Methodology

Approaches for Enhanced Extrapolation of Long-Term Survival Outcomes Using Electronic Health Records of Patients With Cancer

Xiaoliang Wang, PhD, MPH, Blythe J. Adamson, PhD, MPH, Andrew Briggs, DPhil, MSc, Katherine Tan, PhD, Danielle Bargo, MSc, Shuhag Ghosh, MBA, Shrujal Baxi, MD, MPH, Scott Ramsey, MD, PhD

A B S T R A C T

Objectives: This study aimed to demonstrate enhanced survival extrapolation methods using electronic health record-derived real-world data (RWD).

Methods: The study population included patients diagnosed of ER+/HER2− metastatic breast cancer who started first-line treatment with anastrozole or letrozole between November 18, 2014, and November 18, 2015. Two patient cohorts were constructed: a clinical trial cohort from digitized MONARCH-3 clinical trial results and a RWD cohort from a deidentified electronic health record-derived database. RWD patients were weighted to trial baseline covariate distributions. Standard parametric approaches were applied to trial data and a "best-fit" model was selected. We demonstrate traditional and enhanced hybrid (pooling with weighted RWD at start, 75%, or end of trial) extrapolation approaches.

Results: Observed and estimated 5-year progression-free survival (PFS) rates in extrapolating the trial control arm (n = 165) were comparable across all methods. Compared with the observed 5-year mean PFS in the RWD cohort (n = 118) of 20.4 months (95% confidence interval [CI] 16.9-23.8), there was some variation among studied methods. Best-fit standard parametric model (log-normal) had 5-year mean PFS of 21.3 months (95% CI 18.2-24.9), and for the hybrid methods in order of estimate conservativeness was start of trial (20.8 months; 95% CI 18.5-23.2), 75% of trial (21.3 months; 95% CI 18.1-24.5), and end of trial (21.8 months; 95% CI 18.8-25.2).

Conclusions: Our study leverages RWD to enhance long-term survival extrapolation. Future use cases should include applying patient eligibility criteria, weighting on baseline characteristics, and choice of time window to add RWD to trial data.

Keywords: cancer treatment, clinical trial, real-world data, survival extrapolation.

VALUE HEALTH. 2021; ■(■):■–■

Introduction

A common problem for health technology appraisers of newer oncology treatments is that the duration of follow-up in oncology clinical trials is often much shorter than the expected survival of patients. To address this issue, survival extrapolation methods are commonly used to estimate life-years and quality-adjusted life-years for economic analyses and health technology assessments (HTAs).[1] Extrapolations typically show expected outcomes for years or even decades beyond trial time horizons, yet they vary widely in terms of the assumptions and accuracy on the longer-term survival outcomes. In response, HTA agencies such as the National Institute for Health and Care Excellence have endorsed published Decision Support Unit Technical Support Documents for extrapolating survival from clinical trial data.[2,3] Most recommend incorporating real-word data (RWD), such as data from disease registries, to validate the plausibility of extrapolations and better inform long-term survival estimates. In recent years, clinical depth, recency, and long-term data availability have improved with electronic health records (EHRs) of patients with cancer. The use of EHR-derived RWD in HTA assessments can allow for selection of real-world patients similar to those in the trial, extending follow-up time, contextualizing results, and improving evaluation of long-term outcomes observed in trial patients. In addition to more precise matching of cohorts, having an EHR-derived RWD with sufficient follow-up (for currently available treatments) allows for the comparison of alternative methods for extrapolating survival beyond the trial with observed outcomes for real-world patients over an extended time horizon. It is currently unknown how specific choices of matching and extrapolation affect estimates of long-term survival.

To help fill this gap, the objective of this study was to enumerate and demonstrate novel techniques for enhanced trial survival extrapolation using long-term survival estimates from RWD. To illustrate the opportunity, we used data from the control arm of a large randomized clinical trial of patients with metastatic breast

cancer (mBC) and clinically comparable real-world patients identified using EHR-derived data.[4] In the primary analysis, we used RWD patients who were most similar to trial patients and focused on the control arm only. The trade-offs between clinical specificity and representativeness in identifying real-world comparator cohorts for cost-effectiveness research have been discussed elsewhere.[5]

Methods

To demonstrate the application of enhanced hybrid extrapolation methods, we conducted a case study using trial and RWD patients with ER-positive HER2-negative (ER+/HER2−) mBC who started first-line treatment with anastrozole or letrozole.

