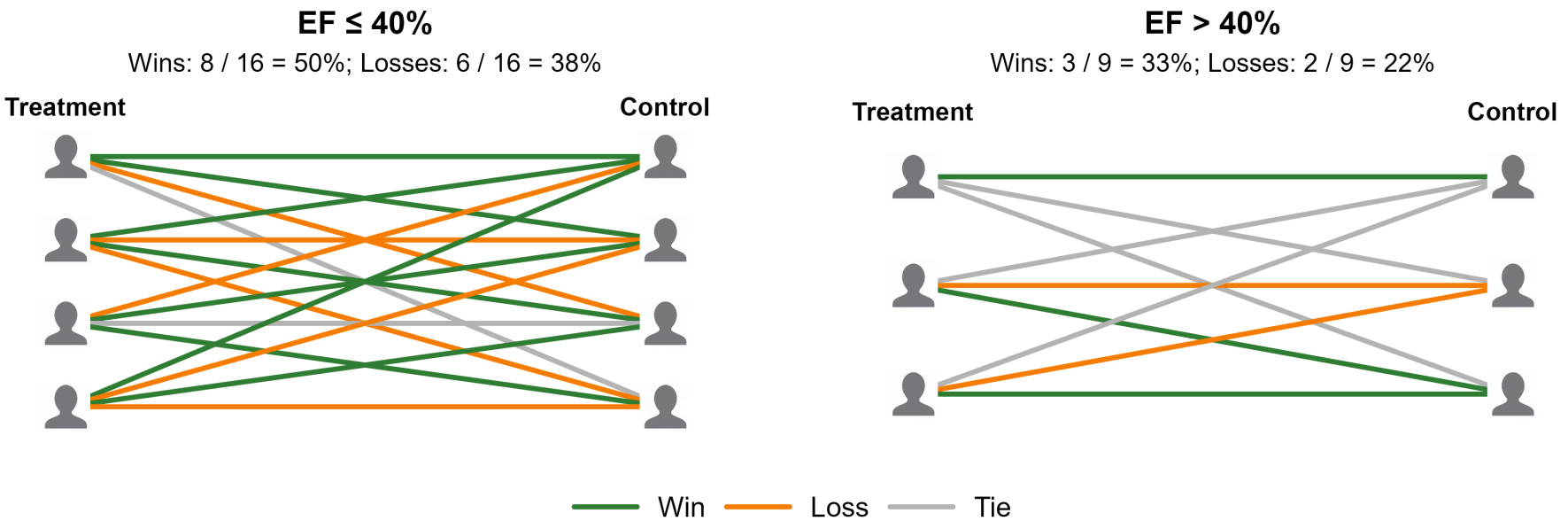
Pairwise Comparison

- **GPC Results**
  - Wins: 8 / 16 = 50%
  - Losses: 3 / 16 = 18.8%
  - Ties: 5 / 16 = 31.2%
- **Effect Sizes**
  - WR = 50% / 18.8% = 2.67
  - NB = 50% - 18.8% = 31.2%
  - WO = (50% + 15.6%) / (18.8% + 15.6%) = 1.72



Stratification

- **GPC within stratum and aggregate** (Dong et al., 2018, 2023)



$$\text{Wins: } 50\% \times 8/(8+6) + 33\% \times 6/(8+6) = 42.7\%$$
$$\text{Losses: } 38\% \times 8/(8+6) + 22\% \times 6/(8+6) = 31.1\%$$

- **Overall WR:** 42.7% / 31.1% = 1.37

Properties and Use Cases

General Setting

- **Notation**
  - n_1, n_0: sample sizes in treatment and control
  - W_{ij}: win-loss indicator of subject i in treatment vs j in control

$$W_{ij} = \begin{cases} 1, & \text{if subject } i \text{ in treatment wins} \\ -1, & \text{if subject } j \text{ in control wins} \\ 0, & \text{if tie} \end{cases}$$

    - Wins : $\hat{w}_{1,0} = (n_1 n_0)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} I(W_{ij} = 1)$
    - Losses : $\hat{w}_{0,1} = (n_1 n_0)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} I(W_{ij} = -1)$
  - $\mathrm{WR} = \hat{w}_{1,0} / \hat{w}_{0,1}$; $\mathrm{NB} = \hat{w}_{1,0} - \hat{w}_{0,1}$; $\mathrm{WO} = (1 + \mathrm{NB})/(1 - \mathrm{NB})$

