Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Oncology Drug Effectiveness from Electronic Health Record Data Calibrated Against RCT Evidence: The PARSIFAL Trial Emulation

David Merola[1,2], Jessica Young[2,3], Deborah Schrag[4], Kueiyu Joshua Lin[1,2,5], Nicholas Robert[6], Sebastian Schneeweiss[1,2]

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [2]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, USA; [3]Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Healthcare Institute, Boston, MA, USA; [4]Department of Medicine, Memorial Sloan Kettering Cancer Center, Weill Cornell Medical School New York, New York, NY, USA; [5]Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [6]Ontada, Irving, TX, USA

Correspondence: David Merola, Email davemerola@gmail.com

Background: The use of electronic health records (EHR) data to assess drug effectiveness in clinical oncology practice is of great interest to regulators, clinicians, and payers. However, the utility of EHR data in clinical effectiveness studies may be limited by missing data, unmeasured confounding, and imperfect outcome surveillance. This study sought to emulate and compare the results of a randomized controlled trial investigating the efficacy of palbociclib with fulvestrant vs letrozole in advanced breast cancer.

Methods: This was a cohort study using longitudinal EHR data derived from outpatient oncology practices in the United States. Eligibility criteria from the PARSIFAL trial were emulated as closely as possible. Patients were included if they had hormone-positive, human epidermal growth factor receptor – 2 (HER-2) negative metastatic breast cancer and had no record of prior treatment for metastatic disease. Patients initiating first-line treatment with palbociclib and fulvestrant following their first record of metastasis were compared to those initiating palbociclib and letrozole on the same day. Treatments were ascertained by oncology medication ordering records in the data source. The primary outcome was death as recorded in the oncologists' EHR systems.

Results: There were 1886 eligible women in the study cohort. Although the 3-year survival was meaningfully lower in clinical practice (59%) compared to the randomized trial (78%), the relative effect size was a hazard ratio (HR) of 1.07 (95% CI: 0.86–1.35), similar to the randomized trial (HR = 1.00; 95% CI: 0.68–1.48).

Conclusion: Despite common challenges encountered in EHR-based studies, it is possible to achieve similar conclusions to emulated randomized trials with the application of analytic approaches that address missing data, confounding, and selection bias. This is a promising finding in light of other emulations and ongoing efforts to improve data from clinical practice and causal analytics.

Keywords: real-world evidence, comparative effectiveness research, oncology, pharmacoepidemiology, healthcare databases

Introduction

The use of data collected from routine care in clinical oncology practice[1,2] to establish drug effectiveness has become of great interest to regulators, clinicians, and other healthcare stakeholders.[3–5] As the quality and availability of these data have increased over the past decade, their utility in generating actionable clinical evidence on the effectiveness of medical products, ie, "real-world evidence" or RWE, has become ever more promising.[6–9]

Databases derived from specialized oncology EHR systems[1,2] contain rich information on patients' treatments and health outcomes, critical for successful comparative effectiveness research of oncology medicines.[10] They draw upon several sources of clinical information, including medication and chemotherapy physician ordering systems, physician notes from outpatient oncology encounters, and molecular diagnostics/biomarkers found in EHRs. Despite advancements in EHR data quality, there remain limitations to these data that may hamper their utility in generating RWE. Among

Received: 4 May 2022
Accepted: 20 July 2022
Published: 10 October 2022

these, missing data, an inability to link records to inpatient records, high-quality tumor registries, or healthcare claims, and a lack of information captured across the health care continuum (ie, "out-of-network" encounters), present major challenges for investigators interested in drawing causal conclusions on treatment effects.[10,11] The extent of bias that may result in the context of these limitations and the utility of oncology EHR databases in clinical investigations is unknown.

To clarify whether data collected from routine care (ie, real-world data (RWD)) can be used for effectiveness research, some investigators have taken the approach of calibrating RWE against randomized clinical trials (RCT).[12–14] In our study, we extend this framework to the oncology setting with the use of a specialized oncology EHR database. In a previous study, we successfully emulated the reduction in time-to-next treatment (TTNT) estimate (HR: 0.64; 95% CI: 0.52–0.78) reported in follow-up analysis of the PALOMA-2 trial (NCT01740427) study cohort using oncology EHR data.[15] The generalizable conclusions of this single trial emulation must be limited due to assumptions made regarding missing data, uncertainty in accurately identifying the first-line advanced breast cancer population, and the potential for unmeasured confounding and differential surveillance.

