Win Ratio in the Presence of Censored Data: Can Probabilistic Variants Improve Robustness?

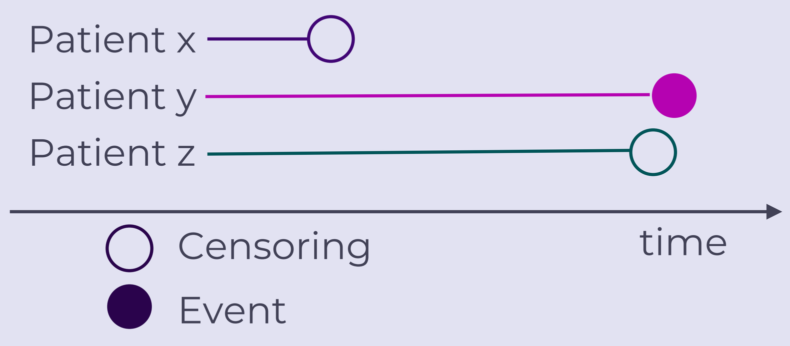
Putnam
Inizio Advisory

Mateusz Nikodem[1], Julita Janik[1,2], Michał Kochmański[1,2], Sylvaine Barbier[3], Paulina Pierzchała[1,2]
[1]Putnam, Poland, [2]AGH University of Science and Technology, Poland, [3]Putnam, France

MSR225

Background

- **Win ratio (WR) – what is it?**
  - Developed for composite hierarchical outcomes in randomized clinical trials (RCTs)
  - Each patient of arm X is compared with each one of arm Y;
  - $\text{WR(X vs Y)} = \frac{\text{Number of wins for X}}{\text{Number of wins for Y}}$
- **Case of interest**
  - Substantial censoring in the top-ranked time-to-event component
- **What's the problem?**

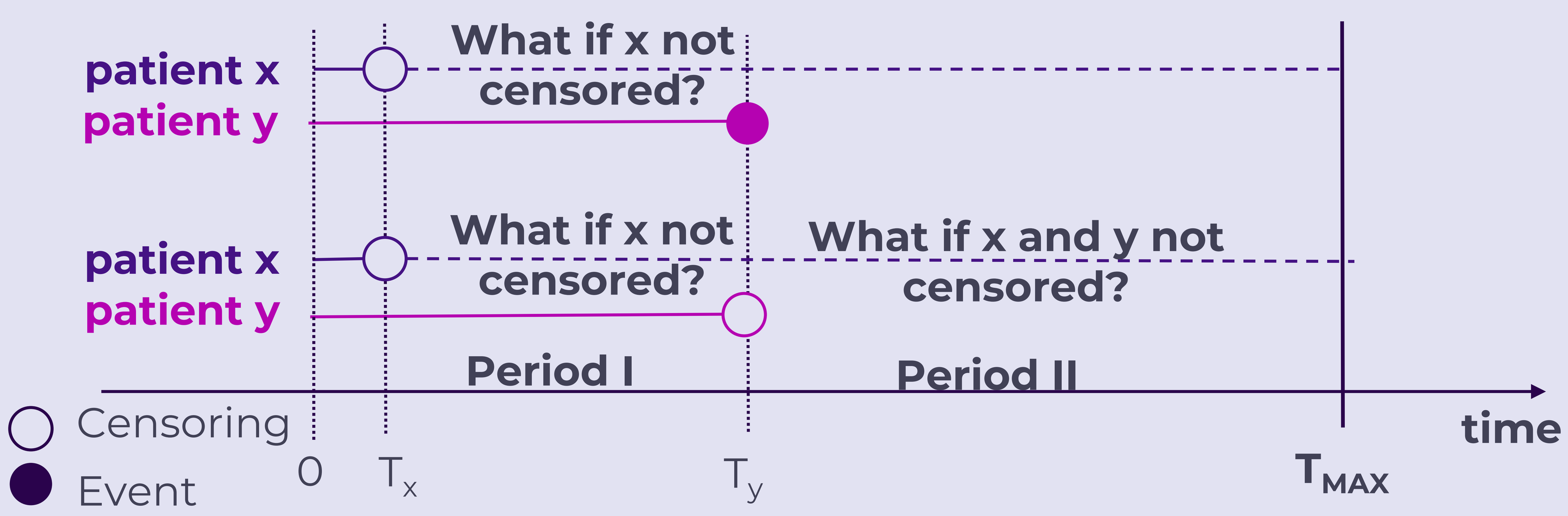


Standard WR treats all pairs as ties
...even though it is likely that:
- x performed worse than y and z
- z performed better than y

WR is sensitive to censoring, which increases bias, accentuates its inherent non-transitivity, and heightens the risk of subgroup paradoxes.