Patient Population—Clinical Trial Cohort

The clinical trial patient cohorts used in this study were reconstructed by digitizing published[6] survival curves and reported patient information available in the primary manuscripts from the MONARCH-3 trial. Details on the study eligibility criteria have been previously published,[4] and a summary is provided in the Appendix Note 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013. In brief, MONARCH-3 is a double-blind, randomized phase III study of abemaciclib or placebo plus a nonsteroidal aromatase inhibitor (AI) in 493 postmenopausal women with ER+/HER2− mBC who had no previous systemic therapy in the advanced setting. Only trial patients who received the placebo treatment of AI only are used in this analysis (n = 165).

Patient Population—Real-World Cohort

Data source

The RWD cohort was constructed using data from the Flatiron Health database, a US nationwide longitudinal, deidentified database derived from EHR data containing patient-level structured and unstructured data curated via technology-enabled abstraction.[6-8] By the end of the observation period used for this analysis (January 31, 2020), the database comprised more than 280 cancer clinics (~800 sites of care). An institutional review board approval of the study protocol was obtained before study conduct and included a waiver of informed consent.

A real-world patient cohort was selected according to trial-specific inclusion and exclusion criteria and trial-specific enrollment windows (Appendix Note 1 and Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013) for the MONARCH-3 trial of patients with mBC.[4] In brief, adult female patients with ER+/HER2− mBC were included if their treatment with either anastrozole or letrozole as first-line treatment initiated any time from November 18, 2014, to November 18, 2015, confirmed in unstructured information. Using MONARCH-3 trial inclusion/exclusion criteria, patients were excluded from the analysis if they had an Eastern Cooperative Oncology Group Performance Status of >1; had inadequate organ function; had evidence or history of central nervous system metastases; received a previous treatment with systemic chemotherapy for advanced disease, everolimus, or a cyclin-dependent kinase 4 and 6 inhibitor (ie, palbociclib, ribociclib, or abemaciclib); or were currently receiving an investigational drug in a clinical trial during the trial enrollment window (Fig. 1). Patients with missing or not documented organ function and functional status were included in the primary analysis.

Baseline characteristics weighting for RWD cohort

To improve the balance in patient characteristics between cohorts, patients in the RWD cohort were weighted on baseline characteristics to reflect the distribution of baseline characteristics of trial patients. We adopted an approach analogous to propensity score modeling that weights RWD patients to achieve cohort-level balance for all measured baseline characteristics.[9] Weights were estimated using the generalized method of moments estimator on trial reported moments (ie, means or proportions of variables).[10-12] Patients in the RWD cohort were weighted on trial median age (≤63 vs >63 years), race (white, other), number of metastatic sites at index date (1, 2, ≥3), site of metastasis (bone only, other), progesterone receptor status (positive, negative/missing), and use of AI (letrozole, anastrozole).

We also performed an exploratory analysis using 3 additional RWD cohorts with less restricted criteria, including no baseline characteristics weighting, no treatment window restriction, and less trial selection criteria.

Follow-up and outcomes

The length of possible follow-up time for each patient was defined as the difference in months between the start date of first-line therapy and the end date of the clinical trial or the most recent EHR activity date (clinic visit, drug administration, or expected end date of most recent oral therapy). The primary outcome for comparison in this study was the mean progression-free survival (PFS) time and PFS rate (%) at 5 years from the start date of first-line treatment. For RWD patients, the real-world PFS (rwPFS) event was disease progression or death, which was based on a clinician-anchored documentation of disease, an approach that has been shown to be scalable, reliable, and meaningful and may provide more context into real-world outcomes measures than response rates.[13,14] The 5-year PFS rate for clinical trial patients was extrapolated from trial outcomes whereas for real-world patients the 5-year rwPFS rate was observed and did not require extrapolation.

Statistical Analysis

We used the clinical trial and RWD cohorts to demonstrate approaches for enhanced extrapolation of patient outcomes combining information from a clinical trial with information from RWD patients and compared results with observed long-term outcomes in RWD (reference; approach 0) and a standard approach for extrapolation using only the clinical trial (approach 1). The 3 approaches for enhanced extrapolation used data from trial pooled with RWD data since index date (approach 2), trial with RWD added at the end of trial (approach 3), and trial with RWD added beyond the point of 75% trial follow-up completion (approach 4). Each extrapolation approach 1 to 4 is detailed below and summarized in Table 1. Potential systematic differences between extrapolated outcomes were assessed quantitatively and qualitatively from mean survival and survival rates at the end of real-world follow-up time and visual inspection of survival curves.