Here, we emulated an alternative randomized clinical trial, the PARSIFAL trial (NCT02491983), which examined the efficacy of letrozole vs fulvestrant when combined with palbociclib for the first-line treatment of advanced breast cancer.[16] Our emulation of the PARSIFAL trial complements our previous emulation, as it compares two dual-therapy regimens with similar indications and effectiveness that began being used in practice around the same time. This mitigates concerns regarding unmeasured confounding, surveillance bias, and differential reasons for missing data that may have been more likely in our PALOMA-2 emulation. Using EHR data from the US Oncology Network, we sought to compare estimates of the relative hazard of all-cause mortality, conditional on confounders, among subjects initiating fulvestrant and palbociclib vs letrozole and palbociclib for first-line treatment of metastatic disease to hazard ratio estimates obtained from the PARSIFAL trial.

Materials and Methods

Data Source

The Ontada *iKnowMed* (iKM) EHR database is a large research database derived from outpatient oncology practices in the US Oncology Network (USON). The USON is comprised of over 400 practice sites and treats over 1,000,000 patients annually. Data in the iKM were drawn from structured fields within electronic health records and include key confounders such as performance status and tumor histology, as well as detailed treatment information (see Data Source Description and eTable 1 in the Supplement). Patients selected into the database had at least one International Classification of Diseases (ICD) diagnosis code indicating breast cancer, had at least 3 months of documented medical history in the database, and were not participants in a clinical trial.

Study Population and Design

Cohort selection criteria were adapted from the PARSIFAL trial (eTable 2). Women at least 18 years old with a diagnosis of metastatic breast cancer and no evidence of prior treatment for metastatic disease were included. Patients with evidence of hormone receptor (HR) negative or human epidermal growth factor receptor-2 (HER-2) positive subtypes of breast cancer were excluded. In contrast to the trial, several eligibility criteria were not emulated due to incomplete capture in the database or limited relevance in RWE studies (eg, safety criteria such as hypersensitivity to study drug or inability to swallow tablets; see details in eTable 2). Follow-up time was initiated on the cohort entry (index) date, which was the day all eligibility criteria were fulfilled, and the treatments of interest were initiated. The follow-up period proceeded until the earliest of the following events: (1) outcome occurrence (all-cause mortality); (2) loss to follow-up, defined by a >90-day period with no treatment, laboratory test result, or vitals recording after last evidence of treatment, or (3) administrative end of data (March 28, 2021).

Treatment Ascertainment

Prescription drug orders by within-network oncologists and associated prescribing dates were fully captured in the data resource and drawn from structured fields in the health record system. The primary exposure of interest was treatment

with palbociclib in combination with fulvestrant, which was compared to palbociclib in combination with letrozole. Treatment groups were operationalized in the database by identifying the occurrence of physician orders with the generic names of interest on the same day. All medications were self-administered, oral therapies.

Outcome Measurement

The primary outcome was overall survival, defined as the time from cohort entry to all-cause mortality. Mortality date was ascertained by provider recording of patients' vital status as "deceased" in a structured field in the health record system. The completeness of mortality data in the database has not been formally assessed.

Baseline Patient Characteristics

Patient demographics (age, geographic region), clinical characteristics (smoking status, BMI, tumor stage, diagnosis date, family history of cancer, Karnofsky/ECOG performance status, site(s) of metastasis, disease-free interval, number of metastatic sites), medication use (anticoagulant use, bone remineralization therapies, antihypertensives, antidepressants, anxiolytics, anti-hyperlipidemics, immunizations, anti-diabetics), and comorbidities (anemia, renal disease, anxiety, arthritis, cardiovascular disease, COPD, diabetes, neutropenia, osteoporosis) were collected to characterize the study cohort and facilitate comparison with the PARSIFAL trial study population. These variables were all ascertained on or before the start date of the treatments of interest.

Missing Data

Missing values were present in five confounding variables: body mass index (BMI) (2%), tumor stage (6%), smoking status (11%), performance status (28%), and number of metastatic sites (63%). Missingness in the data occurred, at least in part, due to changes in reporting standards among oncology practices over time. Based on this information, we assumed that missing data followed a *missing at random* (MAR) mechanism, which permits valid estimation through imputation-based procedures.[17] More specifically, we assumed that the missingness in each variable occurred as a function of practice identifiers, the outcome, treatment, and all confounding variables modeled in our primary analysis, described below.