Binary Case Equivalencies

- **Binary outcome, no censoring**
  - $Y_i^{(a)} = 1$ (response) vs 0 (no-response) for subject i in group a ($a = 0, 1$)
  - Win-loss (1-0) fractions

$$\hat{w}^{(a,1-a)} = (n_a n_{1-a})^{-1} \sum_{i=1}^{n_a} \sum_{j=1}^{n_{1-a}} Y_i^{(a)} (1 - Y_j^{(1-a)}) = \hat{p}_a (1 - \hat{p}_{1-a})$$

    - $\hat{p}_a = n_a^{-1} \sum_{i=1}^{n_a} Y_i^{(a)}$ = response rate in group a

- **GPC stats**
  - WR $= [\hat{p}_1(1 - \hat{p}_0)]/[\hat{p}_0(1 - \hat{p}_1)] =$ response odds ratio
  - NB $= \hat{p}_1 - \hat{p}_0 =$ response rate difference

Variance Estimation

- **Bootstrap**
  - Inital proposal (Pocock et al., 2012)
- **Analytic asymptotic variance**
  - Asymptotic decompostion of U-statistics (Bebu & Lachin, 2016; Dong et al., 2016; Luo et al., 2015)
  - Influence function-based
  - **R packages**: `WR`, `WINS`, `BuyseTest`, `WinRatio`, etc.

