Borrowing Strength Without Borrowing Bias in Oncology Using Power Priors and Commensurate Priors for Long Term Overall Survival Extrapolation

Evidence Pharmaco®
Science-led HEOR & Tech-driven AI

MSR143

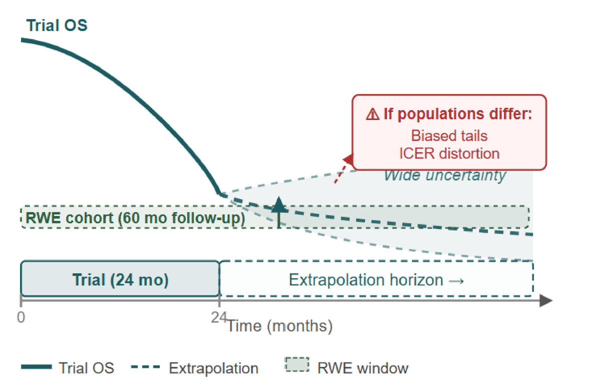
TheoRlytics

Shubhram Pandey[2], Rashi Rani[1], Kushagra Pandey[1], Akanksha Sharma[2]
[1]Heorlytics Pvt. Ltd., Mohali, India; [2] Pharmacoevidence Pvt. Ltd., Mohali, India

INTRODUCTION

- Health Technology Assessment (HTA) submissions in oncology require Overall Survival (OS) extrapolation to a lifetime horizon, yet pivotal trials report median follow-up of 12-30 months leaving survival tails sparsely observed and highly uncertain at the point of decision[1,2]
- Real-world evidence (RWE) from registries and electronic health record (EHR) cohorts can supplement trial data to improve long-term precision, but systematic population differences (prognosis, prior treatment, disease stage) risk distorting extrapolated tails through prior data conflict[3,4]
- The comparative operating characteristics of these strategies and their downstream impact on incremental cost-effectiveness ratio (ICER) estimates have not been formally evaluated in an oncology extrapolation simulation context
- Power priors and commensurate priors provide principled Bayesian frameworks for external borrowing, but fixed-weight approaches do not self-limit under conflict motivating conflict-robust extensions (estimated weights, robust mixture priors)[5,6]

Figure 1: The Borrowing Problem



OBJECTIVE

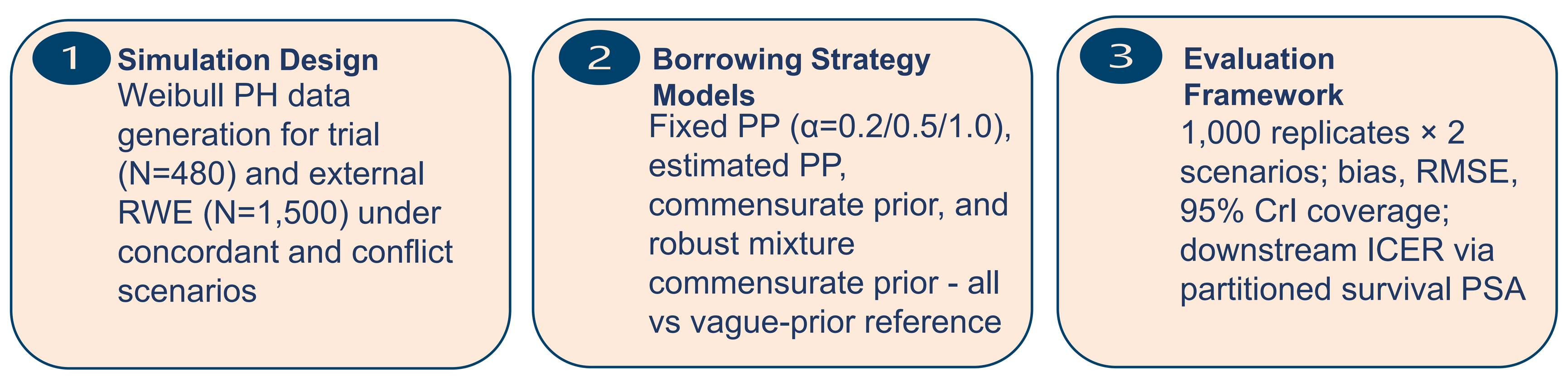
- To compare five Bayesian borrowing strategies fixed and estimated power priors, commensurate prior, and robust mixture commensurate prior for long-term OS extrapolation in oncology, evaluating bias, root mean square error (RMSE), and 95% coverage of five year restricted mean survival time (RMST) under concordant and conflict scenarios across 1,000 simulation replicates, and quantifying downstream ICER consequences via a partitioned survival cost-effectiveness model