Statistical Analysis

Given our assumption of MAR and the variety of variable types (eg, ordinal, continuous, etc.) with missingness, multiple imputation with chained equations (MICE) was used to create 50 imputed datasets.[18] Predictive mean matching, ordered logistic regression, and multinomial logistic regression were used to impute continuous, ordinal, and unordered categorical variables with missing data, respectively. All variables modeled in our outcome regression model were also placed in our imputation models, including the outcome and exposure, as well as indicators for practices associated with varying degrees of missingness. Variables were imputed in the order of their relative missingness—from least to most. The functional forms of the models specified for the imputations are shown in eTable 3. Point and interval estimates estimated within each of the 50 imputed datasets were pooled together using Rubin's Rules.[19,20]

In the primary analysis, we estimated the relative hazard of all-cause mortality among patients treated with palbociclib and fulvestrant vs palbociclib and letrozole under the assumption of a Cox proportional hazards model.[21] The model was adjusted for 18 pre-exposure risk factors for death that were potential confounding variables. These variables were chosen for inclusion in the model because they were available in our data source and deemed to be prognosticators of survival (eTable 4).[22] Notably, not all aforementioned patient baseline characteristics could be adjusted for in the analysis because they did not occur in a high enough frequency in the study population and led to problems with convergence of the statistical model. Schoenfeld residual plots were used to check the proportional hazards assumption. Lastly, using the first imputed dataset, a Kaplan–Meier plot was created in the inverse probability of treatment (IP) weighted study population for qualitative comparison to the overall survival curve reported in the PARSIFAL trial. The distribution of IP weights was evaluated by treatment group to check for the presence of any practical positivity violations, which can result in extreme weights.[23]

Non-Randomized Study Vs Randomized Trial Agreement

To assess the compatibility of our study result with the PARSIFAL trial, standardized difference was calculated to compare the log hazard ratios of overall survival from both studies.[24,25] This measure was chosen because it permits assessment of the magnitude and direction of any difference between the two studies' estimates to facilitate interpretation in the presence of any potential sources of bias. A standardized difference of greater than 1.96 was chosen as a marker of incompatibility of the two study results in alliance with convention in regulatory decisions.[25]

Sensitivity Analyses

We assessed the sensitivity of our results to modeling assumptions made in the primary analysis in several ways. First, the primary analysis was repeated among patients that had no missing data, often called *complete case analysis*.[26] This was done to indirectly evaluate our MAR assumption, since analyses of complete cases would likely differ from the multiple imputation-based analysis under the MAR assumption but may be similar if the data follow a *missing completely at random* (MCAR) mechanism. Next, an IP-weighted Cox proportional hazards model was fit to the complete cases. IP-based estimates (marginal treatment effects) make different modeling assumptions than multivariable-adjusted models (conditional treatment effects) with respect to the relationships between the exposure, outcome, and confounders and also generally target different measures.[27] Furthermore, the marginal treatment effect may resemble the randomized trial result, which was not conditional on all of our measured confounders. Therefore, deviations in IP-based estimates and our primary analysis might indicate a sensitivity of our results to these different modeling approaches or may reflect these different analyses are estimating different effect measures.[28] Note that we refer to a hazard ratio as an "effect measure" even though it is not a causal effect except under the null or other restrictive conditions that are not guaranteed even in a randomized trial with no loss to follow-up. We report hazard ratios here to match what was reported in the PARSIFAL trial.[29,30]

In addition to testing our assumptions on modeling and missingness, we conducted a sensitivity analysis to assess the potential influence of data discontinuity on our estimate. Data discontinuity occurs when patients seek out-of-network care that may not be recorded in our data source, which exclusively contains outpatient records from the US Oncology Network. Encounters occurring outside of the iKM system may result in misclassification bias if they entail new diagnoses, treatments, or procedures.[11] To account for this, a previously validated prediction rule for discontinuity (eTable 5) was applied in the one-year period before cohort entry to characterize study patients in terms of their predicted EHR-continuity.[11] Then, the primary analysis was repeated among patients in the top 25th, 50th, and 75th percentile of predicted EHR-continuity.

Lastly, differential surveillance may indicate the presence of differences in the clinical care of patients between treatment groups and the presence of confounding, which may not be directly observed in the data. To explore the possibility of differential surveillance, the mean rate of imaging procedures and office visits per patient-day were calculated by treatment group as a proxy for unmeasured confounding.

Results

Study Population

Following the application of all study eligibility criteria, 1886 patients were selected into the study cohort—462 initiators of palbociclib and fulvestrant and 1424 initiators of palbociclib and letrozole (eTable 2; Table 1). Relative to the PARSIFAL trial, patients in our study tended to be older, have fewer metastatic sites, a less favorable performance status among palbociclib-fulvestrant initiators, and a more favorable performance status among palbociclib-letrozole users. All these patterns were maintained following missing data imputation. Upon cohort entry, patients in the palbociclib-letrozole group had a median time since initial diagnosis with breast cancer of 1.1 years (IQR: 0.1 years – 8.1 years), while palbociclib-fulvestrant initiators had a median of 4.8 years (IQR: 1.4 years – 9.5 years) since initial diagnosis.