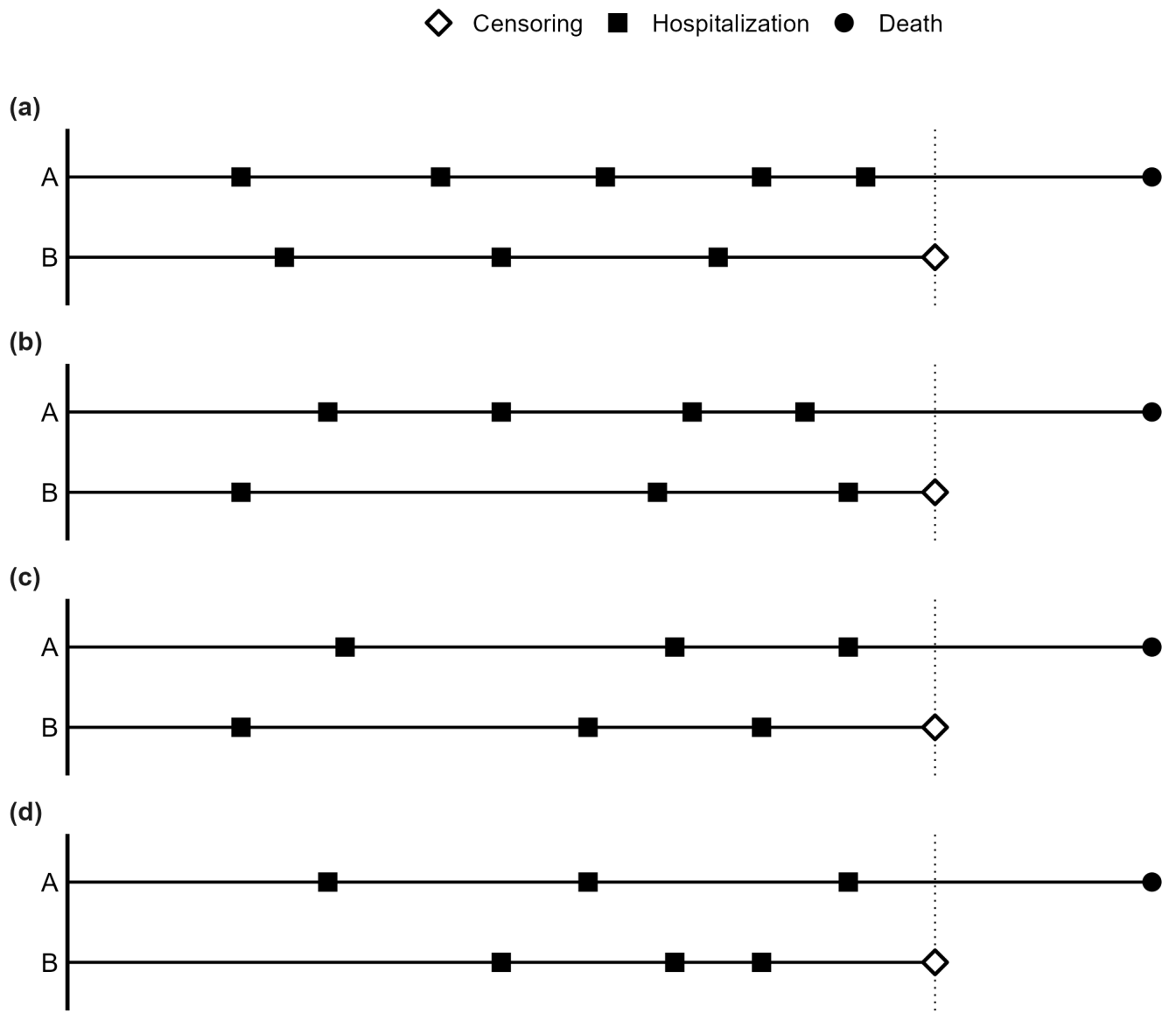
Null vs Alternative Hypotheses

- **Null hypothesis**: No treatment effect on any event
  - WR $= 1$, NB $= 0$, WO $= 1$ regardless of follow-up time
- **Alternative hypothesis**
  - Death and hospitalization times: *stochastically larger* in treatment than in control (Luo et al., 2015; Mao, 2019)
    - Stratified test: Stochastical order in each stratum
  - So WR > 1, NB > 0, WO > 1 regardless of follow-up time
  - Power $\to 1$ as $n = n_1 + n_0 \to \infty$

Recurrent Events (I)

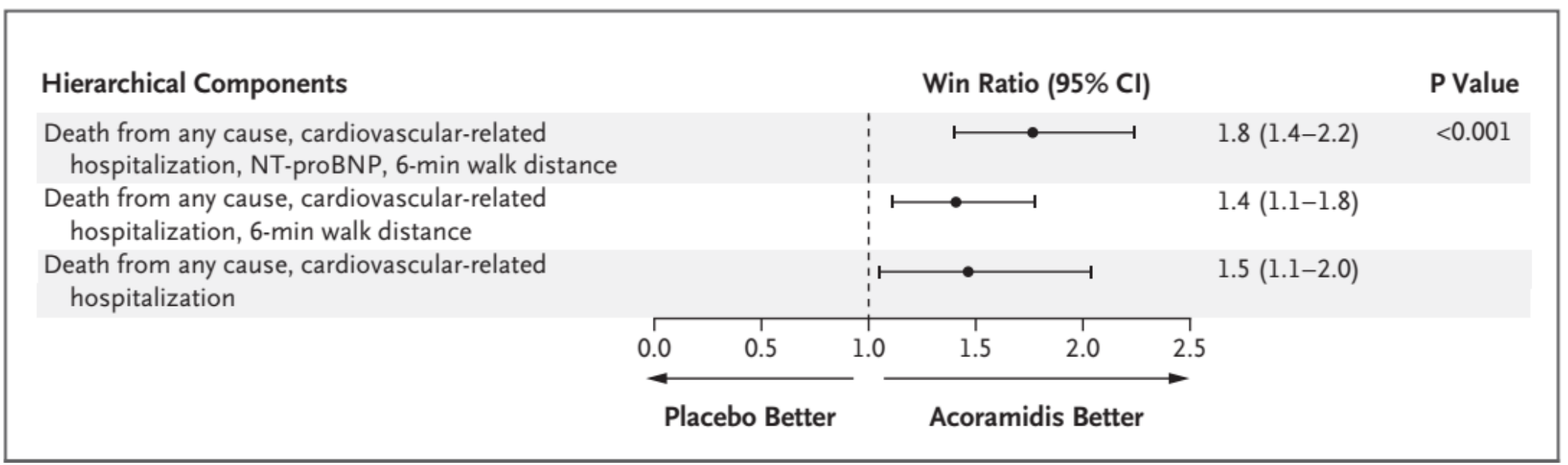
- **GPC to accommodate recurrent events**
  - E.g., hospitalizations, exacerbations, infections, etc.
- **Comparison rules**
  - Death $\rightarrow$ Number of events $\rightarrow$ Time to first/last event (Mao, Kim, & Li, 2022) $\rightarrow$ … (Biomarker/QoL changes)
  - H_A: Treatment stochastically delays all events

