



HHS Public Access

Author manuscript

Pharmacoepidemiol Drug Saf. Author manuscript; available in PMC 2023 April 01.

Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2023 April ; 32(4): 426–434. doi:10.1002/pds.5565.

Effectiveness research in oncology with electronic health record data: A retrospective cohort study emulating the PALOMA-2 trial

David Merola^{1,2}, Jessica Young^{2,3,4}, Deborah Schrag⁵, Kueiyu Joshua Lin^{1,6}, Sarah Alwardt⁷, Sebastian Schneeweiss^{1,2}

¹Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

²Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA

³Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA

⁴CAUSALab, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA

⁵Department of Medicine, Memorial Sloan Kettering Cancer Center, Weill Cornell Medical School, New York, New York, USA

⁶Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

⁷Avalere Health, Washington, District of Columbia, USA

Abstract

Purpose: Oncology electronic health record (EHR) databases have increased in quality and availability over the past decade, yet it remains unclear whether these clinical practice data can be used to conduct reliable comparative effectiveness studies. We sought to emulate a clinical trial with EHR data in the advanced breast cancer population and compare our results against the trial.

Methods: This cohort study used EHR data from US oncology practices. All elements of the study were defined to mimic the PALOMA-2 trial as closely as possible. Patients with hormone-positive, HER-2 negative metastatic breast cancer with no prior treatment for metastatic disease were included. Patients initiating palbociclib and letrozole on the same day following the earliest record of metastasis were compared to those initiating letrozole only. The primary

Correspondence: David Merola, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont Street Suite 3030, Boston, MA 02120, USA. davemerola@gmail.com.

AUTHOR CONTRIBUTIONS

David Merola, Sebastian Schneeweiss, Deborah Schrag, Jessica Young: Conception or design of the work. **David Merola:** Collection and assembly of data. **David Merola:** Data analysis and interpretation. **David Merola:** Drafting the article. **Sebastian Schneeweiss, Jessica Young, Deborah Schrag, Kueiyu Joshua Lin, Sarah Alwardt:** Critical revision of the article. **All authors:** Final approval of the version to be published.

ETHICS STATEMENT

This study was exempt from institutional review board approval as the study was retrospective, non-randomized, and used anonymized data.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

associational measure was the conditional hazard ratio for time-to-next treatment (TTNT). TTNT is well-measured in our data source and amenable for calibration against the randomized study results of the PALOMA-2 trial. We used multiple imputation for several patient characteristics with missing values.

Results: There were 3836 study-eligible women with advanced breast cancer. The hazard ratio for TTNT in the observational study (HR: 0.62; 95% CI: 0.56–0.68) was closely aligned with that of the randomized trial (HR: 0.64; 95% CI: 0.52–0.78).

Conclusions: Under our assumptions on missing data and comparability of the two study populations, results from our non-randomized study closely matched that of the randomized trial. Further studies are needed to determine whether EHR data can yield reliable conclusions on treatment effects in oncology.

Plain Language Summary

Data collected from routine clinical practice have increased in quality and availability over the past decade, yet it remains unclear whether these data can be used to reliably study drug effectiveness. We sought to emulate a randomized clinical trial with such data in the advanced breast cancer population and compare our results against the trial. All eligibility criteria, treatments, and outcome variables were defined to mimic the PALOMA-2 trial as closely as possible. There were 3836 study-eligible women with advanced breast cancer. The results of the observational study were closely aligned with that of the randomized clinical trial. Additional emulations of randomized trials are needed in oncology to gain predictable confidence in when and how treatment effects of oncology products can be studied with electronic health record databases.

Keywords

comparative effectiveness; electronic health records; healthcare databases; metastatic breast cancer; oncology; real-world evidence

1 | INTRODUCTION

Legislative and technological changes over the past decade have given rise to the use of healthcare databases (e.g., administrative claims, electronic health records) in clinical research.^{1,2} Traditionally, the utility of these databases in the context of oncology has been limited due to their poor capture of key clinical characteristics, such as tumor stage, histology, and performance status. However, new specialized electronic health record (EHR) databases^{3,4} containing these prognosticators and other important clinical information are rapidly emerging. Despite recent improvements in data quality, specialized oncology EHR databases have limitations including missing values and incomplete capture of encounters across the healthcare continuum. Consequently, the utility of these data in conducting comparative effectiveness research has yet to be elucidated.