Comparison of Overall Survival in Non-Randomized Study vs PARSIFAL Trial

In the primary analysis, the hazard ratio for overall survival was 1.07 (95% CI: 0.86–1.35), which was congruent with the clinical trial result (HR: 1.00; 95% CI: 0.68–1.48) (Table 2). The crude (unadjusted) hazard ratio was 1.24 (95% CI:

Table 1 Patient and Demographic and Clinical Characteristics at Cohort Entry

Characteristic	Emulation Study[a]		Imputed Data[a]		PARSIFAL Trial	
	Palbociclib-Fulvestrant (n = 462)	Palbociclib-Letrozole (n = 1424)	Palbociclib-Fulvestrant (n = 462)	Palbociclib-Letrozole (n = 1424)	Palbociclib-Fulvestrant (n = 243)	Palbociclib-Letrozole (n = 243)
Age[b]						
Median (range) - yr	69 (32–85)	66 (25–85)	69 (32–85)	66 (25–85)	64 (25–88)	62 (35–90)
ECOG performance status or Karnofsky equivalent - no. (%)						
0	178 (52.8)	587 (57.2)	241 (52.2)[c]	789 (55.4)[c]	151 (62.1)	124 (51.0)
1	134 (39.8)	381 (37.1)	187 (40.5)[c]	545 (38.3)[c]	80 (32.9)	107 (44.0)
2	25 (7.4)	59 (5.7)	34 (7.4)[c]	90 (6.3)[c]	12 (4.9)	12 (4.9)
Data missing[d]	125 (27.1)	397 (27.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recurrence type - no. (%)						
Recurrent	384 (83.1)	929 (65.2)	384 (83.1)	929 (65.2)	141 (58.0)	147 (60.5)
De Novo	78 (16.9)	495 (34.8)	78 (16.9)	495 (34.8)	102 (42.0)	96 (39.5)
Disease site - no. (%)						
Visceral	43 (9.3)	275 (19.3)	43 (9.3)	275 (19.3)	115 (47.3)	118 (48.6)
Non-visceral	87 (18.8)	470 (33.0)	87 (18.8)	470 (33.0)	128 (52.7)	125 (51.4)
Unknown	332 (71.9)	679 (47.7)	332 (71.9)	679 (47.7)	0 (0.0)	0 (0.0)
No. of disease sites - no. (%)						
<3	91 (85.0)	500 (84.6)	394 (85.3)[c]	1230 (86.4)[c]	141 (58.0)	133 (51.4)
≥3	16 (15.0)	91 (15.4)	68 (14.7)[c]	194 (13.6)[c]	102 (42.0)	110 (48.6)
Data missing[d]	355 (76.8)	833 (58.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Previous treatment in early setting						
Chemotherapy	65 (14.1)	142 (10.0)	65 (14.1)	142 (10.0)	98 (40.3)	92 (37.9)
Tamoxifen only	30 (6.5)	125 (8.8)	30 (6.5)	125 (8.8)	48 (19.8)	59 (24.3)
Aromatase inhibitors only	129 (27.9)	167 (11.7)	129 (27.9)	167 (11.7)	26 (10.7)	21 (8.6)
Tamoxifen and aromatase inhibitors	54 (11.7)	47 (3.3)	54 (11.7)	47 (3.3)	39 (16.0)	31 (12.8)

Notes: [a]135 (9%) of palbociclib-letrozole initiators and 57 (12%) of palbociclib-letrozole initiators in emulation study had missing or incomplete biomarker data. [b]Ages were not available for subjects ≥85 years to preserve privacy. Calculations assume these subjects are 85 years old. [c]Provided as average values over 50 imputed datasets. Totals may not add to 100% due to rounding. [d]For missing data, percentages are based on total subjects in the treatment arm.

Table 2 Parameter Estimates of Cox Proportional Hazards Model by Method of Data Analysis

	Parameter Estimate	95% Confidence Interval	Standardized Difference[a]
PARSIFAL Trial Result	1.00	(0.68, 1.48)	–
Following Multiple Imputation (Adjusted by Stratification)	1.07	(0.86, 1.35)	0.04
Following Multiple Imputation (Adjusted by IP-Weighting)	1.13	(0.87, 1.48)	0.07
Complete Cases Only (Adjusted by Stratification)	1.56	(0.98, 2.47)	0.20
Complete Cases Only (Adjusted by IP-Weighting)	1.23	(0.73, 2.09)	0.09

Note: [a]Comparing PARSIFAL Trial Result (top row) to real-world evidence analyses (remaining rows).