Recurrent Events (II)



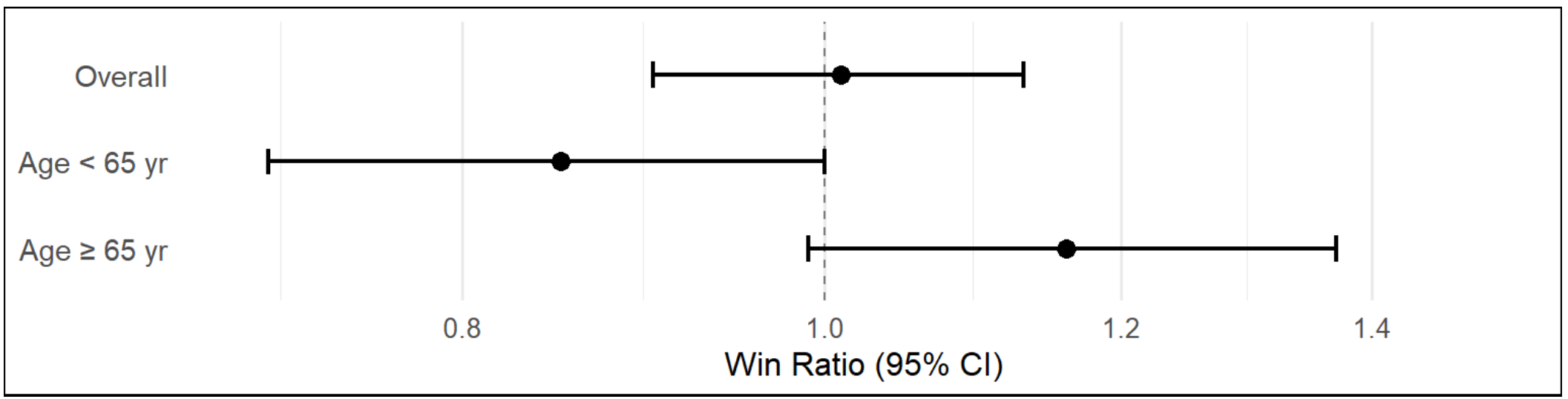
ATTRibute-CM (NEJM 2024)

- **Trial data** (Gillmore et al., 2024)
  - **Design:** Phase 3, double-blind RCT (n = 632; 2:1 randomization)
  - **Population:** Transthyretin amyloid cardiomyopathy (90 % wild-type, mean age 77 yr)
  - **Intervention:** Acoramidis 800 mg BID vs placebo for 30 months
  - **Primary HCE:** Death → CV hospitalization → NT-proBNP → 6-min walk distance
    - Stratified by TTR genotype, NT-proBNP level, eGFR, etc.



DEFENDER (JAMA 2024)

- **Trial data** (Tavares et al., 2024)
  - **Design:** Multicenter, open-label RCT (n = 507; 22 ICUs in Brazil)
  - **Population:** Critically ill adults with an organ dysfunction (respiratory/ cardiovascular/ renal)
  - **Intervention:** Dapagliflozin 10 mg daily vs standard care alone (up to 14 days)
  - **Primary HCE:** Hospital mortality → initiation of KRT → ICU length of stay (28 days)



WR as Primary Endpoint (I)

Gregson et al. (2025) summarizes recent/ongoing RCTs:

Trial	Condition	Interventions	HCE
ATTRibute-CM (2024)	Transthyretin amyloid cardiomyopathy	Acoramidis vs placebo	Death → CV hosp → NT-proBNP/6MWT
COVID-PACT (2022)	Critically ill COVID-19	Full-dose AC vs standard-dose AC	Venous/arterial death → PE → DVT → …
DAPA-MI (2024)	Clinically stable acute MI	Dapagliflozin vs placebo	CV death → non-CV death → HF hosp → …
DEFENDER (2024)	Critical illness with organ failure	Dapagliflozin + SOC vs SOC	In-hospital death → KRT initiation → ICU LOS
EMPULSE (2022)	Acute HF (hospitalized)	Empagliflozin vs placebo	All-cause death → HF events (number/time) → KCCQ change
HEART-FID (2023)	HFrEF with iron deficiency	Ferric carboxymaltose vs placebo	All-cause death → HF hosp → 6MWT change

