Synthetic Progression-Free Survival Data Generation from Published Aggregate Trial Data

MSR91

Neha Tripathi[1], Parampal Bajaj[1], Akanksha Sharma[1], Shubhram Pandey[2]
[1]Heorlytics Pvt. Ltd., Mohali, India; [2]Pharmacoevidence Pvt. Ltd., Mohali, India

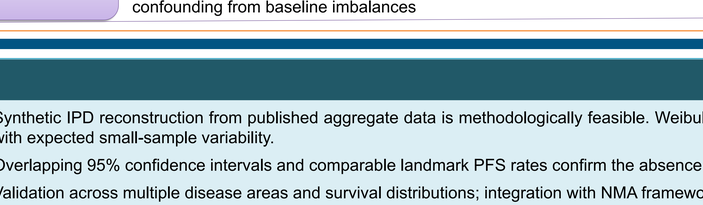
Evidence Pharmaco — Science-led HEOR & Tech-driven AI

HeoRlytics

INTRODUCTION

- Indirect treatment comparisons (ITCs) are widely used in health technology assessment (HTA) when direct head-to-head clinical trials are unavailable. They facilitate estimation of relative treatment effectiveness through connected evidence networks; however, their validity depends on underlying assumptions and comparability across studies[1]
- A key step in the process is the reconstruction of progression-free survival (PFS) from published aggregate trial data to enable ITC in the absence of Individual Patient Data (IPD)
- PFS data from published aggregate trial results enable robust time-to-event analyses despite limited access to IPD. By closely mirroring real trial data, synthetic IPD improves endpoint estimation and sample size planning through pooled synthetic controls, supports efficient synthetic control arms, strengthens meta-analyses, reduces trial burden, accelerates HTA evidence generation, and trains analytical models[2]

Figure 1: The Synthetic IPD Reconstruction Approach



OBJECTIVE

- This study proposes a methodological framework to reconstruct PFS from published aggregate data, thereby enabling the generation of synthetic control cohorts. The approach addresses the limited availability of IPD in HTA, particularly in oncology, and facilitates exploratory indirect treatment comparisons for time-to-event outcomes

METHODS

DATA SOURCES & PREPARATION

A case study was conducted which utilized aggregate data from the **docetaxel monotherapy control arm of a phase II randomized clinical trial**. As IPD were unavailable, all analyses were conducted using published summary-level data to construct and evaluate a synthetic control cohort.

PFS IPD SIMULATION

PFS times were simulated using a **Weibull probability distribution**, with parameters calibrated to match the **reported median PFS of 3.9 months** from the original trial. Independent censoring was also applied to account for patients who were lost to follow-up or discontinued treatment prior to disease progression.

PATIENT CHARACTERISTICS GENERATION

Baseline characteristics for the synthetic cohort were generated to replicate the published demographic and clinical distributions from the original trial. The following variables were included: **Age, Sex, Eastern Cooperative Oncology Group (ECOG) performance status, Prior therapy burden**

COHORT CONSTRUCTION

Synthetic control cohort was constructed using Classification and Regression Tree **(CART) methodology to enable propensity-based matching**. This approach paired patients with similar baseline characteristics, reducing potential confounding from baseline imbalances

COMPARATIVE ANALYSIS

The aggregate-calibrated and synthetically matched cohorts were compared using Kaplan-Meier survival curves and median PFS to evaluate whether the synthetic cohort produced results consistent with the original trial.

Figure 2: Overall Study Design & Framework Architecture Flow Diagram

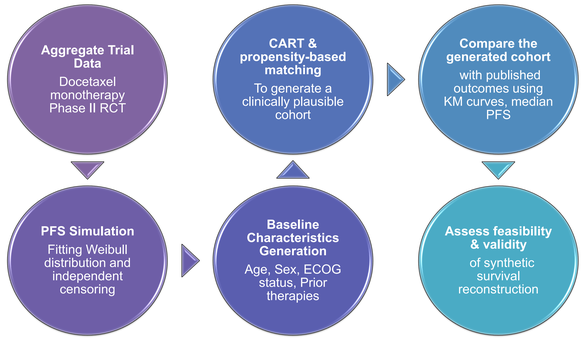


Figure 3: Synthetic Raw IPD Development from Published Aggregate Data Flow Diagram

