

Why Is My Study Underpowered? Limitations of Formula-Based Power Calculations in Real-World Survival Studies and a Practical Correction

Haoyuan Wang MS^{1,2}, Yiwen Lu PhD^{1,2}, Yuqing Lei MS^{1,3}, Qiong Wu PhD⁴, Yong Chen, PhD^{1,2,3}

¹The Center for Health AI and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, Pennsylvania, USA;

²The Graduate Group in Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania, USA;

³Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

⁴Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

NOTE: We are still considering the name of our method, once finalized, I will replace the [Placeholder] with the actual name in the presentation.

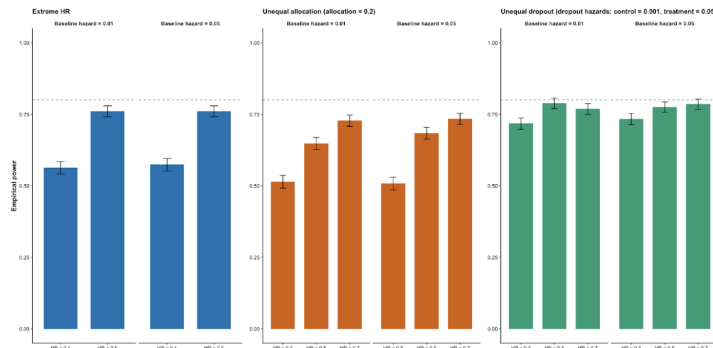
Background

Time-to-event analyses are central to OHDSI and real-world evidence studies, but sample size planning is often still based on closed-form formulas developed for idealized trial settings. Schoenfeld and Freedman-type calculations provide fast approximations for log-rank and Cox comparisons, yet real-world databases often have unequal treatment uptake, differential dropout, variable follow-up, and uncertain baseline hazards.¹⁻³ In these settings, satisfying the expected number of events may not guarantee nominal empirical power. This is especially important for OHDSI studies because multi-database networks and OMOP-formatted cohorts can support rapid feasibility assessment, but feasibility results must be translated into designs that are adequately powered under the data conditions actually expected in each study. We developed [PLACEHOLDER], an empirical power calibration workflow for real-world survival studies, to help investigators move from formula-based sufficiency to simulation-validated sample size planning.

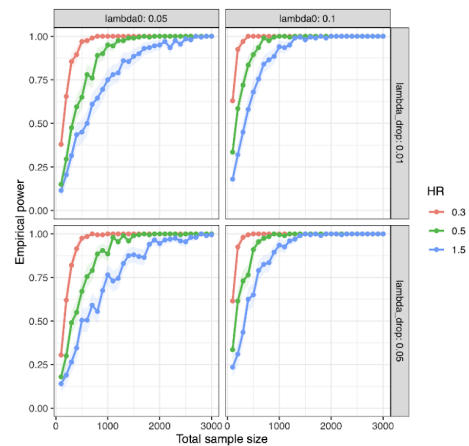
Methods

[PLACEHOLDER] uses the same inputs typically available during real-world data study design: baseline event hazard in the comparator arm, anticipated hazard ratio, planned allocation proportion, annual attrition hazard, fixed follow-up time, type I error rate, and target power (Figure 1). First, a Schoenfeld-based approximation is used to obtain an initial total sample size. Second, Monte Carlo simulations generate exponential event and censoring times under the proposed design and estimate empirical power as the proportion of simulated log-rank tests with $p < 0.05$. Third, a bracketing and integer bisection search iteratively increases or decreases the total sample size until the smallest value achieving the target power is found. Final confirmation simulations are then used to report empirical power and confidence intervals. We evaluated the workflow in simplified simulation scenarios and in a claims-based recombinant zoster vaccine (RZV) effectiveness use case among patients with systemic lupus erythematosus (SLE).⁴