1.02–1.51), which was aligned with our observation of more negative prognosticators in the palbociclib-fulvestrant arm with respect to performance status, disease recurrence, number of metastatic sites, and age. We observed a substantially higher mortality rate in the RWD study than in the randomized trial. In our study, the 3-year overall survival was 59.5% (95% CI: 55.5–63.7) vs 57.7 (95% CI: 49.6–67.1) in the palbociclib/letrozole and palbociclib/fulvestrant groups, respectively (Table 3). This is compared to a 3-year overall survival of 77.1% (95% CI: 70.2–82.5) and 79.4 (95% CI: 73.1–84.4) in the palbociclib/letrozole and palbociclib/fulvestrant arms of the PARSIFAL trial, respectively (Table 3).

Table 3 Estimates of 3-Year Overall Survival

	Palbociclib + Letrozole	Palbociclib + Fulvestrant
PARSIFAL Trial		
Number of Subjects	243	243
Number of Deaths, n (%)	51 (21.0)	51 (21.0)
Follow-up Time, median days (IQR)	960 (726–1191)	
3-Year Survival Probability (95% CI)[a]	79.4 (73.1–84.4)	77.1 (70.2–82.5)
Real-World Evidence Study		
Number of Subjects	1424	462
Number of Deaths, n (%)	372 (26.1)	136 (29.4)
Follow-up Time, median days (IQR)	511 (231–909)	507 (213–880)
3-Year Survival Probability (95% CI)[a]	59.5 (55.5–63.7)	57.7 (49.6–67.1)

Notes: [a]Kaplan–Meier estimate of survival probability at 3 years in the first imputed dataset, adjusted by inverse probability of treatment weights. Estimates from the remaining imputed datasets were very similar.

Kaplan–Meier survival estimates over the study period were aligned between the randomized[16] and non-randomized studies (Figure 1).

Sensitivity Analyses

Relative to our primary analysis, our complete case analyses had point and interval estimates further from the randomized trial result (Table 2) and did not appear to be aligned regardless of adjustment method (ie, IP-weighting or stratification). However, our imputation-based analysis using IP-weighting for confounding adjustment did agree with the trial result (Table 2). In our analysis, adjusting for data continuity, point estimates grew further from the null with wider confidence intervals as higher levels of restriction by continuity ratio were imposed (Table 4). There was no evidence of surveillance bias in our estimates based on the observed rates of imaging procedures over the study period or

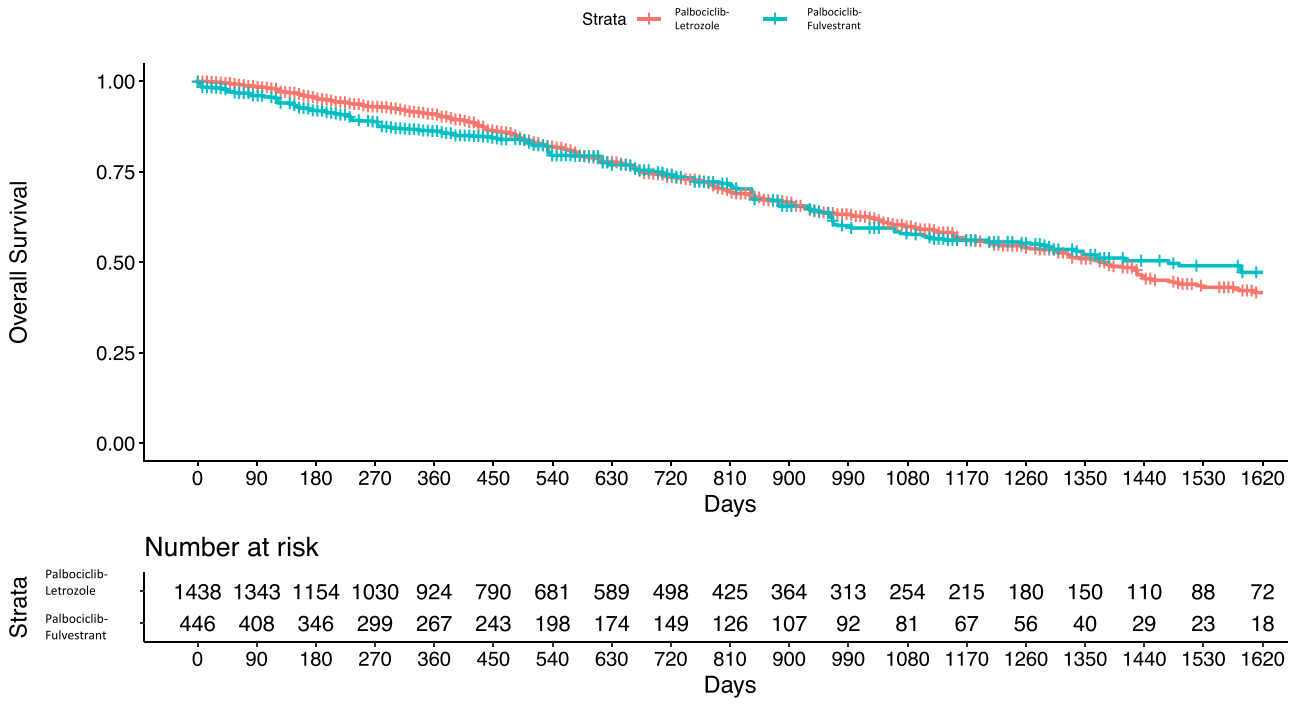

Figure 1 Kaplan–Meier estimates of overall survival by treatment group.