Standard trial extrapolation approach (approach 1)

Standard parametric models were fitted to the clinical trial data to extrapolate the outcomes.[1] The models included exponential, Weibull, Gompertz, log-logistic, log-normal, generalized Gamma, and generalized F distributions. Visual inspection, statistical tests including Akaike information criterion (AIC) and Bayesian information criterion (BIC), and plausibility of survival estimates were used to assess the goodness of fit of parametric distributions. Visual inspection helps to narrow the review of possible survival curves by eliminating those that are clearly not good fits. Statistical tests, in the order of smallest BIC and then smallest AIC, further refine the selection. The plausibility of survival estimates beyond trial follow-up period is the final criterion

Figure 1. Cohort selection diagram for real-world study population.

Step 1: Women with locoregionally recurrent or metastatic breast cancer: n = 18110

Step 2: Adult patients who are HR+/HER2– and receiving AI (letrozole or anastrozole monotherapy) or Abemaciclib as 1L therapy: n = 2939

Step 3: Exclude patients who received prior (neo)adjuvant endocrine therapy including anastrozole, letrozole, exemestane, tamoxifen: n = 2375

Step 4: Excluding patients with prior treatment with any CDK4/6 inhibitor (i.e, palbociclib, ribociclib, abemaciclib): n = 2343

Step 5: Excluding patients with prior treatment with everolimus: n = 2343

Step 6: Excluding patients with ECOG > 1: n = 2091

Step 7: Excluding patients with inadequate organ function (conditional on liver mets): n = 1088

Step 8: Excluding patients with CNS mets: n = 1062

Step 9: Excluding patients with history of HIV: n = 1062

Step 10: Excluding patients with history of hepatitis B and/or C: n = 1062

Step 11: Treatment start before November 11, 2015 and after 2014-11-18: n = 207

1L indicates first-line; AI, aromatase inhibitor; CDK, cyclin-dependent kinase; CNS, central nervous system; ECOG, Eastern Cooperative Oncology Group; HIV, human immunodeficiency virus; mets, metastases.

but sometimes the most difficult to assess because of a lack of appropriate comparator with sufficient follow-up. For example, we considered that it is unlikely that the 5-year PFS rate is <5% among patients with HR+/HER2− mBC. The estimated median survival time-to-event (PFS) from parametric models were also compared with the corresponding trial observed median survival times, if observed. A "best-fit" model was selected based on the criteria and was used to extrapolate longer-term outcomes from the fitted parametric models.

Enhanced methodological approaches (approaches 2-4)

To illustrate approaches for incorporating RWD into long-term survival extrapolation from trial data, we proposed a 2-stage enhanced method by including RWD patients during and beyond the trial time horizon. Parametric distributions selected as the "best-fit" model for trial patients were used for all survival extrapolations, adjusting for the source of patient cohorts (clinical trial vs RWD).

Trial data pooled with RWD since index date (approach 2).

This method fits parametric models, using combined survival data from both trials and RWD from the start of treatment to the end of RWD long-term follow-up (January 31, 2020), adjusting for patient cohorts (clinical trial vs RWD). All trial and RWD patients were included in this analysis with the same entry at the index date (e.g., start of treatment) in this study.

Trial with RWD added at the end of trial (Approach 3).

It is common that survival extrapolation was only performed from the end-of-trial follow-up, which sometimes included only a small number of patients.[15] Therefore, we evaluated whether adding additional RWD patients to this group of patients would help improve survival extrapolation. This method fits parametric models in 2 periods: (1) short-term survival estimates are generated using trial data until the end-of-trial follow-up duration, and (2) long-term survival estimates are generated by adding RWD patients with a duration of follow-up that met or exceeded that of the trial patients (see

Table 1. Approaches for extrapolation of patient outcomes.

Approaches*	Data sources used	Follow-up duration, months	Demonstration number
Traditional extrapolation	Clinical trial	31	1
Enhanced with RWD added at 0% (trial start)	Clinical trial + RWD	60	2
Enhanced with RWD added at 100% (trial end)	Clinical trial + RWD	60	3
Enhanced with RWD added at 75% (during trial)	Clinical trial + RWD	60	4

RWD indicates real-world data.
*All enhanced approaches used the same parametric model as the traditional extrapolation approach.

Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013). The included RWD patients had delayed entry at the end of trial.

Trial with RWD added at 75% trial follow-up completion (Approach 4). Due to the large amount of patients censored at the end of a trial, evidence review groups have proposed to extrapolate survival estimates at an earlier time (eg, 0%-30%) during the trial follow-up.[16] Given that we did not have individual-level data from trial patients, 75% trial completion was considered the closest scenario, based on the risk table published from the MONARCH-3 trial. This method fits parametric models in 2 periods: first, short-term survival estimates are generated using trial data only up until 75% of patients have observed events or were censored, and second, long-term survival estimates are generated by combining the remaining 25% of trial patient and RWD patients with follow-up duration beyond the cutoff time for 75% trial completion (see Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013). In the analysis, trial patients were included for their full follow-up time regardless of cutoff time, and RWD patients were included if they had outcomes or were censored after the cutoff time with a delayed entry at the cutoff time.