METHODS

Phase 1 - Simulation Design

- A randomized trial setting was simulated (N=480; 1:1 allocation; 24-month administrative censoring) with event generation from a Weibull proportional hazards model (shape γ=1.2; baseline hazard λ=0.015; treatment log-HR β=−0.35; HR≈0.70); approximately 214 deaths observed per replication (44.6% event rate)
- An external real-world control cohort (N=1,500; 60-month follow-up) was generated under two pre-specified population scenarios: Concordant (λ_ext=0.015 identical baseline hazard to trial controls) and Conflict (λ_ext=0.008 systematically lower hazard, reflecting a better prognosis registry population with less advanced disease)
- Shape parameter γ was held constant across both scenarios to isolate the effect of scale conflict on borrowing performance; 1,000 independent Monte Carlo replicates were generated per scenario
- Primary estimand: 5-year RMST for the control arm, computed analytically as $\int_0^{\infty} \exp(-\lambda t^{\gamma})\,dt$ from true parameters to serve as the reference for bias and coverage assessment

Figure 2: Overall phases

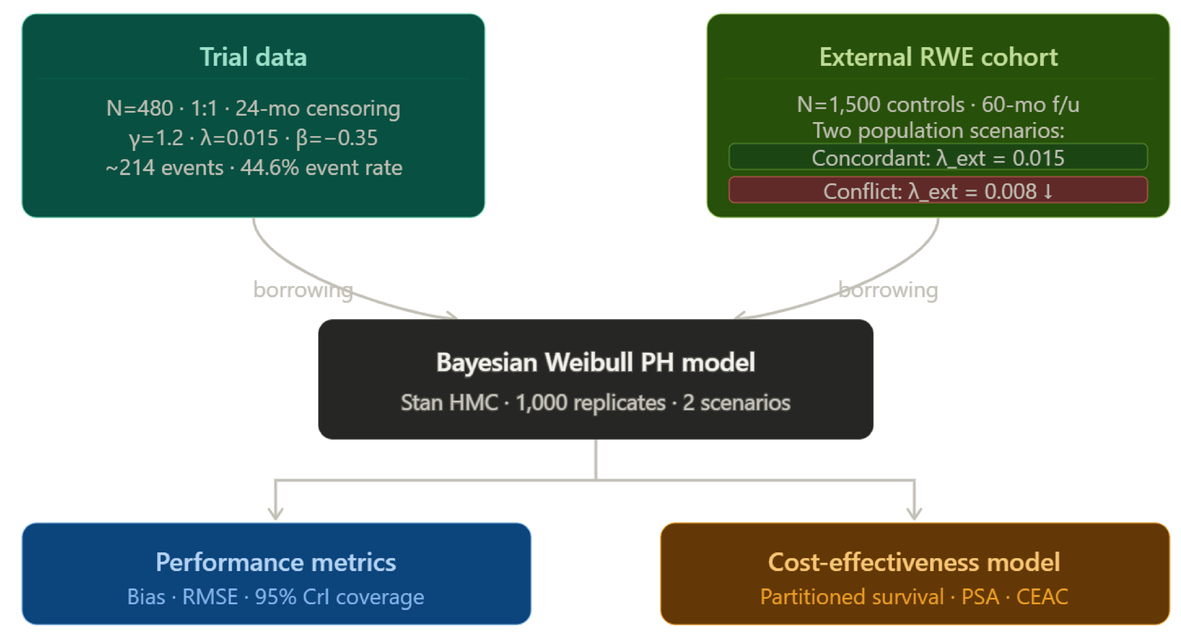


ICER: Incremental Cost-Effectiveness Ratio; PH: Proportional Hazards; PP:Power Prior; RWE: Real-World Evidence;PSA: Probabilistic Sensitivity Analysis; RMSE:root mean square error

Model Structure

- Bayesian Weibull PH likelihood: $h(t_i|x_i) = \lambda\gamma t_i^{(\gamma-1)}\cdot\exp(\beta x_i)$; $S(t_i|x_i) = \exp(-\lambda t_i^{\gamma}\cdot\exp(\beta x_i))$; full likelihood = $\Sigma\,[d_i \cdot \log h(t_i|x_i) + \log S(t_i|x_i)]$, accommodating right-censoring for both trial and external data
- Three parameters: baseline hazard scale λ, shape γ, and treatment log-HR β; vague priors as baseline $\log \lambda \sim N(-4, 2)$; $\log \gamma \sim N(0, 1)$; $\beta \sim N(0, 2)$ **(Table 1)**
- External data exclusively informs (λ, γ); β estimated from trial only across all strategies - borrowing cannot distort the treatment effect estimate