(a) Empirical Power Simulation Result



(b) Concavity of Power Curve



(c) Insufficient Power Due to Concavity of Power Curve

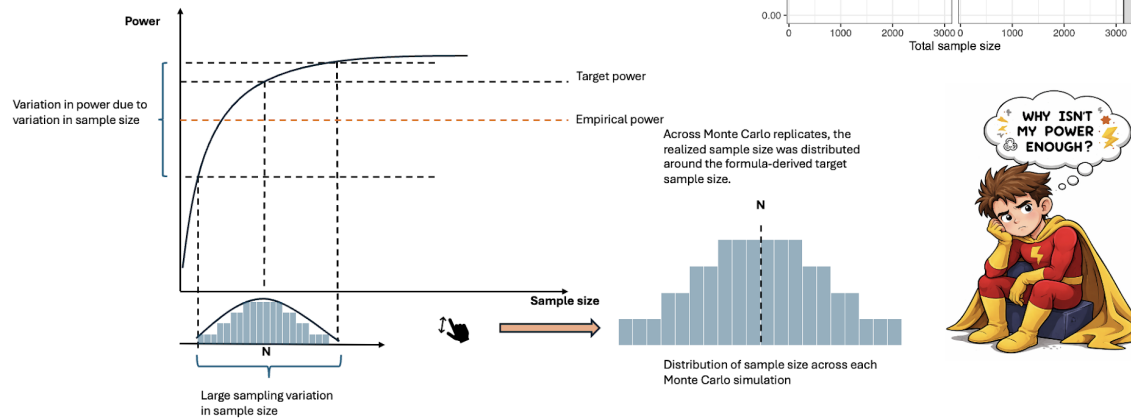


Figure 1: Overall Workflow of [PLACEHOLDER]

Results

Formula-based designs frequently failed to achieve nominal 80% power despite matching the formula-implied expected event counts. In stress-test simulations (Table 1), empirical power fell to 0.56 under an extreme hazard ratio of 0.10, to 0.72 under differential dropout with HR=0.30, and to 0.49 under 20:80 treatment-control allocation with HR=0.30. In the RZV-SLE case study, we parameterized the unvaccinated herpes zoster incidence at 19.8 per 1,000 person-years and examined HR values from 0.20 to 0.50 under a 1:4 vaccinated-to-unvaccinated allocation.⁴ For HR=0.30, the Schoenfeld-based design required 692 total patients and achieved empirical power of approximately 0.49. [PLACEHOLDER] increased the calibrated total sample size to 1,179 and restored empirical power to approximately 0.80. Across HR scenarios, the calibrated designs required larger sample sizes but produced near-nominal power, showing that direct empirical calibration can convert the same clinical assumptions into an operationally reliable design.

HR	Before/After calibration	Sample size	Sample size for treatment arm	Sample size for control arm	Number of events	Empirical power	Confidence Interval
0.2	Before calibration	397	79	318	18.94	0.34	(0.32, 0.36)
	After calibration	843	169	674	40.41	0.80	(0.79, 0.82)
0.3	Before calibration	692	138	554	33.84	0.49	(0.47, 0.51)

	After calibration	1179	236	943	57.57	0.80	(0.79, 0.82)
0.4	Before calibration	1167	233	934	58.43	0.61	(0.59, 0.63)
	After calibration	1732	346	1386	86.56	0.80	(0.79, 0.81)
0.5	Before calibration	1994	399	1595	102.10	0.66	(0.64, 0.68)
	After calibration	2654	531	2123	135.85	0.79	(0.78, 0.80)

Table 1: Results from Real World Data Use Case.

Conclusions

Formula-based power calculations can be optimistic in real-world data survival studies when allocation, censoring, or effect-size conditions depart from ideal assumptions. [PLACEHOLDER] provides a practical OHDSI-relevant workflow that treats standard formulas as a starting point, then calibrates total sample size to the empirical power expected under the proposed data-generating process. This approach can improve protocol planning, feasibility assessment, and sensitivity analysis for observational studies conducted using OMOP-based real-world data. Future work will extend the framework to flexible hazards, non-proportional hazards, and site-specific calibration across distributed data networks.

References

1. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39(2):499-503. doi:10.2307/2531021.
2. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Stat Med*. 1982;1(2):121-129. doi:10.1002/sim.4780010204.
3. Cox DR. Regression models and life-tables. *J R Stat Soc Series B*. 1972;34(2):187-202. doi:10.1111/j.2517-6161.1972.tb00899.x.
4. Mayer SE, Kluberg SA, Spence O, Oraichi D, Seifert H, Ali O, et al. Effectiveness and safety of the recombinant zoster vaccine in adult patients with systemic lupus erythematosus: a claims-based retrospective cohort study in the USA. *RMD Open*. 2025;11(3):e005839. doi:10.1136/rmdopen-2025-005839.