Objectives and concepts

- To highlight how censoring-sensitive the WR can be.
- To propose and evaluate the **probabilistic** alternatives for WR estimations.



Patient x censored before time of event or censoring of y (T_y)

➢ **Standard (deterministic) WR:**
- x just ties with y at this endpoint

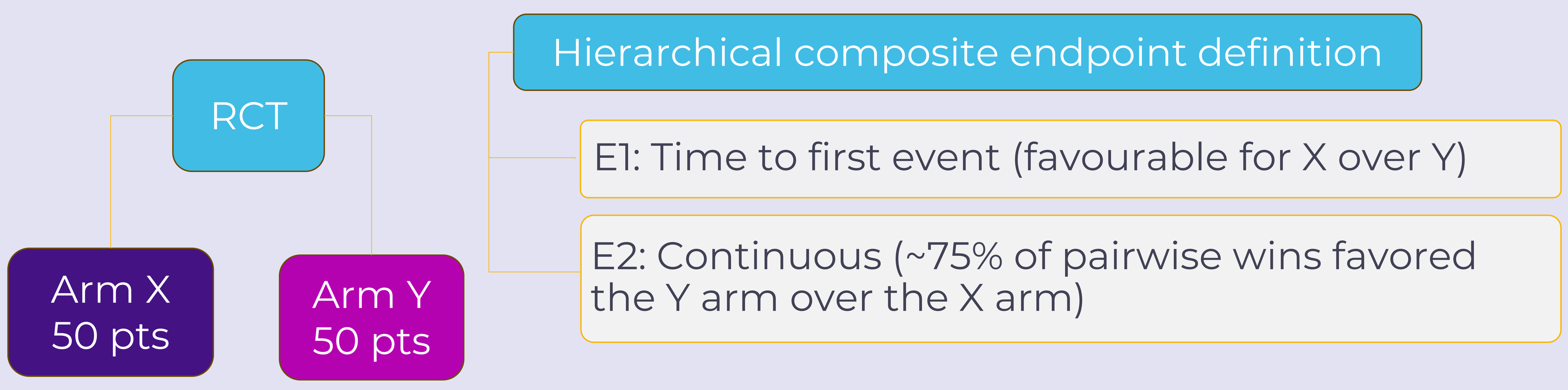
➢ **Probabilistic WR:**
- Estimate the probability that x would experience the event before T_y (i.e. $P(x<T_y)$) under the hypothetical scenario in which x was not censored prior to T_y and use this probability in the WR calculation.
- In case y censored at T_y, estimate $P(x>y\,|x>T_y)$, $P(x<y\,|x>T_y)$ and $P(x=y=T_{MAX}\,|x>T_y)$ assuming neither x nor y were censored before T_{MAX} and use these probabilities in the WR calculation

Methods

Variants of the WR estimation

- **V0**: Standard (deterministic) approach
- **V1:** Probabilistic; uses the Kaplan–Meier (K–M) curve for X in Period I, then follows the V0 approach
- **V2**: Probabilistic; uses the K–M curves for X and Y in both Period I and Period II
- **V3:** Probabilistic; uses the K–M curves for X and Y in both Period I and Period II, accounting for time-varying numbers at risk (down-weighted tails)