One approach to establishing whether specialized oncology EHR databases can be used for drug effectiveness research is to calibrate database studies against randomized clinical trials.^{5,6} If a thoughtfully analyzed observational study's result is congruent with that of the interventional trial, assuming closely emulated treatment, outcome, and eligibility

criteria, then such a finding would support the validity of using the database to carry out effectiveness studies in that particular setting. In the advanced breast cancer setting, some investigators have taken a similar approach. For example, Bartlett and colleagues compared outcomes between the control arm of the PALOMA-2 trial and a similar cohort of advanced breast cancer patients receiving the same treatment (i.e., first-line letrozole) in routine clinical practice.⁷ Their results showed similarity between specially curated outcomes (i.e., real-world progression-free survival [rwPFS] and response rate [rwRR]) and analogous outcomes reported in the trial. Although this is a promising finding, the study employed a single database and had a relatively small study population, limiting the generalizability of conclusions to other data sources and endpoints.

In the present study, we aimed to build upon this work by using a different EHR database to estimate time-to-next treatment (TTNT), an exploratory endpoint reported in a follow-up study of the PALOMA-2 trial (NCT01740427) participants.⁸ We report TTNT as the conditional relative hazard as well as cumulative risk of adding or switching to a second line therapy or all-cause mortality among patients initiating palbociclib and letrozole versus letrozole only. The PALOMA-2 trial was a Phase III study that examined the efficacy of palbociclib in combination with letrozole versus letrozole and placebo for the first-line treatment of estrogen receptor positive (ER+), human epidermal growth factor receptor 2 negative (HER2-) advanced breast cancer.^{8,9} The primary efficacy endpoint of the trial, investigator-assessed progression-free survival, was not measurable in our data source due to a lack of imaging data. Consequently, we chose to estimate the association between treatment and TTNT, which is well-captured in our EHR data source.

2 | METHODS

2.1 | Data sources

This study utilized data from the McKesson *iKnowMed*SM (iKM) EHR database, which is derived from outpatient medical records of over a 100 community oncology practices in the US Oncology Network from January 1, 2004 through March 28, 2021. The data were drawn from various fields in the EHR and compiled into 11 structured tables for analysis. Patient-level information on demographics, biomarkers, diagnoses, treatments, vitals, metastasis, laboratory results, and other key confounders are included in the database (Table S1).

2.2 | Study population and follow-up

Eligibility criteria for our study were adapted from the PALOMA-2 trial. Within the iKM database, women at least 18 years old diagnosed with metastatic breast cancer between January 1, 2005 and March 28, 2021 and no evidence of prior treatment for metastatic disease were included. To evaluate the first-line advanced disease setting, cohort entry was defined by the first date in which palbociclib or letrozole was ordered following an initial record of metastatic disease. Patients were excluded if they had a record of systemic breast cancer treatment between the first metastasis record and initiation of palbociclib or letrozole. Patients with evidence of ER- or HER-2+ subtypes of breast cancer were excluded, while patients with confirmed HR+/HER-2- disease or missing biomarker data were included in order to reduce the chance of a small sample size and insufficient power to detect an effect

size similar to that observed in the PALOMA-2 trial. Biomarker data was drawn from prior to the date of treatment initiation, and for patients with multiple biomarker test results, the result closest in time to the date of treatment initiation was used. Other eligibility criteria are listed alongside the PALOMA-2 trial criteria in Table S2.⁹ Follow-up began on the day after cohort entry and continued until the first of the following events: (1) outcome occurrence (i.e., addition of a second line therapy or death due to any cause), (2) loss to follow-up, defined by a 90-day period following the last treatment with no evidence of treatment, a laboratory test result, or vitals recording, or (3) administrative end of data (March 28, 2021).

2.3 | Treatment ascertainment

Treatment exposures were ascertained by identifying generic names of prescription drug orders by within-network oncology providers, which were fully captured in the database. When patients were prescribed dual therapy with palbociclib and letrozole, the orders were recorded on the same day. Therefore, following the first record of metastasis, patients with incident orders for palbociclib and letrozole on the same day were compared to those with incident order(s) of letrozole only.

2.4 | All-cause mortality

Mortality 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.

2.5 | Subsequent treatment measurement

The date of the first order for a systemic anti-cancer therapy that was not the primary treatment regimen following the index date was deemed the "subsequent treatment," and used to define the TTNT outcome described below.