WR as Primary Endpoint (II)

Trial	Condition	Interventions	HCE
PARTNER 3 (2023)	Low-risk symptomatic AS	TAVR vs SAVR	Death → stroke → rehospitalization days
REDUCE LAP-HF II (2022)	HFpEF/HFmrEF	Atrial shunt device vs sham	CV death/stroke → HF hosp → KCCQ change
TRILUMINATE (2023)	Symptomatic tricuspid regurgitation	TEER vs OMT	Death/surgery → HF hosp → KCCQ change
VIP-ACS (2022)	Acute coronary syndrome	High-dose influenza vaccine vs standard	Death → MI → stroke → …

…

Extensions and Variations

Sample Size Calculation

- **Rank-based approximation** (Yu & Ganju, 2022)

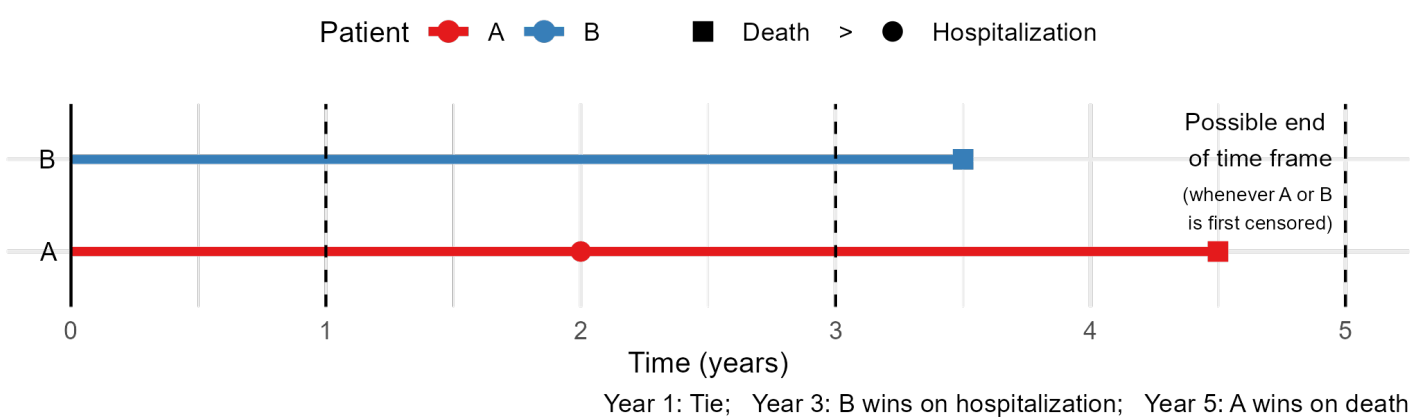
$$n \approx \frac{4(1+p_{\text{tie}})(z_{1-\alpha/2}+z_{1-\beta})^2}{3q(1-q)(1-p_{\text{tie}})\log^2(\text{WR})}$$

  - $\alpha = 0.05$: type I error; $1-\beta = 80\%, 90\%$: desired power
  - $q = n_1/n$; p_{tie}: proportion of ties
  - Elicit effect size $\log(\text{WR})$ and p_{tie} from component-wise win-loss measures (Barnhart et al., 2025a, 2025b)

- **Parametric approach** (Mao, Kim, & Miao, 2022)
  - Gumbel–Hougaard copula + Numerical integration (R package `WR`)

Estimand Issue

- **Without censoring**
  - WR, NB, WO = functions of joint distribution of events
  - Change with follow-up time (Mao, 2024; Oakes, 2016)



- **With censoring**
  - Weighted average of win-loss probabilities over time
  - Trial-dependent estimand (ICH, 2020)

Time Restriction

- **Specify time horizon τ**
  - Imagine all subjects followed up to τ (1 yr, 3 yrs, etc.)
- **GPC under consistent time frame**
  - **Estimand**: WR(τ), NB(τ), WO(τ)
  - **Interpretation**: win ratio/net benefit/win odds up to time τ
- **Estimation with censored data**
  - Inverse probability censoring weighting (Dong et al., 2020b, 2021)
    - **R-package**: `WINS`
  - Multiple imputations (T. Wang et al., 2023, 2024)