- Each borrowing strategy replaces the vague prior on (log λ, log γ) with an externally-informed prior; the likelihood structure and β prior remain unchanged **(Figure 3)**

Figure 3: Simulation architecture



CEAC: Cost-Effectiveness Acceptability Curve; PH: Proportional Hazards; PSA: Probabilistic Sensitivity Analysis; RMSE: Root mean square error

Borrowing Strategies

- No borrowing: Vague priors retained; external data excluded - serves as precision and coverage reference
- Fixed power prior (α = 0.2, 0.5, 1.0): Posterior $\propto L(\theta|D\_trial) \times L(\theta|D\_ext)^{\alpha} \times \pi_0(\theta)$; fixed α does not adapt to population compatibility - full borrowing (α=1.0) propagates conflict directly into the survival tail
- Estimated power prior: α ~ Beta(1,1) jointly estimated with model parameters; posterior $\hat{\alpha}$ self limits toward zero under conflict, providing automatic down weighting and a quantifiable borrowing diagnostic
- Commensurate prior (Hobbs et al., 2012): External data fitted first; posterior summary (μ_ext, σ_ext) used as location prior - $\log \lambda \sim N(\mu\_ext, \sigma^2\_ext/\tau)$; commensurability τ ~ Half Normal(0,1) estimated from trial data; large τ = tight borrowing, small τ = conflict-protective
- Robust mixture commensurate (Schmidli et al., 2014): $\pi(\log \lambda) = 0.70 \times N(\mu\_ext, \sigma^2\_ext/\tau) + 0.30 \times N(-4, 9)$; vague component ensures prior mass away from external location under conflict, vague component dominates and attenuates bias propagation

Table 1: Model Parameters

Strategy	Key parameter	Mathematical form	Conflict-robust
No borrowing	-	Vague prior on (λ, γ)	Reference
PP α = 0.20	α = 0.20, fixed	$L(\theta\|D\_ext)^{0.2} \times \pi_0(\theta)$	No
PP α = 0.50	α = 0.50, fixed	$L(\theta\|D\_ext)^{0.5} \times \pi_0(\theta)$	No
PP α = 1.00	α = 1.00, full	$L(\theta\|D\_ext)^{1} \times \pi_0(\theta)$	No - high risk
PP $\hat{\alpha}$ estimated	α ~ Beta(1,1)	α jointly estimated with θ	Partial

