Non- and semi-parametric approaches to estimand construction for composite endpoints

Penn Conference on Statistical Issues in Clinical Trials: Advances in Time to Event Analyses

4-17-2023

Lu Mao

Associate Professor of Biostatistics

University of Wisconsin-Madison

School of Medicine and Public Health
UNIVERSITY OF WISCONSIN–MADISON

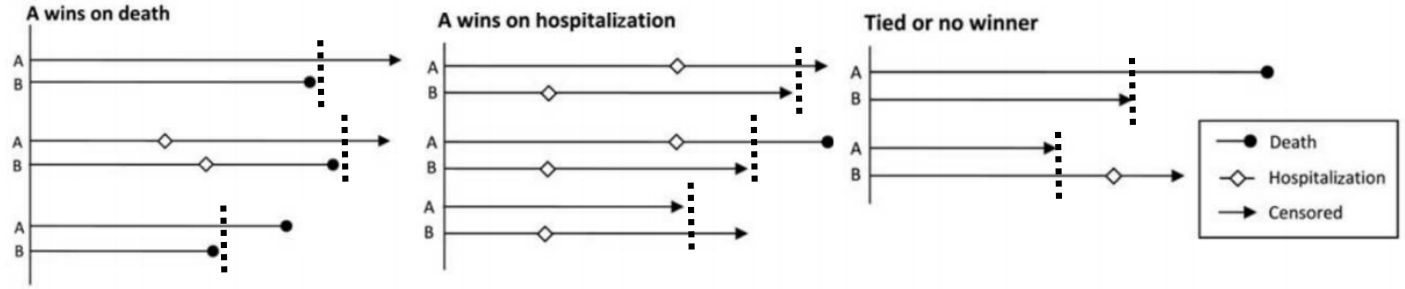
Introduction – Composite Endpoints

- **Composite endpoints**: those that combine mortality with nonfatal events like cardiovascular (CV) hospitalization and tumor progression
  - Traditional: time to first event
  - Limitations:
    - Statistical efficiency
    - Death vs nonfatal events
- **General pairwise comparisons (GPC)**: compare every patient in the treatment with every one in the control
  - Involve more events
  - Flexible ranking of event types (e.g., death > hospitalization)

Introduction – GPC

- **General framework**: for each pair, determine a *winner*, *loser*, or tie
  - E.g., compare the two at the earlier of their death/censoring times

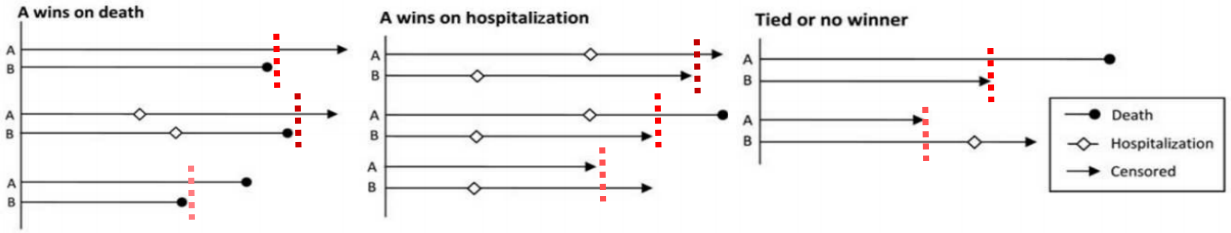
Death > hospitalization



  - $\widehat{w}_1$: proportion of pairs where treated wins; $\widehat{w}_0$: proportion of pairs where untreated wins
  - <u>GPC statistics</u>
    - *Win ratio* (WR): $\widehat{w}_1/\widehat{w}_0$ (Pocock et al., 2012)
    - *Proportion in favor* (PIF) of treatment (or *net benefit*): $\widehat{w}_1 - \widehat{w}_0$ (Buyse, 2010)
    - *Win odds* (WO): $(\widehat{w}_1 + 0.5\hat{O})/(\widehat{w}_0 + 0.5\hat{O})$ (Dong et al., 2020a), where $\hat{O} = 1 - \widehat{w}_1 - \widehat{w}_0$