Case study

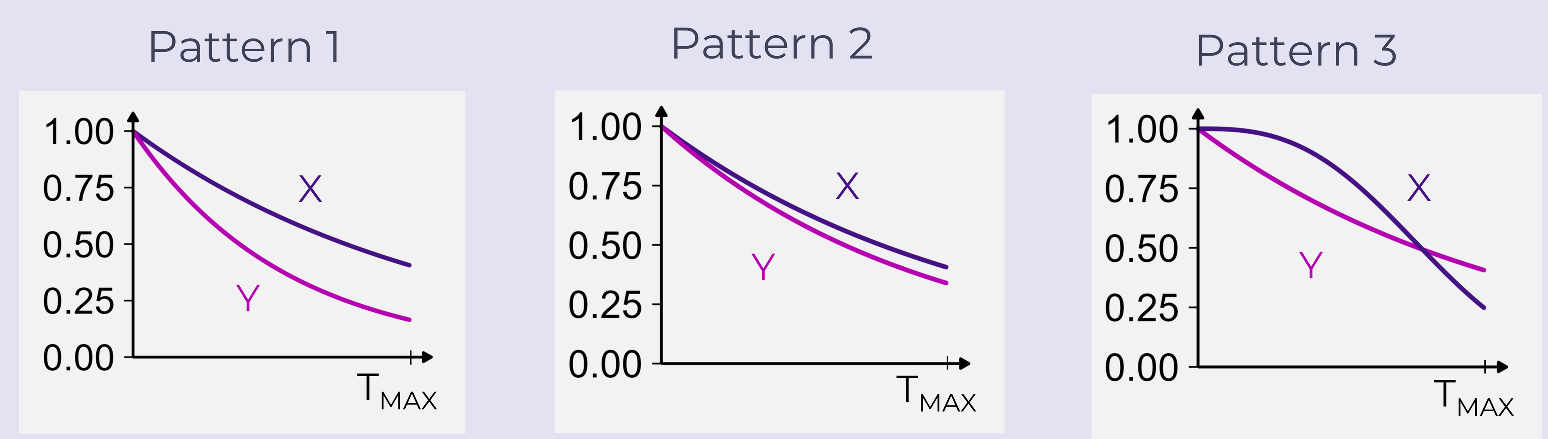


Scenarios: 9 scenarios as a combination of:

- **3 levels of censoring rates:**
  - X = 30%, Y = 30%
  - X = 60%, Y = 30%
  - X = 60%, Y = 60%

✖

- **3 survival curve patterns**:
  - Pattern 1: curves substantially separated
  - Pattern 2: curves close to each other
  - Pattern 3: curves intersecting