PP: Power Prior

RESULTS

- 5-year RMST (control arm): 29.2 months (2.43 years); all borrowing strategies recovered this with near-zero bias under concordance
- Under concordance, all borrowing strategies maintained 93-95% credible interval coverage no meaningful calibration penalty for informative borrowing when populations are compatible
- Posterior uncertainty (mean SD of RMST estimates) reduced 31-38% relative to no borrowing; commensurate and robust mixture achieved 34% reduction; full PP achieved 38%
- RMSE reduced 12-18% across strategies; gains were modest because variance reduction was partially offset by small prior-induced shrinkage toward the external posterior mean
- Mean posterior $\hat{\alpha}$ = 0.74 under concordance (estimated PP) data correctly identified the external cohort as highly compatible and assigned high borrowing weight automatically **(Table 2)**

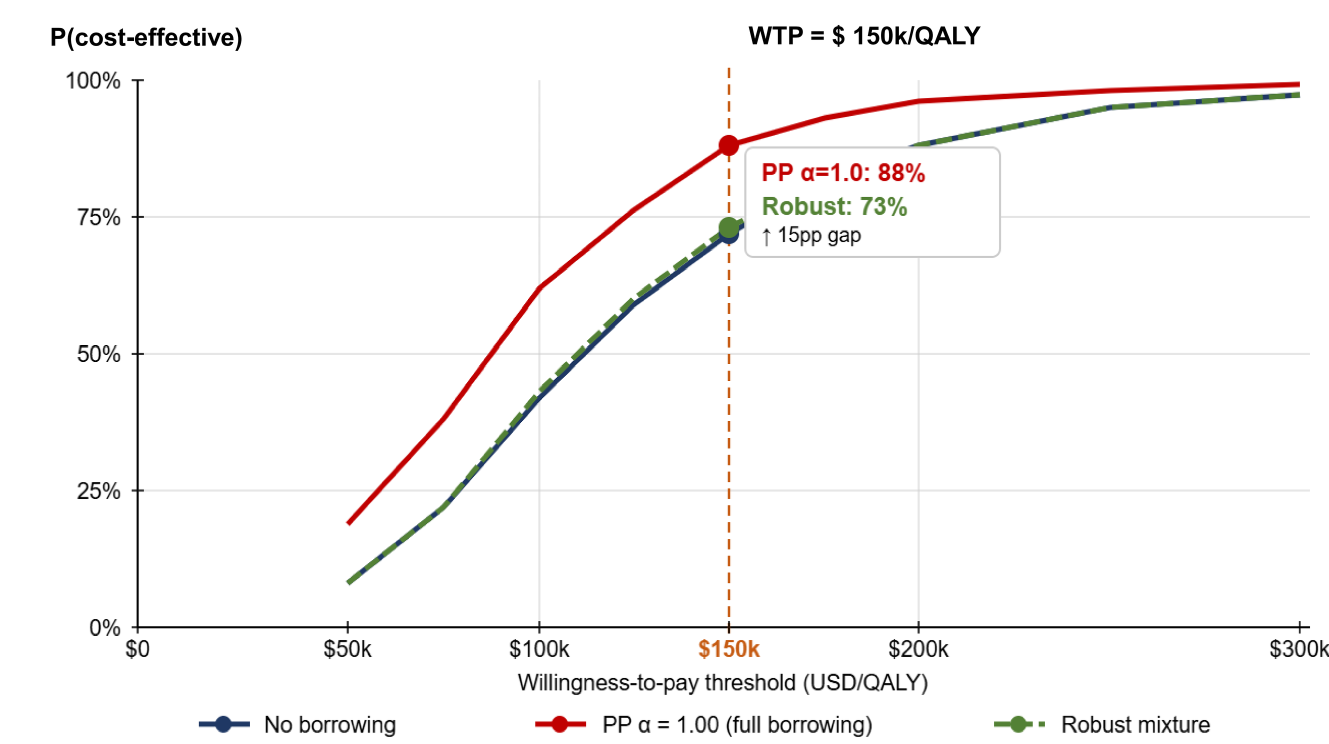
Table 2: Model Performance Summary

Strategy	Bias (LY)	SD (mo)	SD reduction	RMSE (mo)	RMSE reduction	Coverage
No borrowing	−0.01	3.18	—	3.18	—	95.2%
PP α = 0.20	0.00	2.19	31%	2.80	12%	95.1%
PP α = 0.50	+0.01	2.07	35%	2.72	14%	94.7%
PP α = 1.00	+0.02	1.97	38%	2.61	18%	93.8%
PP $\hat{\alpha}$ estimated	+0.01	2.11	34%	2.69	15%	94.3%
Commensurate	0.00	2.09	34%	2.67	16%	94.5%
Robust mixture	+0.01	2.09	34%	2.67	16%	94.3%

LY: Life Years; mo: months; PP: Power Prior; RMSE: Root mean square error; SD: Standard Deviation

- Under concordance, borrowing narrowed ICER credible interval width by ~30% without shifting the mean ICER ($128,000/QALY) - precision gain at zero cost to validity
- Under conflict, full Power Prior (α=1.0) distorted mean ICER to $99,000/QALY - a $29,000 downward shift crossing the $100k/QALY reimbursement threshold; conflict-robust strategies held mean ICER at $124,000–$130,000/QALY, within $6,000 of the true baseline
- At WTP $150,000/QALY, full Power Prior under conflict inflate d the probability of cost-effectiveness from 72% to 88% - a 16 percentage-point overstatement of treatment value that would mislead reimbursement decisions **(Figure 4)**