Introduction – GPC

- **Limitation**: the *estimands* of win/loss proportions depend on censoring distribution



  - $\widehat{w}_a$ *mixes* comparisons made at *different* times (Luo et al., 2015; Bebu & Lachin, 2016; Oakes, 2016)
    - $\widehat{w}_a \rightarrow w_a$ $(a = 1, 0)$
    - Heavy censoring → shorter follow-up → less events → $w_a \downarrow$
    - In fact (Oakes, 2016)

$$w_a = \int_0^\infty w_a(t)\, \mathrm{d}G(t)$$

      - $w_a(t) = \mathrm{pr}(\text{Group } a \text{ wins against group } 1 - a \text{ by time } t)$
      - $G(t)$: Distribution of the *minimum* of the two group-specific *censoring times*

Introduction – Testing or Estimation?

- **Hypothesis testing** (qualitative): test

$$H_0\colon\ w_1(t) \equiv w_0(t), \qquad \forall\ t \geq 0$$

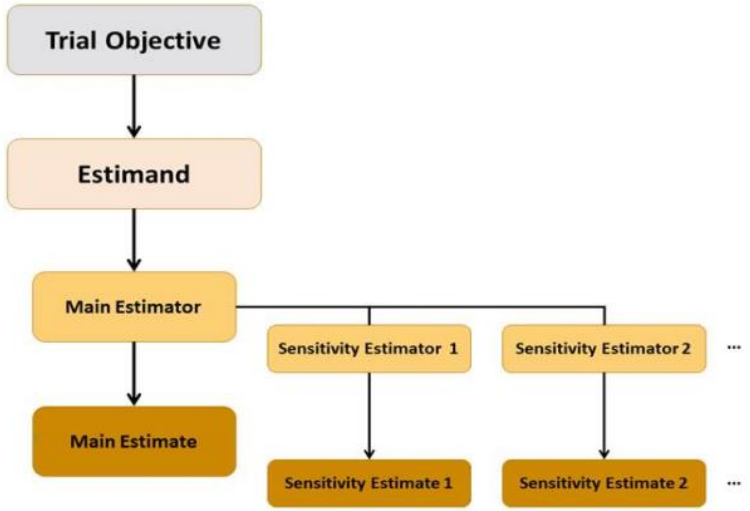
against

$$H_A\colon\ w_1(t) \geq w_0(t),\ \forall\ t \geq 0 \text{ with strict inequality for some } t$$

  - GPC statistics generally yield valid tests
    - E.g., Reject H_0 if $\log(\widehat{w}_1/\widehat{w}_0) > \hat{c}_\alpha$ (determined by variance of log-WR and type I error α)
    - As sample size increases, $\text{pr}\{\log(\widehat{w}_1/\widehat{w}_0) > \hat{c}_\alpha\} \overset{H_A}{\to} 1$ (consistency)
- **Estimation** (quantitative): how much is the treatment better than control?
  - WR, PIF, and WO are functions of censoring distribution
  - *Generalization* to target population questionable (Luo and Quan, 2020)

Introduction – ICH-E9(R1) Addendum

- **ICH-E9 (R1)**: "Estimands and Sensitivity Analysis in Clinical Trials" (ICH, 2020)
  - "A central question for drug development and licensing is to quantify treatment effects."
  - Define effect-size estimands that are meaningful and generalizable (Akacha et al., 2017a; Akacha et al., 2017b; Akacha et al., 2021; McCaw et al., 2021; Ionan et al., 2022)
    - Remove the influence of censoring
  - The guidelines have been adopted by European Medicines Agency (EMA) and FDA (Lynggaard et al., 2022)



Estimand Construction

- **General strategy**
  - Full (*uncensored*) outcome on a patient from group a : $\mathcal{H}^{(a)}(\infty) \quad (a = 1, 0)$
    - $\mathcal{H}^{(a)}(t) = \left\{N_D^{(a)}(u), N_1^{(a)}(u), \ldots, N_K^{(a)}(u): 0 \le u \le t\right\}$
    - $N_D^{(a)}(u) = I\left(D^{(a)} \le u\right)$; $D^{(a)}$ = Overall survival time
    - $N_k^{(a)}(u)$ = Counting process for kth type of (possibly recurrent) nonfatal event $(k = 1, \ldots, K)$
  - Win/loss process

$$w_a(t) = \text{pr}\left(\mathcal{H}^{(a)}(t) \text{ wins against } \mathcal{H}^{(1-a)}(t)\right)$$