Regression Analysis

- **A proportional win-fractions (PW) model**

$$\frac{\mathrm{pr}(W_{ij}(t) = 1 \mid Z_i, Z_j)}{\mathrm{pr}(W_{ij}(t) = -1 \mid Z_i, Z_j)} = \exp\{\beta^{\mathrm{T}}(Z_i - Z_j)\}$$

  - $W_{ij}(t)$: win-loss indicator by time t
  - Z_i, Z_j: covariate vectors for subjects i and j
  - β: log-WRs for one-unit increases in covariates
  - U-statistic estimation; residual analysis to check WR constancy (Mao & Wang, 2021, R package `WR`); stratification (T. Wang & Mao, 2022)
  - Elastic net regularization for high-dimensional Z (Mao, 2025, R tool `WRNet`)

Win-Loss Times

- **Restricted mean time** in favor of treatment

$$\mu(\tau) = w_{1,0}(\tau) - w_{0,1}(\tau)$$

  - $w_{a,1-a}(\tau)$: average *time* subject in group a wins over subject in group $1-a$ up to time τ
  - Interpretation: net favorable time gained/lost by treatment (Mao, 2023, R package `rmt`)

- **Other win time measures**
  - Expected win time ($\tau \to$ maximum length of follow-up), etc. (Troendle et al., 2024)

DOOR

- **Desirability of outcome ranking** (Evans et al., 2015)
  - Pre-defined ordinal categories of clinical outcome $Y^{(a)}$
    1. Dead
    2. Alive with severe AE
    3. Alive with mild AE
    4. Alive without AE
- **Estimand**
  - Tie-adjusted *win probability* from GPC ranked by DOOR

$$\theta_{\text{DOOR}} = \Pr(Y^{(1)} > Y^{(0)}) + \tfrac{1}{2}\Pr(Y^{(1)} = Y^{(0)})$$

  - R package `door`

Challenges and Open Questions

Critiques

- **Some caveats** (Butler et al., 2024)
  - WR ignores ties
    - Win relative to *loss* rather than *"not win"*
    - Some consider this to overstate effect size
  - Inclusion of low-priority biomarker
    - Surrogate measures drives treatment effect
  - Subjectivity in defining HCE
    - Clinical input needed
    - Post hoc construction inflates type I error

Covariate Adjustment

- **Different from regression**
  - **Marginal estimands** between groups, not conditioning on Z
  - **Gain efficiency** when outcome-covariate model is true, otherwise still valid (robustness)
  - **Standard endpoints**
    - Continuous, binary, univariate survival, etc. (Tsiatis et al., 2008; B. Wang et al., 2021; Ye et al., 2023)
    - FDA recommendation (FDA, 2023)
- **Challenges with WR**
  - U-statistic structure
  - Lack of likelihood structure

Interim Analysis

- **Purpose**
  - Analyze interim data for evidence of efficacy/futility $\rightarrow$ stop trial early
  - Univariate survival: information accrued $\propto$ number of events
- **Challenges with WR**
  - Information time not event-driven
    - Case study: RMST (Lu & Tian, 2021; Luo et al., 2019)
  - Correlations between component events
    - *Dependent* increments

Meta Analysis

- **Challenges**
  - Primary studies *not* reporting win-loss measures
  - Primary studies over *different* time spans
  - Primary studies with *different* definitions of win/loss
- `WinKM`: A toolkit to start
  - Calculate win-loss statistics based on
    - KM estimates for OS and EFS
    - At-risk table at selected time points
    - Total event counts (reported in the CONSORT diagram or results section)

Conclusion

Key Takeaways (I)

- **Win ratio and its variants**
  - Intuitive, clinically interpretable effect sizes for HCEs
  - Support recurrent events, stratification, etc.
  - Time-resticted estimands, regression analysis, win times, DOOR, etc.
- **Good practice**
  - Pre-specify HCE definition and analysis plan
  - Caution in including surrogate measures
    - Anticipated clinical event rates
    - Set thresholds for meaningful differences
  - Sensitivity analyses

Key Takeaways (II)

- **R packages**
  - `WR`, `WINS`, `BuyseTest`, `WinRatio`, `rmt`, `WRNet`, `door`, `WinKM`, etc.
- **Open questions**
  - Covariate adjustment, interim analysis, meta analysis, etc.

Acknowledgments

- **Funding**
  - R01HL149875 (11/2019 – 7/2028)
  - DMS2015526 (7/2020 – 6/2024)

- **Collaborators**
  - KyungMann Kim, Tuo Wang, Gaohong Dong, Bo Huang, etc.