In the primary analysis for both end-of-trial and 75% completion methods, the duration of trial follow-up (months) was used to decide the cutoff time. Sensitivity analysis was conducted using the exact trial end date as the cutoff time. To check whether survival outcomes between trial and RWD cohorts were similar, we applied the same parametric model selection process to the RWD cohort in sensitivity analysis. We further tested the differences in PFS between trial and RWD cohort under the "best-fit" baseline hazard function in the pooled method. In addition, we performed exploratory extrapolation analysis using 30-year as a proxy of lifetime time horizon across all approaches.

All analyses are conducted using R version 3.6.1.

Results

Study Cohorts

The digitized trial data represented the experience of 165 patients with mBC who received first-line AI in MONARCH-3. Among the 18 110 patients with mBC in the real-world database, 1 062 patients met the inclusion and exclusion criteria and received AI as the first-line treatment. After further restriction to the same enrollment window as the trial, the RWD cohort included 207 patients with mBC (Fig. 1). After weighting on baseline characteristics of the RWD cohort to match the trial cohort, the effective sample size of the RWD cohort was 118 patients (Table 2). In brief, compared with the trial patients, RWD patients who met the selection criteria were more likely to be older, be white, have bone-

only metastasis, have lower number of metastases, and receive letrozole. Balance in the distributions of baseline characteristics was achieved in most characteristics after weighting with the exception of the number of metastatic sites and visceral metastatic involvement. The mean 5-year PFS for the long-term RWD cohort (approach 0; reference) was 20.4 months (95% CI 16.9-23.8), and the observed 5-year PFS rate was 11.0% (95% CI 6.1-19.8; Table 3).

Standard Trial Extrapolation (Approach 1)

In mBC using approach 1, the log-normal distribution was selected as the "best-fit" for MONARCH-3 patients who received an AI (model performance summarized in Appendix Fig. 1 and Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013). Using log-normal distribution in parametric modeling, the approach 1 model estimated median PFS for patients in the trial was 14.1 months (95% confidence interval [CI] 11.5-17.3), compared with the trial reported median PFS of 14.8 months (95% CI 11.7-19.2). The mean 5-year PFS for trial cohort (approach 1) was 21.3 months (95% CI 18.2-24.9), using MONARCH-3 trial extrapolation only with log-normal distribution (Table 3 and Fig. 2A). The 5-year PFS rate extrapolated from the trial data was 11.0%, the same as the observed 5-year PFS rate from long-term RWD.

Hybrid Methodological Approaches (Approaches 2-4)

RWD-enhanced survival extrapolation results from hybrid methods approaches 2 to 4 are reported in Table 3 and Figure 2B to D. Among the 3 hybrid methods, the mean 5-year PFS estimates from the approach 2 pooled method were the lowest and had the narrowest uncertainty (mean PFS 20.8 months; 95% CI 18.5-23.2), followed by the estimates from approach 4 adding RWD at 75% of trial follow-up completion (mean PFS 21.3 months; 95% CI 18.1-24.5). The approach 3 method with RWD added at the end of trial had the highest estimates for mean PFS at 21.8 months (95% CI 18.8-25.2). The approach 3 hybrid method with RWD added at the end of trial had the highest 5-year PFS rate at 12.0% (95% CI 7.2-17.2) compared with the other 2 hybrid methods. Visual inspection suggested that there might be discrepancy between the parametric model using log-normal distribution and the Kaplan-Meier survival curves for pooled and end-of-trial methods, but not for 75% hybrid (Fig. 2). In particular, the number of patients at risk dropped significantly for the end-of-trial method, which may partially explain the observed discrepancy.

In the sensitivity analysis, we applied the same parametric model selection process using the RWD patients only and found that log-normal distribution was also selected as the "best-fit" model for the RWD cohort. In addition, we did not observe a statistically significant difference of PFS between the trial and RWD patients under log-normal baseline hazards (P=.74). Using 30 years as a proxy for a lifetime time horizon, we found that extrapolation from trial alone resulted in a slightly higher estimate of the mean lifetime PFS (27.8 months) than extrapolation from

Table 2. Baseline characteristics of patients in real-world cohort and MONARCH-3 clinical trial.