  - E.g., $w_a(t) = \text{pr}\Big(\underbrace{D^{(1-a)} < D^{(a)} \wedge t}_{\text{Win on death}} \text{ or } \underbrace{\left\{D^{(1)} \wedge D^{(0)} > t, T^{(1-a)} < T^{(a)} \wedge t\right\}}_{\text{Win on first nonfatal event}}\Big)$,

($b \wedge c = \min(b, c)$, $T^{(a)}$: time to first nonfatal event)

Estimand Construction – Two Approaches

- **Nonparametric**: Specify a time horizon τ (e.g., 5 years) (Oakes, 2016; Finkelstein & Schoenfeld, 2019)
  - Restricted WR: $w_1(\tau)/w_0(\tau)$
  - Restricted PIF: $w_1(\tau) - w_0(\tau)$
  - Restricted WO: $\{w_1(\tau) + 0.5O(\tau)\}/\{w_0(\tau) + 0.5O(\tau)\}$, where $O(\tau) = 1 - w_1(\tau) - w_0(\tau)$
- **Semiparametric**: Impose a temporal model on relationship between $w_1(t)$ and $w_0(t)$
  - Proportional win-fractions model: $\frac{w_1(t)}{w_0(t)} \equiv \theta$ (time-invariant win ratio) (Mao & Wang, 2021)
- **Estimation** with *censored* data: $\{\mathcal{H}^{(a)}(X^{(a)}), X^{(a)}\}$
  - $X^{(a)} = D^{(a)} \wedge C^{(a)}$, where $C^{(a)}$ is (independent) censoring time

Estimand Construction – Nonparametric

- **Goal**: estimating $w_a(\tau)$ using $\left\{\mathcal{H}_i^{(a)}\left(X_i^{(a)}\right), X_i^{(a)}\right\}$ $(i = 1, \ldots, n_a)$
  - No censoring before τ:

$$\widehat{w}_a(\tau) = (n_1 n_0)^{-1} \sum_{i=1}^{n_a} \sum_{j=1}^{n_{1-a}} I\left\{\mathcal{H}_i^{(a)}(\tau) \text{ wins against } \mathcal{H}_j^{(1-a)}(\tau)\right\}$$

  - In general, inverse probability censoring weighting (IPCW; Dong et al., 2020b)
    - Weight the kernel by, e.g., $\frac{I\left(C_i^{(a)} \geq D_i^{(a)} \wedge \tau, C_j^{(1-a)} \geq D_j^{1-a} \wedge \tau\right)}{G_a\left(D_i^{(a)} \wedge \tau\right) G_{1-a}\left(D_j^{(1-a)} \wedge \tau\right)}$ to correct for censoring bias, where $G_a(t) = \text{pr}\left(C^{(a)} \geq t\right)$, or $\text{pr}\left(C^{(a)} \geq t \mid Z\right)$ if censoring depends on covariates Z (Dong et al., 2021)
    - R-package: `WINS` (Cui and Huang, 2022; CRAN: https://CRAN.R-project.org/package=WINS)

Estimand Construction – Nonparametric

- **A variation**: restricted mean time in favor (RMT-IF) (Mao, 2023, *Biometrics*)

$$\mu(\tau) = w_1(\tau) - w_0(\tau)$$

  - $w_a(\tau) = E\left(\text{Time } \mathcal{H}^{(a)}(\cdot) \text{ is better than } \mathcal{H}^{(1-a)}(\cdot) \text{ over } [0, \tau]\right)$
  - Re-expressed in terms of survival functions of component events
    - Plug-in Kaplan—Meier estimator, avoid IPCW
  - R-package: https://CRAN.R-project.org/package=rmt
  - Example: levamisole+fluorouracil versus control in a colon caner trial (Moertel, et al., 1990)

	$\tau = 2.5$ years			$\tau = 5.0$ years			$\tau = 7.5$ years		
	Est	**SE**	***p*-value**	**Est**	**SE**	***p*-value**	**Est**	**SE**	***p*-value**
Pre-relapse	2.09	0.44	<0.001	3.41	0.70	<0.001	4.15	0.86	<0.001
Survival	0.56	0.57	0.321	3.64	1.53	0.018	7.44	2.56	0.004
Overall	2.65	0.83	0.001	7.05	1.93	<0.001	11.59	3.03	<0.001

Note Est, estimate; SE, standard error.