Table 4 Parameter Estimates of Cox Proportional Hazards Model by Varying Levels of Restriction by Continuity Ratio

	Hazard Ratio	95% Confidence Interval	Standardized Difference[a]
PARSIFAL Trial Result	1.00	(0.68, 1.48)	–
Following Multiple Imputation and Restriction to Top 25th Percentile CR	1.35	(0.87, 2.11)	0.14
Following Multiple Imputation and Restriction to Top 50th Percentile CR	1.16	(0.82, 1.65)	0.08
Following Multiple Imputation and Restriction to Top 75th Percentile CR	1.08	(0.85, 1.37)	0.05
Following Multiple Imputation (Not Restricted by CR)	1.07	(0.86, 1.35)	0.04

Notes: All estimates were adjusted by stratification (ie, multivariable adjustment only). [a]Comparing PARSIFAL Trial Result (top row) to real-world evidence analyses (remaining rows).

office visits in each treatment arm (eTable 6). Notably, imaging frequency data were missing in approximately 83% of patients in either study arm.

Discussion

In this non-randomized study comparing the association of palbociclib and fulvestrant to palbociclib and letrozole on overall survival, we found very similar estimates and reached the same clinical conclusion as a recently completed randomized Phase 2 trial. This finding not only lends support to the utility of RWE in examining oncology treatment effectiveness but also confirms the findings of the PARSIFAL Trial. Our study had 3.8 times more patients than the trial leading to more precise estimates and was representative of clinical practice in the US. An array of sensitivity analyses lends support to the validity of our statistical modeling assumptions and the mechanisms of missing data. A major strength of this study was that it compared the outcomes of two regimens that had an equivalent evidence base for their use in the first-line advanced breast cancer setting. Consequently, provider prescribing preference is thought to be a stronger deciding force of treatment choice than patient characteristics and disease severity, resembling random treatment allocation of randomized trials. This can reduce the potential for unmeasured confounding and differential treatment of patients by design, which is supported by the non-differential rate of imaging procedures observed between the treatment groups. Despite an element of randomness being potentially introduced to treatment selection in routine practice, it should be noted that some providers might prefer to initiate treatments that their patients have not previously failed. Particularly, patients receiving fulvestrant/palbociclib may have been more likely to have progressed while receiving adjuvant aromatase inhibitor.

This study had a similar result to our previous emulation of the PALOMA-2 trial with fewer indicators of potential bias. Both trial emulations exhibited some differences between the randomized trial and EHR study populations. These differences, however, did not appear to influence the observed estimates, assuming the trial estimate was unbiased. One reason for this could be that baseline characteristics in the RWD are poorly captured in structured fields, leading to apparent discrepancies in study populations when they were in fact similar. Alternatively, it is possible that the patient characteristics that did differ were not strong effect measure modifiers. In fact, there was no evidence of strong effect measure modification in the PARSIFAL trial in any pre-specified subgroups.[16] This context could explain why our results were so similar to the PARSIFAL trial's results.

Our study has several limitations. First, due to the non-randomized nature of our study, unmeasured or residual confounding always remains a possibility, particularly for non-oncology related prognosticators of survival that were poorly captured in our database. We estimated that an unmeasured confounder would have to have an independent association of 1.64 or 0.61 with both the exposure and outcome to explain our null finding if the true hazard ratio for overall survival was 1.36 or 0.85, respectively, beyond the bounds of the 95% confidence interval of the primary analysis results.[31,32] Despite this possibility, we believe that an unmeasured confounder of this magnitude is unlikely, particularly because its association with the outcome and exposure would have to be this strong independent of all other measured confounders. Second, some types of measurement error, such as non-differential exposure misclassification, can result in a bias towards the null. Although this could also explain our observed result, our prior emulation of the PALOMA-2 trial demonstrated a consistent conclusion for a non-null estimate, which strengthens

our confidence in this study's findings. Third, our study only investigated effectiveness measures in the first-line advanced breast cancer population, limiting the generalizability of our study's results to this setting. It is possible that studies investigating different treatment settings and/or outcomes may have a greater sensitivity to confounders that were not measured in this study, for example cardiovascular morbidity, rendering our conclusions less relevant to those contexts. In particular, this may be more important in studies of earlier-stage disease, where cancer may not be the most probable cause of death for patients. Furthermore, our analytic strategy may also not be generalizable to studies that employ alternative data sources, such as unstructured data (eg, physician notes), as the quantity and mechanism of missing information may vary. Lastly, exposures could only be ascertained in the data through physician ordering. Thus, it cannot be confirmed whether patients were actively taking their therapies as medication administration was not captured. Although the extent of exposure misclassification due to this fact is unknown, it may be less extreme in this study due to the life-threatening nature of breast cancer.