2.6 | Time-to-next treatment (outcome) measurement

TTNT was defined as a composite outcome of all-cause mortality or initiation of a subsequent treatment. TTNT was emulated because it was well-observed in the database and approximated the hazard ratio for progression-free survival in an analysis of the trial data with extended follow-up.⁸ In the metastatic breast cancer setting, there are several efficacious treatment choices available following failure of a first-line therapy, which further supports TTNT as a reasonable proxy for disease progression and treatment efficacy.^{10,11} As with all clinical endpoints, TTNT has limitations. Particularly, extreme cases of treatment success and treatment failure may both contribute to long periods prior to initiation of subsequent lines of therapy. Furthermore, we implicitly assume that the reasons for initiating subsequent lines of therapy in our study match those observed in the randomized trial.

2.7 | Baseline patient characteristics

Patient demographics (age, geographic region), clinical characteristics (smoking status, body mass index (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, chronic obstructive pulmonary disease, diabetes, neutropenia, osteoporosis) were collected to characterize the study cohort, adjust for confounding, and/or facilitate comparison with the PALOMA-2 trial study population. These variables were all ascertained on or prior to the date of treatment start.

2.8 | Missing data

Five key confounding variables had missing values, including BMI (2%), tumor stage (13%), smoking status (17%), performance status (27%), and number of metastatic sites (63%). The missing values were believed to be due to changes in EHR reporting standards that occurred among practices participating in the Oncology Care Model, which could be indirectly observed in the data through a practice identifier variable.^{12,13} In particular, physician entry of key variables, such as smoking status and ECOG performance status, into structured portions of the EHR became required following entry into the OCM. Therefore, we assumed that these variables followed a *missing at random* (MAR) mechanism and, particularly, that missingness was a function of practice identifiers, the exposure, the outcome, and all confounders adjusted for in the analysis.

2.9 | Statistical analysis

Multiple imputation with chained equations (MICE) was used to impute missing values since this method is suitable to address data that are MAR¹⁴ and permits imputation of ordinal, nominal, and continuous variables.¹⁴ The functional forms of the models specified for the imputations are shown in Table S3. All variables included in the outcome regression model were also included in the imputation models, in addition to predictors of missingness to reduce bias.^{15,16} Predictive mean matching was used to estimate values of BMI, while ordered logistic and multinomial logistic models were used to estimate missing values of ordinal and nominal variables, respectively. These models were used to generate 50 imputed datasets, which were analyzed individually using the methods described below. Variables were imputed in the order of their degree of missingness (from least to most). To account for the uncertainty in estimates due to missingness, all 50 point and interval estimates were pooled using Rubin's Rules.^{17,18}

For the primary analysis, a multivariable Cox proportional hazards model was used to calculate the relative hazard of TTNT among patients initiating palbociclib and letrozole versus letrozole alone, conditional on measured baseline confounders. The model was adjusted for 18 confounding variables believed to be prognostic of the outcome (Table S4). All these variables were measured on or before the date of treatment initiation. The proportional hazards assumption was checked graphically with Schoenfeld residual plots. Lastly, using the first imputed dataset, a Kaplan–Meier plot was created in the inverse probability (IP) of treatment weighted study population for qualitative comparison to the event-free survival curve produced in the PALOMA-2 trial. The distribution of IP weights in the study population was examined by treatment group to identify extreme weights which suggest positivity violations.¹⁹

2.10 | Non-randomized study versus randomized trial agreement

We qualitatively assessed the magnitude and direction of any difference between the two studies' point and interval estimates of TTNT in the context of any potential sources of bias. The standardized difference between the log hazard ratio of TTNT from our emulation study with that reported in the PALOMA-2 follow-up study was used because it provides a measure of magnitude and direction of any deviation between the two studies, facilitating interpretation of the results.^{6,20}

2.11 | Sensitivity analysis I: Approach to missing data and conditional versus marginal hazard ratios

To assess the robustness of our outcome model assumptions in the primary analysis, we conducted sensitivity analyses. First, only complete cases were analyzed in the same manner as the primary analysis. Then, a Cox proportional hazards model weighted by IP weights was used to estimate the marginal hazard ratio of TTNT in the complete cases only and "imputed" study populations. Analysis of complete cases only offers a way of gaining insight regarding our assumption of the missing data mechanism. In particular, the complete case analysis is expected to differ from the imputation-based analysis under the MAR assumption but may be similar if the data follow a *missing completely at random* (MCAR) mechanism. Additionally, the marginal hazard ratio, calculated with IP weights, was hypothesized to align more with the randomized trial result since the estimate produced by the trial investigators was not conditional on the confounders in this study and non-collapsibility of the hazard ratio.²¹