Estimand Construction – Semiparametric

- **Proportional win-fractions (PW) model**
  (Mao & Wang, 2021, *Biometrics*)
  - $\frac{w_1(t)}{w_0(t)} \equiv \theta$ (time-invariant WR) for all $t \geq 0$ (can include covariates Z)
    - For Pocock's rule of pairwise comparison, satisfied under a Lehmann model (Oakes, 2016)
    - Treatment θ times as likely to win as compared to control (regardless of the restricting time)
  - No IPCW is needed as WR is constant under proportionality
  - R-package: https://CRAN.R-project.org/package=WR

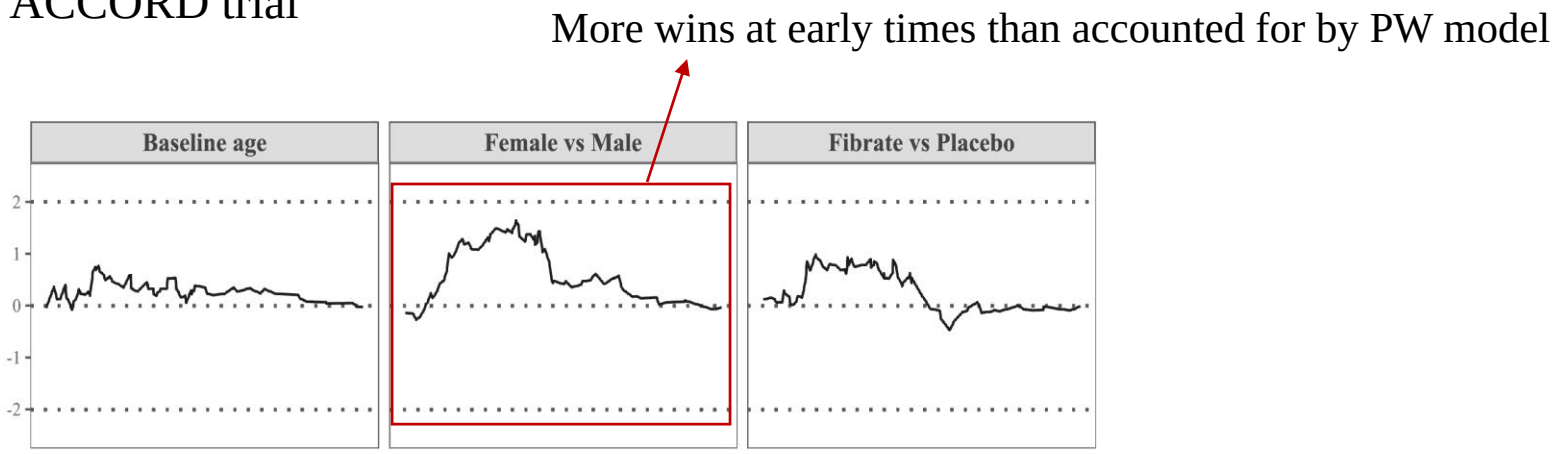
Example: PW regression analysis of HF-ACTION Trial.

Covariate	Win Ratio	95% CI	*P*-value
Training vs Usual	1.06	(0.95-1.19)	.275
Non-Ischemic vs Ischemic	1.15	(1.02-1.31)	.027
Age (decade)	1.02	(0.97-1.07)	.468
Male vs Female	0.72	(0.63-0.82)	<.001
CPX Duration (minute)	1.11	(1.09-1.13)	<.001
Canada vs USA	1.34	(1.09-1.66)	.007
France vs USA	1.95	(1.32-2.89)	.001
Atrial Fibrillation (y vs n)	0.80	(0.70-0.92)	.002
Diabetes (y vs n)	0.98	(0.87-1.11)	.726

Note: CI, confidence interval.