Conclusions

Common challenges of using EHR databases for comparative effectiveness research are highlighted by our study. Despite these challenges, we demonstrated that it may be possible to achieve similar conclusions to randomized trials in line with several other emulation projects when we apply analytic tools and rigorous study designs that address missing data, confounding, and selection bias. As more RCT emulation studies with oncology EHR data are becoming available, we will gain confidence in RWE studies in oncology and will be able to differentiate how to conduct such studies and when they will likely be leading to valid findings.

Abbreviations

BMI, Body mass index; EHR, Electronic health record; HER-2, human epidermal growth factor receptor-2; HR, Hormone receptor; iKM, iKnowMed; IP, Inverse probability of treatment; IQR, Interquartile range; MAR, Missing at random; MCAR, Missing completely at random; MICE, Multiple imputation with chained equations; RCT, Randomized controlled trial; TTNT, Time to next treatment; USON, United States Oncology Network.

Data Sharing Statement

The data that support the findings of this study are available from Ontada, a McKesson Corporation business, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Ethics Approval and Informed Consent

This study utilized retrospective data and was therefore deemed exempt from review by the Mass General Brigham Institutional Review Board. The data accessed in this study comply with all relevant data protection and privacy regulations.

Acknowledgment

This work is a component of the following published thesis: Merola, David. 2022. The Use of Oncology Electronic Health Record Databases to Assess the Effectiveness of Breast Cancer Treatment. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences. Available from: https://dash.harvard.edu/handle/1/37371149

Funding

This project was funded by an Innovation in Regulatory Science Award from the Burroughs Wellcome Fund (#1020021).

Disclosure

Dr. Merola owns equity in and is an employee of Aetion Inc., a software manufacturer and reports grants from Burroughs Wellcome Fund. Dr. Robert is a shareholder of McKesson Corporation. Dr. Schneeweiss (ORCID# 0000-0003-2575-467X) is participating in investigator-initiated grants to the Brigham and Women's Hospital from Boehringer Ingelheim unrelated to the topic of this study. He is a consultant to and shareholder of Aetion Inc. His

interests were declared, reviewed, and approved by the Brigham and Women's Hospital in accordance with their institutional compliance policies. Dr Schrag reports personal fees from JAMA, Served as site PI on a clinical trial that was sponsored by GRAIL, personal fees, intellectual property, and travel from Dana-Farber Cancer Institute, personal fees from University of North Carolina, outside the submitted work. The authors report no other conflicts of interest in this work.