Baseline characteristics	RWD (I/E aligned) (n = 207)	RWD (weighted) (n = 118.33)	MONARCH-3 control (n = 165)
Age (years, by trial median age), %			
≤63	24	50	50
>63	76	50	50
Race, %			
White	75	62	62
Other	25	38	38
PR status, %			
Positive	76	77	77
Negative	24	23	22
Site of metastasis, %			
Bone only	43	24	24
Visceral	23	35	76
Number of metastasis, %			
1	57	29	28
2	19	26	26
≥3	9	18	46
AI, %			
Letrozole	46	79	79
Anastrozole	54	21	21

AI indicates aromatase inhibitor; I/E, inclusion/exclusion; PR, progesterone receptor; RWD, real-world data.

RWD only (25.7 months). A similar pattern was observed among hybrid methods, where the pooled methods had the lowest estimate (26.8 months) compared with end-of-trial and 75% hybrid with the narrowest 95% CI (see Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013).

Discussion

In this study, we demonstrated the potential for pooling information from clinical trial patient cohorts with RWD cohorts to enhance survival extrapolation with longer follow-up time. The clinical depth, real-world endpoints, and longitudinality of EHR-derived RWD facilitated matching and weighting of patient populations in the 3 hybrid approaches illustrated here. Among the enhanced survival extrapolation methods explored, we found that outcomes were similar between the clinical trial and selected RWD cohort with rigorous application of patient eligibility criteria, weighting, and choice of time window. Among the 3 hybrid methods, the pooled method with full follow-up time from both cohorts provided a weighted 5-year mean PFS in between the trial and RWD cohort-specific estimates, with a narrower 95% CIs from larger sample size.

The variations in PFS estimates were because of different contributions from short-term and long-term survivors and underlying assumptions. Although both end-of-trial and 75% completion methods only assume that long-term survival after a certain cutoff point follows the same distribution, regardless of short-term survival outcomes, the pooled method assumes that both short-term and long-term survival outcomes among trial and RWD patients follow the same distributions of hazards. There is also a trade-off between sample size and conservativeness of estimations dependent on the absolute and relative number of RWD patients included in the analysis with trial patients. As was shown in our analysis, including all RWD patients in the pooled analysis

yielded a larger sample size and provided an estimation with the narrowest CI compared with other hybrid methods. Moreover, qualitative evaluation such as visual inspection can be helpful in evaluating different approaches and their assumptions. The Kaplan-Meier curve showed a larger dip for the end-of-trial method. This dip was possibly due to the small number of patients at risk toward the end of the trial, which could have an influential impact on survival extrapolation. This observed discrepancy between the Kaplan-Meier curve and the parametric model suggested that the end-of-trial method (Fig. 2C) is relatively an unpreferred approach given the unstable estimates toward the end of trial. Other factors, such as the completeness of trial survival outcomes and length of the trial compared with the natural history of the disease in the patient population, may also be considered for comparing different approaches.

We selected a RWD cohort that closely matched the clinical trial eligibility criteria, weighted by baseline characteristics, and restricted to a similar treatment initiation time window. It is possible that matching to the trial I/E criteria is not required when the objective is not emulating the clinical trial or the purpose is intended to be more generalizable. The details of selecting a real-world comparator group that is relevant to the research question have been previously discussed.[5] In our exploratory analysis, we found that all 3 RWD cohorts with less restricted selection criteria had larger sample size (see Appendix Table 5 and Appendix Figure 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013). Relaxed RWD cohorts also had longer median rwPFS than the primary RWD cohort as expected, because patients in the RWD cohorts were more likely to be white and less likely to have >1 site of metastasis (Table 3). Given that the focus of this study is to demonstrate the opportunity to enhance trial survival extrapolation with external RWD, we selected the restricted RWD cohort that had the most similar survival patterns, with the trade-off of a smaller sample size and shorter follow-up. In addition, we only demonstrated the selection of standard parametric models using trial data in the primary analysis. Nevertheless, it is possible to reselect the "best-fit" model using information from both trial and RWD cohorts. In our exploratory analysis of refitting the standard parametric models, survival curves of log-normal distribution and AIC/BIC showed that it fitted the combined data well in both pooled and 75% methods, but not in "end-of-trial" approach (see Appendix Fig. 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013). In the visual exploration of the hazard functions from all cohorts, we have found that both RWD only and all combined cohorts met the hazards assumption of log-normal distribution (increases over time from 0 to reach a maximum and then decreases monotonically), whereas there was an upward increase at around month 20 in the trial cohort, possibly due to small number of patients in the risk set (see Appendix Fig. 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013). Our further exploration using flexible spline-based models (1-5 knots under either proportional hazards or proportional odds assumption) did not yield better fitted hazards or survival curves than log-normal distribution (see Appendix Figs. 5 and 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.08.013).[17,18] These exploratory findings suggest visual inspection and additional expert clinical opinions might be needed.