Estimand Construction – Semiparametric

- **Checking proportionality**
  - Plot residuals (observed vs model-based win-fractions) over time
    - Should see *no* systematic trend if proportionality is satisfied
    - Nonproportionality → Estimand becomes a censoring mix of time-dependent WRs
    - Example: ACCORD trial



    - Stratify on nonproportional covariates (Wang & Mao, 2022, *Statistics in Medicine*)

Estimand Construction – Semiparametric

- Similar **global models** for PIF and WO?
  - Because $w_a(0) = 0$, difficult/impossible for
    - $w_1(t) - w_0(t)$ or
    - $\{w_1(t) + 0.5O(t)\}/\{w_0(t) + 0.5O(t)\}$

    to be constant over t (unless under the null)
  - How to realistically constrain PIF/WO over time?
- Alternative: **local models** for a restricting time τ
  - E.g., $w_1(\tau \mid Z) - w_0(\tau \mid Z) = g\left(\theta^{\mathrm{T}} Z\right)$
  - IPCW (Dong et al., 2020b; 2021)? Pseudo-observation (Andersen & Pohar Perme, 2010)?

Informative Censoring

- Both approaches assume independent censoring
  - Random loss of follow-up
  - Study termination with (unselective) staggered entry
- Two types of dependent (informative) censoring
  - Dropout influenced by factors (e.g., baseline covariate) unaccounted for in the model
    - Target: a population where such dropout did not occur
    - Solution: covariate-adjusted IPCW (Dong et al., 2021)
  - "Intercurrent events" (ICH, 2020): treatment discontinuation, death from related causes
    - Target: a population where such events do occur
    - Solution: "Composite strategy" (e.g., death > treatment-discont. > minor symptoms); "While-on-treatment strategy" (adjusting for the time patient is on treatment)

Open Problems

- **Efficiency of IPCW**
  - Utilize as many “complete cases” as possible
  - Complete case ← win/loss determinate at τ
  - Depends on outcome types and rule of comparison (Dong et al., 2020b)
  - Augmentation using baseline and interim data (Tsiatis et al., 2008)?
- **Regression of (local) win/loss estimands**
  - A general formulation

$$h\{w_1(\tau \mid Z), w_0(\tau \mid Z)\} = \theta^{\mathrm{T}} Z$$

    - $h(\cdot,\cdot)$ is some link function
    - IPCW? Pseudo-observations?

Summary

- **GPC** is useful in full utilization and ranking of outcomes
- **ICH-E9 (R1) Addendum** → clearly specify the estimand
  - Time-dependent win/loss fractions: $w_a(t)$ $(a = 1, 0)$
  - Comparing an *uncensored* observation from treatment to one from control
- **Two approaches**
  - Nonparametric: e.g., $w_1(\tau)/w_0(\tau)$ (IPCW)
  - Semiparametric: e.g., $w_1(t)/w_0(t) \equiv \theta$ for all t (model checking)
- **Future work**
  - Improve the efficiency of IPCW
  - More flexible regression methods

Acknowledgments

- **Funding source**

  - NIH-NHLBI grant **R01HL149875** (12/01/2019—11/30/2022)

    Novel Statistical Methods for Complex Time-to-Event Outcomes in Cardiovascular Clinical Trials

- **Data source**

  - HF-ACTION and ACCORD study data are provided by ACCORD study data are provided by BioLINCC of NHLBI