References

Barnhart, H., Lokhnygina, Y., Matsouaka, R. A., & Rockhold, F. W. (2025b). Trial design with win statistics for multiple time-to-event endpoints with hierarchy. *Statistics in Biopharmaceutical Research*, *17*(2), 197–210.

Barnhart, H., Lokhnygina, Y., Matsouaka, R., Halabi, S., Yanez, D., Mentz, R. J., & Rockhold, F. (2025a). Sample size and power calculations with win measures based on hierarchical endpoints. *Statistics in Medicine*, *44*(10-12), e70096.

Bebu, I., & Lachin, J. M. (2016). Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics*, *17*(1), 178–187. https://doi.org/10.1093/biostatistics/kxv032

Brunner, E., Vandemeulebroecke, M., & Mütze, T. (2021). Win odds: An adaptation of the win ratio to include ties. *Statistics in Medicine*, *40*(14), 3367–3384. https://doi.org/10.1002/sim.8967

Butler, J., Stockbridge, N., & Packer, M. (2024). Win ratio: A seductive but potentially misleading method for evaluating evidence from clinical trials. *Circulation*, *149*(20), 1546–1548.

Buyse, M. E. (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine*, *29*(30), 3245–3257. https://doi.org/10.1002/sim.3923

Buyse, M. E., Verbeeck, J., De Backer, M., Deltuvaite-Thomas, V., Saad, E. D., & Molenberghs, G. (2025). *Handbook of generalized pairwise comparisons: Methods for patient-centric analysis*. CRC Press, Taylor & Francis Group.

Dong, G., Hoaglin, D. C., Huang, B., Cui, Y., Wang, D., Cheng, Y., & Gamalo-Siebers, M. (2023). The stratified win statistics (win ratio, win odds, and net benefit). *Pharmaceutical Statistics*,

22(4), 748–756. https://doi.org/10.1002/pst.2293
Dong, G., Huang, B., Wang, D., Verbeeck, J., Wang, J., & Hoaglin, D. C. (2021). Adjusting win statistics for dependent censoring. *Pharmaceutical Statistics*, *20*(3), 440–450. https://doi.org/10.1002/pst.2086
Dong, G., Li, D., Ballerstedt, S., & Vandemeulebroecke, M. (2016). A generalized analytic solution to the win ratio to analyze a composite endpoint considering the clinical importance order among components. *Pharmaceutical Statistics*, *15*(5), 430–437. https://doi.org/10.1002/pst.1763
Dong, G., Mao, L., Huang, B., Gamalo-Siebers, M., Wang, J., Yu, G., & Hoaglin, D. C. (2020b). The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of independent censoring. *Journal of Biopharmaceutical Statistics*, *30*(5), 882–899. https://doi.org/10.1080/10543406.2020.1757692
Dong, G., Qiu, J., Wang, D., & Vandemeulebroecke, M. (2018). The stratified win ratio. *Journal of Biopharmaceutical Statistics*, *28*(4), 778–796. https://doi.org/10.1080/10543406.2017.1397007
Evans, S. R., Rubin, D., Follmann, D., Pennello, G., Huskins, W. C., Powers, J. H., Schoenfeld, D., Chuang-Stein, C., Cosgrove, S. E., Fowler Jr, V. G., et al. (2015). Desirability of outcome ranking (DOOR) and response adjusted for duration of antibiotic risk (RADAR). *Clinical Infectious Diseases*, *61*(5), 800–806.
FDA. (2023). *Guidance document: Adjusting for covariates in randomized clinical trials for drugs and biological products*. US Food and Drug Adminstration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products
Gillmore, J. D., Judge, D. P., Cappelli, F., Fontana, M., Garcia-Pavia, P., Gibbs, S., Grogan, M., Hanna, M., Hoffman, J., Masri, A., Maurer, M. S., Nativi-Nicolau, J., Obici, L., Poulsen, S. H., Rockhold, F., Shah, K. B., Soman, P., Garg, J., Chiswell, K., … Fox, J. C. (2024). Efficacy and