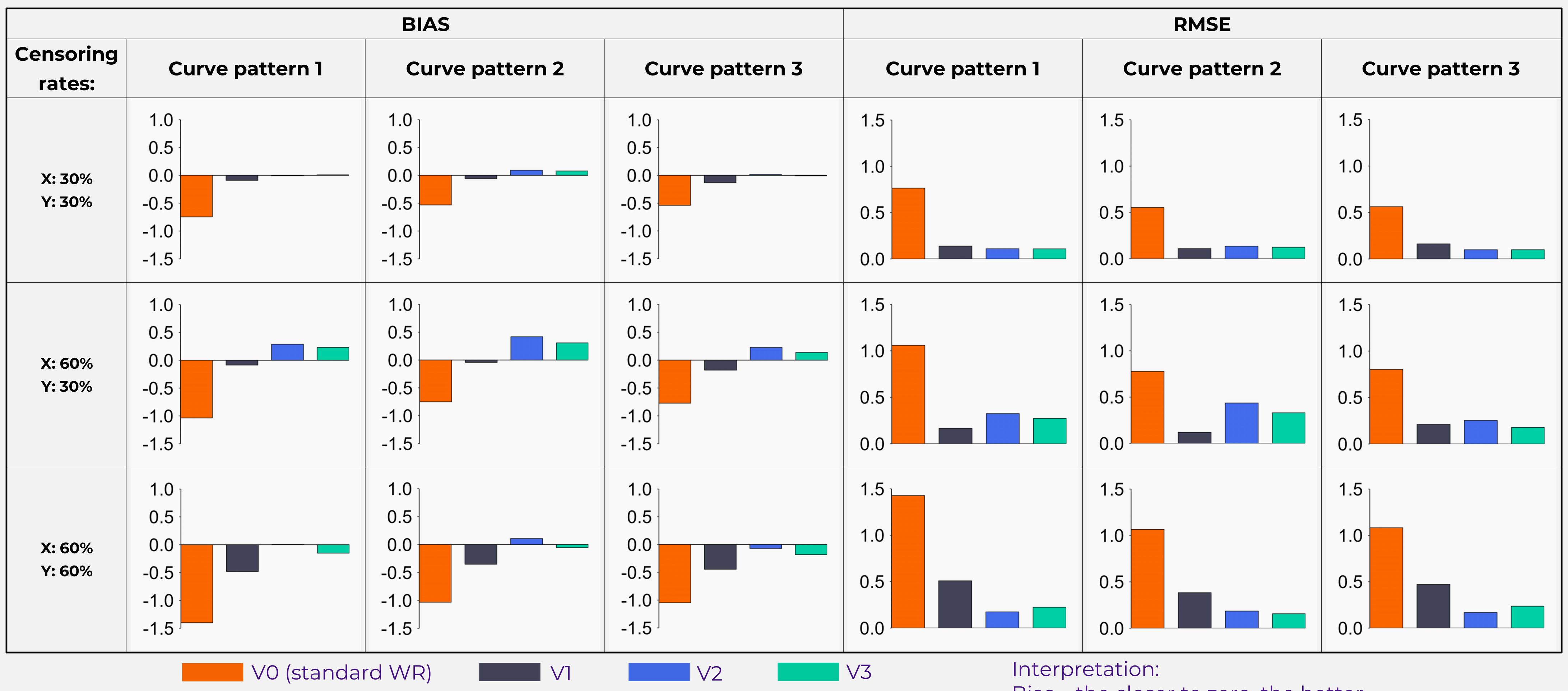
Methods of testing

- Simulation approach: 1,000 random simulations of events and censoring on E1, and of the score on E2
- Gold standard: standard WR calculated under a hypothetical scenario with no censoring
- Measure of method fit:
  - Bias of log(WR)
  - Root Mean Squared Error (RMSE) of log(WR)

Results

Win ratio estimations in different variants of probabilistic approach, compared to standard WR calculation

Scenario		Reference WR (no censoring)	Standard (deteministic) WR	Probabilistic WR		
Distribution of the 1st endpoint	Censoring rate per arm			Variant 1	Variant 2	Variant 3
Curve pattern 1	X: 30%, Y: 30%	1.77	0.83	1.61	1.75	1.78
	X: 60%, Y: 30%		0.62	1.62	2.36	2.22
	X: 60%, Y: 60%		0.43	1.09	1.78	1.51
Curve pattern 2	X: 30%, Y: 30%	1.01	0.59	0.95	1.10	1.09
	X: 60%, Y: 30%		0.47	0.96	1.53	1.38
	X: 60%, Y: 60%		0.36	0.71	1.12	0.96
Curve pattern 1	X: 30%, Y: 30%	1.04	0.60	0.91	1.05	1.03
	X: 60%, Y: 30%		0.48	0.87	1.30	1.20
	X: 60%, Y: 60%		0.36	0.66	0.97	0.87



Interpretation:
Bias – the closer to zero, the better
RMSE – the smaller the better

Discussion & Conclusion

- Standard WR estimates can be highly misleading; higher censoring rates lead to substantially increased bias and RMSE.
- Probabilistic methods evaluated here show potential to reduce both bias and RMSE.
- Performance varied by scenario: the V1 method, which applied the probabilistic approach only in Period I, performed better under imbalanced censoring, whereas V2 and V3 showed better suitability with balanced censoring.
- Further investigation and testing of methods applicable for non-random censoring patterns (e.g. incorporating IPCW modeling) merit further exploration.

References

1. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J. 2012 Jan;33(2):176-82. doi: 10.1093/eurheartj/ehr352. Epub 2011 Sep 6. PMID: 21900289.
2. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Stat Med. 2010 Dec 30;29(30):3245-57. doi: 10.1002/sim.3923. PMID: 21170918.
3. Péron J, Buyse M, Ozenne B, Roche L, Roy P. An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. Stat Methods Med Res. 2018 Apr;27(4):1230-1239. doi: 10.1177/0962280216658320. Epub 2016 Aug 2. PMID: 27487842.
4. De Backer M, Legrand C, Péron J, Lambert A, Buyse M. On the use of extreme value tail modeling for generalized pairwise comparisons with censored outcomes. Pharm Stat. 2023 Mar;22(2):284-299. doi: 10.1002/pst.2271. Epub 2022 Nov 2. PMID: 36321470.
5. Stute W. The statistical analysis of Kaplan-Meier integrals, IMS Lecture Notes Monogr. Ser., 1995: 231-254 (1995) doi: 10.1214/lnms/1215452223
6. Parner, E. T., & Overgaard, M. (2025). Estimation of win, loss probabilities, and win ratio based on right-censored event data. Scandinavian Journal of Statistics, 52(1), 170–184. doi: 10.1111/sjos.12734
7. Cui, Y., & Huang, B. — WINS: The R WINS Package. R package version 1.5.1 (2025). https://CRAN.R-project.org/package=WINS

Abbreviations

IPCW, Inverse Probability of Censoring Weights; RCT, Randomized Clinical Trials; RMSE, Root Mean Square Error; WR, Win Ratio

Contact
Mateusz Nikodem
Mateusz.Nikodem@putassoc.com

Find out more at putassoc.com

All information is © 2025 Putnam LLC.