- **Collaborators & Students**

  - KyungMann Kim, PhD
  - Tuo Wang

Bibliography

- ACCORD Study Group (2010) Effects of intensive blood-pressure control in type 2 diabetes mellitus. *New England Journal of Medicine*, 362, 1575–1585.
- Akacha, M., Bretz, F., Ohlssen, D., Rosenkranz, G., & Schmidli, H. (2017a). Estimands and their role in clinical trials. *Statistics in Biopharmaceutical Research*, 9, 268—271.
- Akacha, M., Bretz, F., & Ruberg, S. (2017b). Estimands in clinical trials–broadening the perspective. *Statistics in Medicine*, 36, 5—19.
- Akacha, M., Bartels, C., Bornkamp, B., Bretz, F., Coello, N., Dumortier, T., ... & Vong, C. (2021). Estimands—What they are and why they are important for pharmacometricians. CPT: *Pharmacometrics & Systems Pharmacology*, 10, 279—282.
- Andersen, P. K. & Pohar Perme, M. (2010). Pseudo-observations in survival analysis. Statistical methods in medical research, 19(1), 71-99.
- Bebu, I. and Lachin, J.M. (2015) Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics*, 17, 178–187.
- Buyse, M. (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine*, 29, 3245–3257.
- Dong, G., Huang, B., Chang, Y. W., Seifu, Y., Song, J., & Hoaglin, D. C. (2020a). The win ratio: Impact of censoring and follow-up time and use with nonproportional hazards. *Pharmaceutical Statistics*, 19, 168–177.
- 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, 882—899.

Bibliography

- Dong, G., Huang, B., Wang, D., Verbeeck, J., Wang, J., & Hoaglin, D. C. (2021). Adjusting win statistics for dependent censoring. *Pharmaceutical Statistics*, 20, 440—450.
- Finkelstein, D. M. & Schoenfeld, D. A. (2019). Graphing the win ratio and its components over time. *Statistics in Medicine*, 38, 53—61.
- 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.
- Ionan, A. C., Paterniti, M., Mehrotra, D. V., Scott, J., Ratitch, B., Collins, S., ... & Bretz, F. (2022). Clinical and Statistical Perspectives on the ICH E9 (R1) Estimand Framework Implementation. *Statistics in Biopharmaceutical Research*, accepted.
- Lynggaard, H., Bell, J., Lösch, C., Besseghir, A., Rantell, K., Schoder, V., & Lanius, V. (2022). Principles and recommendations for incorporating estimands into clinical study protocol templates. *Trials*, 23, 1—15.
- Luo, X., & Quan, H. (2020). Some meaningful weighted log-rank and weighted win loss statistics. *Statistics in Biosciences*, 12, 216—224.
- 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.
- McCaw, Z. R., Tian, L., Wei, J., Claggett, B. L., Bretz, F., Fitzmaurice, G., & Wei, L. J. (2021). Choosing clinically interpretable summary measures and robust analytic procedures for quantifying the treatment difference in comparative clinical studies. *Statistics in Medicine*, 40, 6235—6242.

Bibliography

- Mao, L. (2019). On the alternative hypotheses for the win ratio. *Biometrics*, 75, 347–351.
- Mao, L. (2022). Nonparametric inference of general while-alive estimands for recurrent event. *Biometrics*, https://doi.org/10.1111/biom.13709.
- Mao, L. (2023). On restricted mean time in favor of treatment. *Biometrics*, 79, 61–72.
- Mao, L. and Kim, K. (2021). Statistical models for composite endpoints of death and non-fatal events: a review. *Statistics in Biopharmaceutical Research*, 13, 260–269.
- Mao, L., Kim, K. and Li, Y. (2022). On recurrent-event win ratio. *Statistical Methods in Medical Research*, 31, 1120–1134.
- Mao, L., Kim, K., and Miao, X. (2022). Sample size formula for general win ratio analysis. *Biometrics*, 78, 1257–1268.
- Mao, L., and Lin, D. (2016). Semiparametric regression for the weighted composite endpoint of recurrent and terminal events. *Biostatistics*, 17, 390–403.
- Mao, L. and Wang, T. (2021). A class of proportional win-fractions regression models for composite outcomes. *Biometrics*, 77, 1265–1275.
- Moertel, C.G., Fleming, T.R., Macdonald, J.S., Haller, D.G., Laurie, J.A., Goodman, P.J. et al. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 322, 352–358.
- Oakes, D. (2016). On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika*, 103, 742–745.

Bibliography

- O’Connor, C. M., Whellan, D. J., Lee, K. L., Keteyian, S. J., Cooper, L. S., Ellis, S. J. et al. (2009) Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. *Journal of the American Medical Association*, 301, 1439–1450.
- 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, 4658–4677.
- Wang, T. and Mao, L. (2022). Stratified proportional win-fractions regression analysis. *Statistics in Medicine*, 41, 5305–5318.