To the best of our knowledge, this is the first study illustrating different approaches incorporating external real-world EHR-derived data to enhance survival extrapolation from a clinical trial to reduce uncertainty. By using high-quality EHR data, we were able to select RWD patients who were comparable with trial patients based on demographic and clinical characteristics. We further balanced the baseline characteristics between the trial and RWD patient cohorts using propensity score weighting. Clinical

Table 3. Five-year PFS and survival rates, estimated by standard trial extrapolation and RWD-enhanced methods.

Baseline characteristics	Approach to extrapolation				
		Traditional	RWD-enhanced		
	0 RWD long-term observation	1 Trial data extrapolated estimate	2 Trial data with RWD pooled from the start	3 Trial data with RWD bolted onto the end	4 Trial date with RWD added at 75% trial completion
Total number of patients	207	165	371	217	234
Effective number of patients	118	165	207	186	197
Mean PFS, months	20.4 (16.9-23.8)	21.3 (18.2-24.9)	20.8 (18.5-23.2)	21.8 (18.8-25.2)	21.3 (18.1-24.5)
Mean PFS rate at 5-years	11.0% (6.1-19.8)	11.0% (6.4-16.5)	10.4% (7.2-13.9)	12.0% (7.2-17.2)	11.1% (7.2-16.4)

Note. Mean survival times were estimated at the end of RWD follow-up: 61 months for mBC.
mBC indicates metastatic breast cancer; PFS, progression-free survival; RWD, real-world data.

trials such as the MONARCH-3 study that only report PFS and have limited follow-up are vulnerable to bias in extrapolating mean survival.[14] Nevertheless, RWD can observe long-term outcomes.

Our sensitivity analysis on overall survival extrapolation using RWD patients with short-term follow-up as a proxy of immature trial data suggested potential bias to overestimate long-term

Figure 2. Survival curves of log-normal distribution extrapolation of progression-free survival among patients with mBC by different methods. (A) Trial-only extrapolation (approach 1). (B) Pooled trial + RWD extrapolation (approach 2). (C) Hybrid trial + RWD (end-of-trial) extrapolation (approach 3). (D) Hybrid trial + RWD (at 24 months of trial, ie, 75% of trial completion) extrapolation (approach 4). Note: gray band represents the 95% confidence interval of progression-free survival estimates using log-normal distribution.
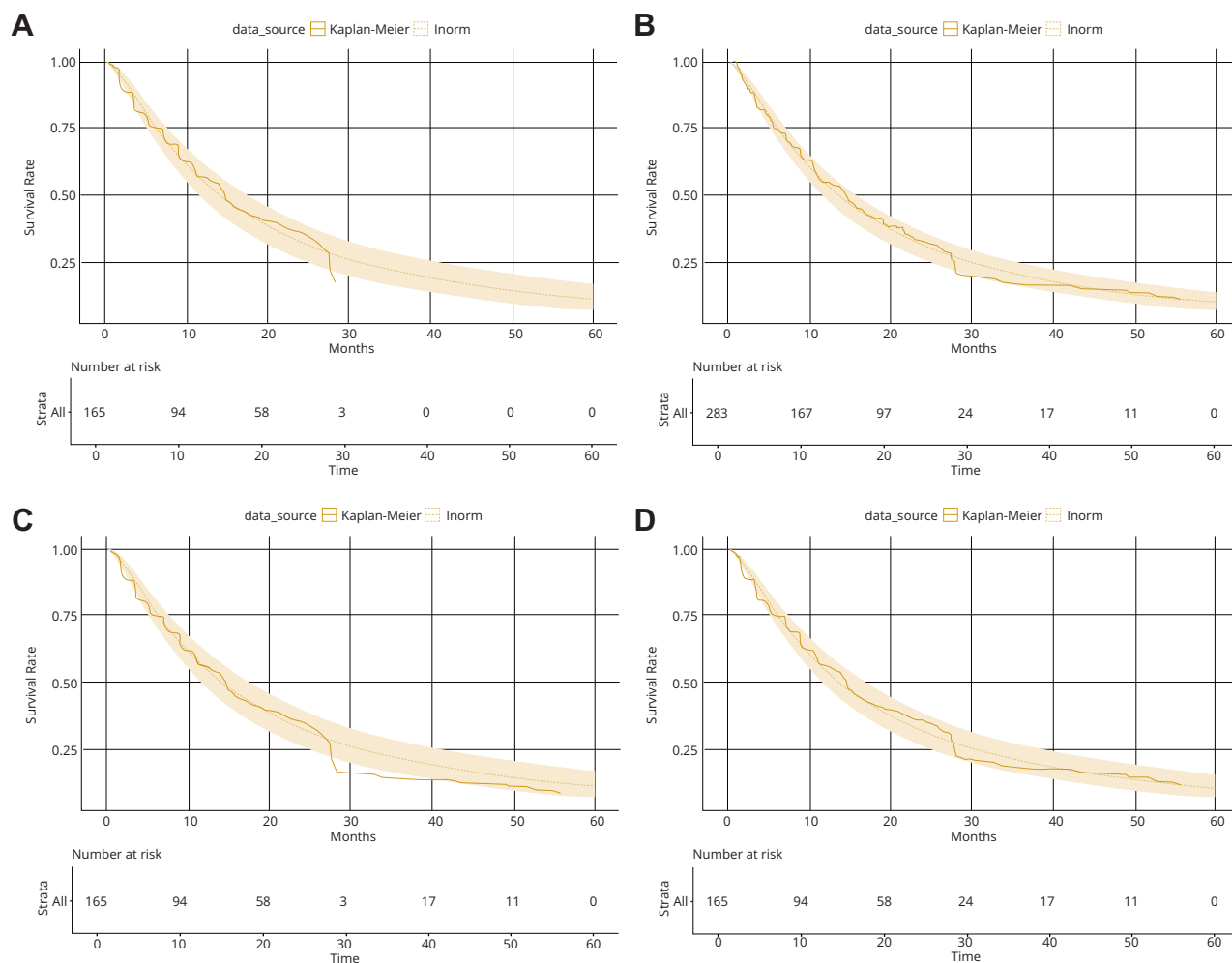
mBC indicates metastatic breast cancer; RWD, real-world data.