safety of acoramidis in transthyretin amyloid cardiomyopathy. *New England Journal of Medicine*, *390*(2), 132–142. https://doi.org/10.1056/NEJMoa2305434
Gregson, J., Taylor, D., Owen, R., Collier, T., J. Cohen, D., & Pocock, S. (2025). Hierarchical composite outcomes and win ratio methods in cardiovascular trials: A review and consequent guidance. *Circulation*, *151*(22), 1606–1619.
ICH. (2020). *ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials*, *step 5*. London: European Medicines Evaluation Agency.
Lu, Y., & Tian, L. (2021). Statistical Considerations for Sequential Analysis of the Restricted Mean Survival Time for Randomized Clinical Trials. *Statistics in Biopharmaceutical Research*, *13*(2), 210–218. https://doi.org/10.1080/19466315.2020.1816491
Luo, X., Huang, B., & Quan, H. (2019). Design and monitoring of survival trials based on restricted mean survival times. *Clinical Trials*, *16*(6), 616–625. https://doi.org/10.1177/1740774519871447
Luo, X., Tian, H., Mohanty, S., & Tsai, W. Y. (2015). An Alternative Approach to Confidence Interval Estimation for the Win Ratio Statistic. *Biometrics*, *71*(1), 139–145. https://doi.org/10.1111/biom.12225
Mao, L. (2019). On the Alternative Hypotheses for the Win Ratio. *Biometrics*, *75*(1), 347–351. https://doi.org/10.1111/biom.12954
Mao, L. (2023). On restricted mean time in favor of treatment. *Biometrics*, *79*(1), 61–72. https://doi.org/10.1111/biom.13570
Mao, L. (2024). Defining estimand for the win ratio: Separate the true effect from censoring. *Clinical Trials*, *21*(5), 584–594.
Mao, L. (2025). Regularized win ratio regression for variable selection and risk prediction, with an application to a cardiovascular trial. *BMC Medical Research Methodology*, *25*(1). https://doi.org/10.1186/s12874-025-02554-w
Mao, L., Kim, K., & Li, Y. (2022). On recurrent-event win ratio. *Statistical Methods in Medical*

Research, *31*(6), 1120–1134. https://doi.org/10.1177/0962280221084134
Mao, L., Kim, K., & Miao, X. (2022). Sample size formula for general win ratio analysis. *Biometrics*, *78*(3), 1257–1268. https://doi.org/10.1111/biom.13501
Mao, L., & Wang, T. (2021). A class of proportional win-fractions regression models for composite outcomes. *Biometrics*, *77*(4), 1265–1275. https://doi.org/10.1111/biom.13382
Oakes, D. (2016). On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika*, *103*(3), 742–745. https://doi.org/10.1093/biomet/asw026
Pocock, S. J., Ariti, C. A., Collier, T. J., & Wang, D. (2012). The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal*, *33*(2), 176–182. https://doi.org/10.1093/eurheartj/ehr352
Tavares, C. A. M., Azevedo, L. C. P., Rea-Neto, Á., Campos, N. S., Amendola, C. P., Kozesinski-Nakatani, A. C., David-João, P. G., Lobo, S. M., Filiponi, T. C., Almeida, G. M. B., Bergo, R. R., Guimarães-Júnior, M. R. R., Figueiredo, R. C., Castro, J. R., Schuler, C. J., Westphal, G. A., Carioca, A. C. R., Monfradini, F., Nieri, J., … Zampieri, F. G. (2024). Dapagliflozin for critically ill patients with acute organ dysfunction: The DEFENDER randomized clinical trial. *JAMA*, *332*(5), 401–411. https://doi.org/10.1001/jama.2024.10510
Troendle, J. F., Leifer, E. S., Yang, S., Jeffries, N., Kim, D.-Y., Joo, J., & O'Connor, C. M. (2024). Use of win time for ordered composite endpoints in clinical trials. *Statistics in Medicine*, *43*(10), 1920–1932. https://doi.org/10.1002/sim.10045
Tsiatis, A. A., Davidian, M., Zhang, M., & Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*, *27*(23), 4658–4677. https://doi.org/10.1002/sim.3113
Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M., & Rosenblum, M. (2021). Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment. *Journal of the American Statistical Association*, *118*(542), 1152–1163. https://doi.org/10.1080/01621459.2021.1981338
Wang, T., Li, Y., & Qu, Y. (2024). Restricted time win ratio: from estimands to estimation. *Statistics in Biopharmaceutical Research*, 1–18. https://