References

1. Cartwright TH, Clayton M, Garey JS, Boehm KA. Use of an electronic health record (iKnowMed), in conjunction with evidence-based pathways, for data capture and outcome measurement in colorectal cancer patients treated with first-line therapy in the US Oncology Network. *Am J Clin Oncol*. 2010;28(15_suppl):3626. doi:10.1200/jco.2010.28.15_suppl.3626
2. Parikh RB, Galsky MD, Gyawali B, et al. Trends in checkpoint inhibitor therapy for advanced urothelial cell carcinoma at the end of life: insights from real-world practice. *Oncologist*. 2019;24(6):e397–e399. doi:10.1634/theoncologist.2019-0039
3. Administration FaD. Real-World Evidence; 2021. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence. Accessed September 29, 2022.
4. Califf RM, Robb MA, Bindman AB, et al. Transforming evidence generation to support health and health care decisions. *N Engl J Med*. 2016;375(24):2395–2400. doi:10.1056/NEJMsb1610128
5. Eichler HG, Koenig F, Arlett P, et al. Are novel, nonrandomized analytic methods fit for decision making? The need for prospective, controlled, and transparent validation. *Clin Pharmacol Ther*. 2020;107(4):773–779. doi:10.1002/cpt.1638
6. Brufsky A, Liu X, Li B, McRoy L, Layman RM. Real-world tumor response of palbociclib plus letrozole versus letrozole for metastatic breast cancer in US clinical practice. *Target Oncol*. 2021;16(5):601–611. doi:10.1007/s11523-021-00826-1
7. Carrigan G, Whipple S, Capra WB, et al. Using electronic health records to derive control arms for early phase single-arm lung cancer trials: proof-of-concept in randomized controlled trials. *Clin Pharmacol Ther*. 2020;107(2):369–377. doi:10.1002/cpt.1586
8. Paul SK, Montvida O, Best JH, Gale S, Pethoe-Schramm A, Sarsour K. Effectiveness of biologic and non-biologic antirheumatic drugs on anaemia markers in 153,788 patients with rheumatoid arthritis: new evidence from real-world data. *Semin Arthritis Rheum*. 2018;47(4):478–484. doi:10.1016/j.semarthrit.2017.08.001
9. Tan K, Bryan J, Segal B, et al. Emulating control arms for cancer clinical trials using external cohorts created from electronic health record-derived real-world data. *Clin Pharmacol Ther*. 2021;111:168–178. doi:10.1002/cpt.2351
10. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol*. 2019;16(5):312–325. doi:10.1038/s41571-019-0167-7
11. Lin KJ, Glynn RJ, Singer DE, Murphy SN, Lii J, Schneeweiss S. Out-of-system care and recording of patient characteristics critical for comparative effectiveness research. *Epidemiology*. 2018;29(3):356–363. doi:10.1097/EDE.0000000000000794
12. Jemielita T, Li XN, Piperdi B, Zhou W, Burke T, Chen C. Overall survival with second-line pembrolizumab in patients with non-small-cell lung cancer: randomized phase III clinical trial versus propensity-adjusted real-world data. *JCO Clin Cancer Inform*. 2021;5:56–65. doi:10.1200/CCI.20.00099
13. Jemielita T, Widman L, Fox C, Salomonsson S, Liaw KL, Pettersson A. Replication of oncology randomized trial results using Swedish registry real world-data: a feasibility study. *Clin Pharmacol Ther*. 2021;110(6):1613–1621. doi:10.1002/cpt.2424
14. Soni PD, Hartman HE, Dess RT, et al. Comparison of population-based observational studies with randomized trials in oncology. *J Clin Oncol*. 2019;37(14):1209–1216. doi:10.1200/JCO.18.01074
15. Rugo HS, Finn RS, Dieras V, et al. Palbociclib plus letrozole as first-line therapy in estrogen receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer with extended follow-up. *Breast Cancer Res Treat*. 2019;174(3):719–729. doi:10.1007/s10549-018-05125-4
16. Llombart-Cussac A, Perez-Garcia JM, Bellet M, et al. Fulvestrant-palbociclib vs letrozole-palbociclib as initial therapy for endocrine-sensitive, hormone receptor–positive, ERBB2 -negative advanced breast cancer. *JAMA Oncol*. 2021;7:1791. doi:10.1001/jamaoncol.2021.4301
17. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581–592. doi:10.1093/biomet/63.3.581
18. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20(1):40–49. doi:10.1002/mpr.329
19. Rubin DB. Multiple Imputation after 18+ Years. *J Am Stat Assoc*. 1996;91(434):473–489. doi:10.1080/01621459.1996.10476908
20. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: Wiley-Interscience; 2004.
21. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Methodol*. 1972;34:187–220.
22. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149–1156. doi:10.1093/aje/kwj149
23. Zhu Y, Hubbard RA, Chubak J, Roy J, Mitra N. Core concepts in pharmacoepidemiology: violations of the positivity assumption in the causal analysis of observational data: consequences and statistical approaches. *Pharmacoepidemiol Drug Saf*. 2021;30(11):1471–1485. doi:10.1002/pds.5338
24. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28(25):3083–3107. doi:10.1002/sim.3697
25. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. *Clin Pharmacol Ther*. 2020;107(4):817–826. doi:10.1002/cpt.1633
26. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med*. 2010;29(28):2920–2931. doi:10.1002/sim.3944
27. Schulte PJ, Mascha EJ. Propensity score methods: theory and practice for anesthesia research. *Anesth Analg*. 2018;127(4):1074–1084. doi:10.1213/ANE.0000000000002920

28. Sjolander A, Dahlqwist E, Zetterqvist J. A note on the noncollapsibility of rate differences and rate ratios. *Epidemiology*. 2016;27(3):356–359. doi:10.1097/EDE.0000000000000433
29. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615–625. doi:10.1097/01.ede.0000135174.63482.43
30. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernan MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med*. 2020;39(8):1199–1236. doi:10.1002/sim.8471
31. Mathur MB, Ding P, Riddell CA, VanderWeele TJ. Web site and R package for computing E-values. *Epidemiology*. 2018;29(5):e45–e47. doi:10.1097/EDE.0000000000000864
32. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167(4):268–274. doi:10.7326/M16-2607