2.12 | Sensitivity analysis II: Data discontinuity

Since EHR databases typically only contain information from a particular healthcare network, patients' seeking out-of-network care may lead to misclassification bias.²² We addressed this by employing a published prediction rule to identify patients with high data-continuity in health records and restrict the study population to these patients with higher data completeness.²³ Therefore, in an exploratory analysis, we repeated our primary analysis among patients within the 25th, 50th, and 75th percentile of predicted EHR-continuity calculated during the 365 days prior to cohort entry (Table S5). The continuity calculation used in this study was developed previously using an oncology cohort derived from a linked claims-EHR database.

2.13 | Sensitivity analysis III: Surveillance bias

Outcome assessment among patients in the PALOMA-2 trial occurred every 3 months after randomization. However, in the emulation study it is possible that patients were surveilled at different rates among the treatment arms. This may lead to bias by allowing more opportunity for patients in one treatment arm to experience the outcome relative to the other. We assessed the potential for surveillance bias by estimating the mean rate of imaging procedures and office visits (proxied by vitals measurements) per patient-day during the follow-up period for each treatment group.

2.14 | Sensitivity analysis IV: Misclassification bias due to missing/incomplete biomarker data

Approximately 29% of patients receiving letrozole alone and 10% of patients receiving palbociclib-letrozole in the final study cohort had missing or incomplete biomarker data. These patients were included in the primary analysis to conserve sample size under the implicit assumption that they had HR+/HER-2- disease. However, it is possible that some or all these patients that received letrozole alone were in fact HER-2+ since letrozole may be used among patients with this subtype, while palbociclib is typically not. Given that HER-2+ disease is associated with a poorer prognosis²⁴ than HER-2-, our assumption may have resulted in a bias away from the null in the primary analysis, favoring the palbociclib-letrozole regimen. Considering this, a sensitivity analysis was carried out by repeating our analyses among patients with complete and confirmed biomarker data.

3 | RESULTS

3.1 | Study cohort selection

Among 246 752 women 18 years or older with a breast cancer diagnosis, 1299 palbociclib-letrozole users and 2537 letrozole only users were study-eligibility (Table S2). The trial population differed substantially from the averaged imputed study populations, with emulation study participants tending to be classified as having newly metastatic disease, a shorter disease-free interval, Stage IV disease at initial diagnosis, and only one site of metastasis to a much greater extent than trial participants (Table 1). Upon cohort entry, patients in the letrozole only group had a median time since initial diagnosis with breast cancer of 1.5 years (IQR: 0.15 years—7.5 years), while palbociclib and letrozole initiators had a median of 0.8 years (IQR: 0.1 years—8.1 years) since initial diagnosis.

3.2 | Primary analysis

The hazard ratio for TTNT estimated in our primary analysis was 0.62 (95% CI: 0.56–0.68), which aligned with the PALOMA-2 trial result of 0.64 (95% CI: 0.52–0.78; Table 2). The unadjusted hazard ratio was closer to the null (HR: 0.71; 95% CI: 0.64–0.78), which was consistent with the greater presence of negative prognostic factors observed in the palbociclib-letrozole arm prior to adjustment. Median event-free survival for TTNT in the emulation study was shorter than in the trial at 23.1 months (95% CI: 20.8–24.7) in the palbociclib-letrozole arm versus 14.2 months (95% CI: 12.8–15.9) in the letrozole only arm after adjustment using IP weights (Table 3, Figure 1).

The relationship between Kaplan–Meier estimates of event-free survival in each treatment arm was similar between the non-randomized (Figure 1) and randomized trial.⁸

3.3 | Sensitivity analyses

The IP weight-based analysis in the imputed data also resembled the primary analysis (Table 2). However, both, the IP weight-based and stratification-based complete case analyses were not in agreement and further from the null than the clinical trial result. When conducting the primary analysis among patients within the top 75th, 50th, and 25th percentiles of CR, our results were not appreciably altered (Table 4).