Figure 4: Cost-effectiveness acceptability curves



PP:Power Prior; QALY: Quality-Adjusted Life Year; RMSE:root mean square error; WTP: Willingness-to-Pay

CONCLUSION

- Bayesian borrowing from external RWE meaningfully improves OS extrapolation precision under population concordance -reducing posterior uncertainty by 31–38% and RMSE by 12–18% while maintaining nominal 95% coverage, with no distortion to mean ICER estimates
- Full fixed borrowing (PP α=1.0) is unsafe under conflict - introducing −0.41 LY bias, collapsing coverage to 62%, and shifting mean ICER by $29,000/QALY in a direction that would incorrectly favor reimbursement; fixed partial borrowing (α=0.5) also degrades materially (coverage 89%),conflict-robust strategies - estimated PP and robust mixture commensurate - self-limit under population heterogeneity, recovering bias to −0.07 LY and coverage to 90–92% without requiring manual specification of the borrowing weight; posterior $\hat{\alpha}$ and τ distributions provide transparent, interpretable diagnostics
- The choice of borrowing strategy is not a statistical technicality - it is a decision-relevant input: a 15 percentage-point overstatement of cost-effectiveness probability at a $150k/QALY threshold has direct implications for reimbursement conclusions in NICE, CADTH, and EMA-HTA submissions
- Recommended practice for HTA submissions: apply robust mixture commensurate or estimated power prior; report posterior $\hat{\alpha}$ / τ distributions as borrowing diagnostics; pre-specify conflict thresholds in the SAP; and conduct scenario analyses under both concordant and conflict assumptions to bound extrapolation uncertainty transparently

References
1. NICE DSU TSD 14, 2014
2. Latimer NR et al. Med Decis Making. 2013
3. Ciani O et al. Value Health. 2022
4. Guyot P et al. BMC Med Res Methodol. 2012
5. Ibrahim JG, Chen M-H. Lifetime Data Anal. 2000
6. Schmidli H et al. Biometrics. 2014

Correspondence:
Shubhram Pandey (Shubhram.Pandey@pharmacoevidence.com)

Disclosure:
PB, AS, KP, SM, and SP, the authors declare that they have no conflict of interest