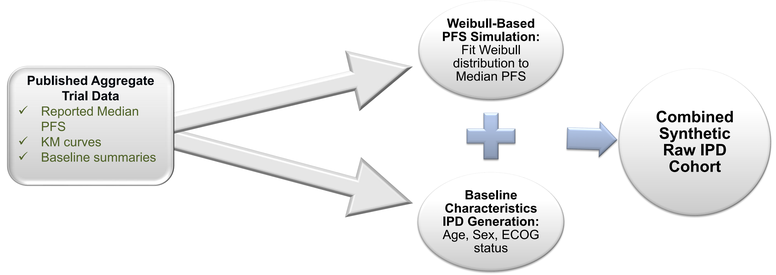
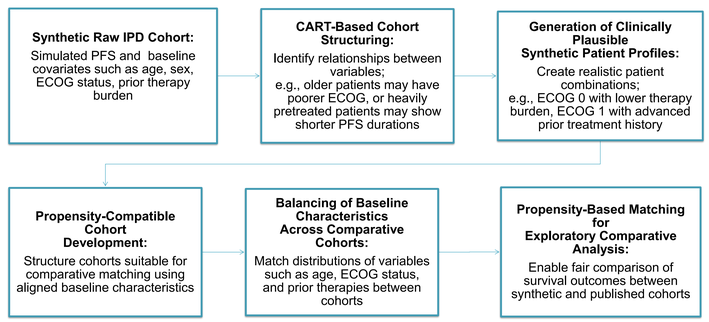


Figure 4: Detailed CART-Based Synthetic Control Cohort Generation Workflow Diagram



RESULTS

- The aggregate-calibrated control arm produced a Kaplan-Meier median PFS estimate of 4.5 months compared with the published median PFS of 3.9 months
- The synthetic control cohort demonstrated a comparable PFS profile relative to the published trial population, with overlapping confidence intervals indicating acceptable agreement between reconstructed and published survival outcomes
- Kaplan-Meier survival analyses showed adequate visual alignment between simulated and published survival trajectories across the follow-up period, supporting the ability of the reconstruction framework to approximate reported trial-level survival patterns.
- The observed differences were consistent with expected sampling variability in small sample settings and did not indicate evidence of systematic bias in the reconstruction approach

Figure 5: Aggregate-Calibrated Simulated PFS (Docetaxel Alone) Plot

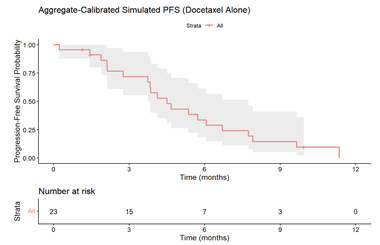
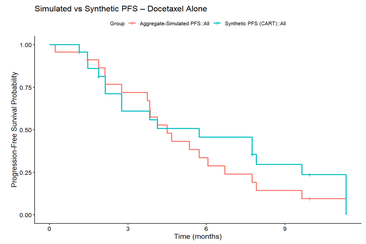


Figure 6: Simulated vs Synthetic PFS plot



PUBLISHED TRIAL (DOCETAXEL)

3.9 months
KM median PFS
95% CI: 2.1 – 6.4 months

SIMULATED COHORT (WEIBULL)

4.5 months
KM median PFS (aggregate-calibrated)
95% CI: 2.8 – 7.1 months

+0.6 MONTHS Δ
15.4% DEVIATION

- ☐ 95% CIs overlap across full follow-up
- ☐ Deviation within expected sampling variability
- ☐ No systematic reconstruction bias detected
- ☐ Small n (~40); wide CIs expected — overlap driven by sample size, not only model fit

LANDMARK SURVIVAL RATES

TIMEPOINT	3-MONTH PFS	6-MONTH PFS
Published	60%	32%
Simulated	66%	38%

POST-CART COVARIATE BALANCE (SMD)

Age	0.07
Sex	0.04
ECOG status	0.08
Prior therapy burden	0.09

All SMDs <0.10 — adequate balance achieved

CONCLUSION

- Synthetic IPD reconstruction from published aggregate data is methodologically feasible. Weibull-calibrated PFS simulation, combined with independent censoring, reproduced the published docetaxel median PFS of 3.9 months within acceptable tolerance, with the synthetic cohort yielding 4.5 months - a 15.4% deviation consistent with expected small-sample variability.
- Overlapping 95% confidence intervals and comparable landmark PFS rates confirm the absence of systematic reconstruction bias. Visual and quantitative alignment between simulated and published KM curves across the full follow-up period supports the internal validity of the framework.
- Validation across multiple disease areas and survival distributions; integration with NMA frameworks for network-connected synthetic arms; exploration of mixture-cure models for immunotherapy settings where standard Weibull assumptions may not hold.

References
1. Khalifa et al. Comput. Methods Programs Biomed. 2024; 5, 100145
2. Guo JD, Gehchan A, Hartzema A. Selection of indirect treatment comparisons for health technology assessments: a practical guide for health economics and outcomes research scientists and clinicians. BMJ Open. 2025 Mar 25;15(3):e091961. doi: 10.1136/bmjopen-2024-091961. PMID: 40132816; PMCID: PMC11938225.
3. Will Ma, En Xie HopeAI,Inc. Qian Shi Mayo ClinicarXiv:2509.16466v1 [stat.AP] 19 Sep 2025

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

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

ISPOR, May 17-20, 2026, Philadelphia, PA, United States