There was some evidence of differential surveillance, with the mean rate of imaging procedures and office visits much greater in the letrozole only arm (0.027 imaging procedures/patient-day; 0.287 office visits/patient-day) vs. the palbociclib-letrozole arm (0.012 procedures/patient-day; 0.122 office visits/patient-day; Table S6); however, imaging data were missing for most patients. Lastly, in our sensitivity analysis restricting to patients with complete HR+/HER-2- biomarker data ($n = 2972$), all estimates shifted slightly further from the null relative to our main analyses, appearing robust to our assumption regarding missing biomarker data (Table S7).

4 | DISCUSSION

In this clinical trial emulation study, we successfully emulated the TTNT endpoint reported in the PALOMA-2 trial. These findings were largely consistent with prior work by Barlett et al, which demonstrated concordance of disease progression, survival, and response rates observed in comparator groups derived from real-world data vs. the PALOMA-2 trial.⁷ Our results were robust to changes in analytic methods, supporting the soundness of our modeling assumptions. Our study sample was over five times larger than the clinical trial, supporting adequate statistical power to emulate the treatment effect observed in the clinical trial and may allow for the analysis of more subgroups. This study addressed data discontinuity by applying a prediction rule in the baseline period, and assessing the sensitivity of the results within subsets of patients with various levels of predicted continuity. These results suggested that data discontinuity may be less prevalent in patients with advanced malignancy receiving active treatments. This is not surprising, as oncology care is typically integrated within a single network and patients are less likely to seek cancer treatment across different health systems simultaneously. In the complete cases analysis, a different result was observed compared to the primary analysis. This is consistent with data that are MAR, where patients with complete data are systematically different than those with missing values.

Despite the advantages of our study, our confidence in the results of our emulation is tempered by the potential presence of differential surveillance and several assumptions that were made to account for missing values. Our analysis of imaging procedures and office visits revealed a greater than twofold higher rate of surveillance among letrozole only patients. Based on this, we would expect a greater rate of outcome events in the letrozole only arm, resulting in a bias away from the null. However, approximately 7% of study patients had imaging data available and office visits do not directly indicate surveillance for disease progression. Therefore, it is difficult to say whether surveillance bias could explain our results and more reliable markers of surveillance during follow-up are needed.

In addition to assumptions concerning missing data and differential surveillance, it is possible that cancelation of biases, random chance, outcome misclassification, and emulation failures could explain our findings. The clinical trial participants tended to have younger age, fewer Stage IV diagnoses, less favorable performance status, a shorter disease-free interval, and more metastatic sites. Many of these differences are conflicting with respect to prognosis, and their cumulative influence on the outcome is unknown. Furthermore, the reasons underlying therapy change may differ among treating physicians

in the PALOMA-2 trial versus routine practice. For instance, affordability of treatment and insurance coverage may not influence therapeutic decisions in the clinical trial, as treatments are typically provided by study sponsors. Another possible bias could arise from misclassification of mortality in our data. Fewer than five mortality events were recorded in this study, with treatment change predominantly driving our TTNT outcome. If patients experienced mortality in a differential pattern with respect to treatment group, then it is possible that more event-free person-time could be included in that particular group due to our censoring criteria (namely, the 90-day gap rule regarding structured data activity). Lastly, since treatment indication is not directly observed in EHR data, it is possible that patients selected into our study were not consistent with our target population. Our concerns here, however, are at least partially alleviated due to the robustness of our results to the sensitivity analysis of patients with confirmed HR+/HER-2- disease, as well as the low percentage (~7%) of patients excluded for having HER-2+ disease in the original analysis (Table S2). Overall, routine capture of imaging data, biomarkers, survival, and strong prognosticators of disease progression, such as comorbidities, number of metastatic sites, and locations of metastases, could vastly improve confidence in our study's findings and permit investigation of a wider array of clinically relevant outcomes.

To strengthen the use of oncology EHR data for comparative effectiveness research, several directions should be explored further. First, gaining a better understanding of the mechanisms that lead to missing data in EHRs can inform decisions on how to address this issue in data analysis. The source of missingness is not always apparent and sometimes, as in our experience, discussions with the data vendor on data provenance can be quite instructive. In addition, validation studies of key variables can be helpful in quantifying potential bias that may arise from misclassification. The nature and extent of information bias can help contextualize the plausibility of results. Likewise, as non-oncology comorbidities and drugs are often not observed in the EHRs of oncology practices, quantification of the potential magnitude of residual confounding this may introduce can also facilitate the interpretation of oncology-based EHR studies and strengthen confidence in the use of these data.