overall survival compared with RWD patients with full long-term follow-up. In such circumstances where trial data are immature, it might be more important to incorporate survival information from a comparable RWD cohort with sufficient follow-up to improve long-term survival estimates. Furthermore, the proposed hybrid approaches allow flexibility for adding RWD patients at various time points, with the trade-off between analytical sample size and survival estimation.

There are also limitations. Our case study was limited to the control arm in the MONARCH-3 study and did not estimate the treatment effects between experimental and control treatments. This was due to the consideration that long-term follow-up of a new treatment is less likely to be available among RWD patients at the time of HTA appraisal submission. In this case, future analysis using trial-only extrapolation of experimental treatment and enhanced trial-RWD extrapolation of placebo treatment may be an option to estimate treatment effects. If long-term follow-up of a new treatment is available among RWD patients, proposed hybrid methods can be applied to both treatment arms for treatment effect estimation. In addition, we were not able to look into several aspects because of the lack of individual-level data from the trial patients. We selected the cutoff time point (0%-100% trial completion) in the enhanced hybrid approach arbitrarily, which may have an impact on the survival estimation. If individual-level data of trial patients are available in other use cases, censoring patterns should be considered in choosing a more informed cutoff time point. The trial cohort had a dip in survival probability around month 27. This sudden change in survival could be due to the small number of patients left in the trial (n = 7 at month 21), but patient-level censoring data are needed for further investigation. It is also unclear whether the pivot in the hazards function among trial patients was due to censoring or biological toxicity. Given that it was not observed in the RWD data, we choose to use the log-normal distribution as the best fitted model. More informed model choices, especially if complex models are needed, may be able to be made with more detailed patient-level data, clinical expert opinions, or other external data sources. Moreover, EHR-derived RWD is limited to information documented during the course of routine care, and therefore, it may not capture all relevant information. Real-world treatment and progression were retrospectively captured from EHR using a clinician-anchored abstraction approach. Although previous studies have shown the consistency between the real-world progression variable with progression published in clinical trials,[14,19] there is potential misclassification in both treatment lines and real-world outcome variables. RWD used in our case study is United States based, which may not be generalizable to patient experiences outside of the United States. Nevertheless, it is possible to select a US patient cohort that has similar baseline characteristics and treatment as a non-US patient population. We were only able to compare the extrapolated survival estimates with the observed PFS in the same RWD cohort, given that PFS is not available in commonly used external databases. It will be helpful to evaluate the robustness of the approaches using other external data sources if possible. Finally, the usability of this enhanced approach, including inclusion criteria and characteristic weighting, may vary by specific use cases across different tumor types with different treatment patterns and survival trajectories.

Conclusions

Our case study demonstrated the strengths and limitations of several methods to enhance survival extrapolation by adding additional information from external RWD patients. We have also shown great potential in leveraging RWD with clinical depth and longitudinality. Further EHR-based studies using RWD are needed to confirm our findings and to extend beyond this use case for other cancer types and antineoplastic therapies.

Supplemental Material

Article and Author Information

Author Affiliations: Flatiron Health, Inc., New York, NY, USA (Wang, Adamson, Tan, Bargo, Ghosh, Baxi); The Comparative Health Outcomes, Policy, and Economics Institute, University of Washington, Seattle, WA, USA (Adamson, Ramsey); Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, England, UK (Briggs); Hutchinson Institute for Cancer Outcomes Research, Fred Hutchinson Cancer Research Center, Seattle, WA, USA (Ramsey).

Correspondence: Xiaoliang Wang, PhD, MPH, Flatiron Health, Inc, 233 Spring St, 5th Floor, New York, NY 10013, USA. Email: wendy.wang@flatiron.com

REFERENCES

1. Latimer NR. Survival analysis for economic evaluations alongside clinical trials–extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Mak*. 2013;33(6):743–754.
2. Latimer N. NICE DSU technical support document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. National Institute for Health and Care Excellence Decision Support Unit. http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf. Accessed June 19, 2020.
3. Woods B, Sideris E, Palmer S, Latimer N, Soares M. DSU technical support document 19: Partitioned survival analysis for decision modelling in health care: a critical review. National Institute for Health and Care Excellence Decision Support Unit. http://nicedsu.org.uk/wp-content/uploads/2017/06/Partitioned-Survival-Analysis-final-report.pdf. Accessed June 19, 2020.

4. Johnston S, Martin M, Leo AD, et al. MONARCH 3 final PFS: a randomized study of abemaciclib as initial therapy for advanced breast cancer. *NPJ Breast Cancer*. 2019;5(1):5.

5. Ramsey SD, Adamson BJ, Wang X, et al. Using electronic health record data to identify comparator populations for comparative effectiveness research. *J Med Econ*. 2020;23(12):1618–1622.

6. Abernethy AP, Gippetti J, Parulkar R, Revol C. Use of electronic health record data for quality reporting. *J Oncol Pract*. 2017;13(8):530–534.

7. Ma X, Long L, Moon S, Adamson BJS, Baxi SS. Comparison of population characteristics in real-world clinical oncology databases in the US: flatiron health, SEER, and NPCR. Preprint. Posted online May 30, 2020. medRxiv 2020.03.16.20037143. https://doi.org/10.1101/2020.03.16.20037143.

8. Birnbaum B, Nussbaum N, Seidl-Rathkopf K, et al. Model-assisted cohort selection with bias analysis for generating large-scale cohorts from the EHR for oncology research. Preprint. Posted online January 13, 2020. ArXiv 2001.09765.

9. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55.

10. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. 2017;186(8):1010–1014.

11. Segal BD, Bennette CS. Re: "Transportability of trial results using inverse odds of sampling weights.". *Am J Epidemiol*. 2018;187(12):2716–2717.

12. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health*. 2012;15(6):940–947.

13. Griffith SD, Tucker M, Bowser B, et al. Generating real-world tumor burden endpoints from electronic health record data: comparison of RECIST, radiology-anchored, and clinician-anchored approaches for abstracting real-world progression in non-small cell lung cancer. *Adv Ther*. 2019;36(8):2122–2136.

14. Griffith SD, Miksad RA, Calkins G, et al. Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced Non-small-cell lung cancer data set. *JCO Clin Cancer Inform*. 2019;3(3):1–13.

15. Abiraterone for castration-resistant metastatic prostate cancer previously treated with docetaxel-containing regimen. Technology appraisal guidance. National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/ta259/resources/abiraterone-for-castrationresistant-metastatic-prostate-cancer-previously-treated-with-a-docetaxelcontaining-regimen-pdf-82600499923909. Accessed April 13, 2021.

16. Talimogene laherparepvec for treating unresectable metastatic melanoma. Technology appraisal guidance. National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/ta410/resources/talimogene-laherparepvec-for-treating-unresectable-metastatic-melanoma-pdf-8260459650733. Accessed April 13, 2021.

17. Royston P, Lambert PC. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. College Station, TX: Stata Press; 2011.

18. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002;21(15):2175–2197.

19. Johnson KR, Ringland C, Stokes BJ, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol*. 2006;7(9):741–746.