Under our assumptions regarding missing data and comparability of the two study populations, our non-randomized study finding was similar to that of the randomized trial. Although our study is promising, large-scale emulations of multiple randomized trials are needed in oncology similar to those in other fields^{3,4,16} to gain predictable confidence in when and how treatment effects of oncology products can be studied with EHR databases.²⁵

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

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

CONFLICT OF INTEREST

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 Aetion Inc., a software manufacturer of which he owns equity. His interests were declared, reviewed, and approved by the Brigham and Women's Hospital in accordance with their institutional compliance policies. Dr. Alwardt reports that she is a shareholder of McKesson Corporation. Dr. Merola owns equity in and is an employee of Aetion, Inc.

Funding information

Burroughs Wellcome Fund, Grant/Award Number: #1020021

DATA AVAILABILITY 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.

REFERENCES

1. Bonamici RS. H.R. 34–114th congress (2015–2016): 21st Century Cures Act. 2016 <https://www.congress.gov/bill/114th-congress/house-bill/34>
2. Platt R, Brown JS, Robb M, et al. The FDA sentinel initiative—an evolving National Resource. *N Engl J Med.* 2018;379(22):2091–2093. [PubMed: 30485777]
3. 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. *J Clin Oncol.* 2010;28(15_suppl):3626.
4. 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. [PubMed: 30944183]
5. Franklin JM, Paterno E, Desai RJ, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. *Circulation.* 2021;143(10):1002–1013. [PubMed: 33327727]
6. 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. [PubMed: 31541454]
7. Huang Bartlett C, Mardekian J, Cotter MJ, et al. Concordance of real-world versus conventional progression-free survival from a phase 3 trial of endocrine therapy as first-line treatment for metastatic breast cancer. *PLoS One.* 2020;15(4):e0227256. [PubMed: 32315295]
8. 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. [PubMed: 30632023]
9. Finn RS, Martin M, Rugo HS, et al. Palbociclib and Letrozole in advanced breast cancer. *N Engl J Med.* 2016;375(20):1925–1936. [PubMed: 27959613]
10. Walker B, Boyd M, Aguilar K, et al. Comparisons of real-world time-to-event end points in oncology research. *JCO Clin Cancer Inform.* 2021;5:45–46. [PubMed: 33411622]
11. National Comprehensive Cancer Network. Breast Cancer. Version 4. https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf (Accessed December 2, 2021).
12. Rubin DB. Inference and missing data. *Biometrika.* 1976;63(3): 581–592.
13. Services USCfMM. Oncology Care Model|CMS Innovation Center. <https://innovation.cms.gov/innovation-models/oncology-care> (Accessed December 9, 2021).
14. 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. [PubMed: 21499542]

15. Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behav Res.* 1998;33(4):545–571. [PubMed: 26753828]
16. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393. [PubMed: 19564179]
17. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc.* 1996;91(434):473–489.
18. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* Wiley-Inter-science; 2004.
19. 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. [PubMed: 34375473]
20. 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. [PubMed: 19757444]
21. Sjolander A, Dahlgvist E, Zetterqvist J. A note on the noncollapsibility of rate differences and rate ratios. *Epidemiology.* 2016;27(3):356–359. [PubMed: 26680297]
22. 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. [PubMed: 29283893]
23. Lin KJ, Singer DE, Glynn RJ, Murphy SN, Lii J, Schneeweiss S. Identifying patients with high data completeness to improve validity of comparative effectiveness research in electronic health records data. *Clin Pharmacol Ther.* 2018;103(5):899–905. [PubMed: 28865143]
24. Toikkanen S, Helin H, Isola J, Joensuu H. Prognostic significance of HER-2 oncoprotein expression in breast cancer: a 30-year follow-up. *J Clin Oncol.* 1992;10(7):1044–1048. [PubMed: 1351537]
25. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther.* 2019;105(4):867–877. [PubMed: 30636285]

Key Points

- Oncology electronic health record (EHR) databases have increased in quality and availability over the past decade, yet it remains unclear whether these clinical practice data can be used to conduct reliable comparative effectiveness studies.
- In this non-randomized cohort study modeled after the PALOMA-2 trial, the hazard ratio for time-to-next treatment was closely aligned with that of the randomized trial.
- Large-scale emulations of randomized trials are needed in oncology to gain predictable confidence in when and how treatment effects of oncology products can be studied with EHR databases.

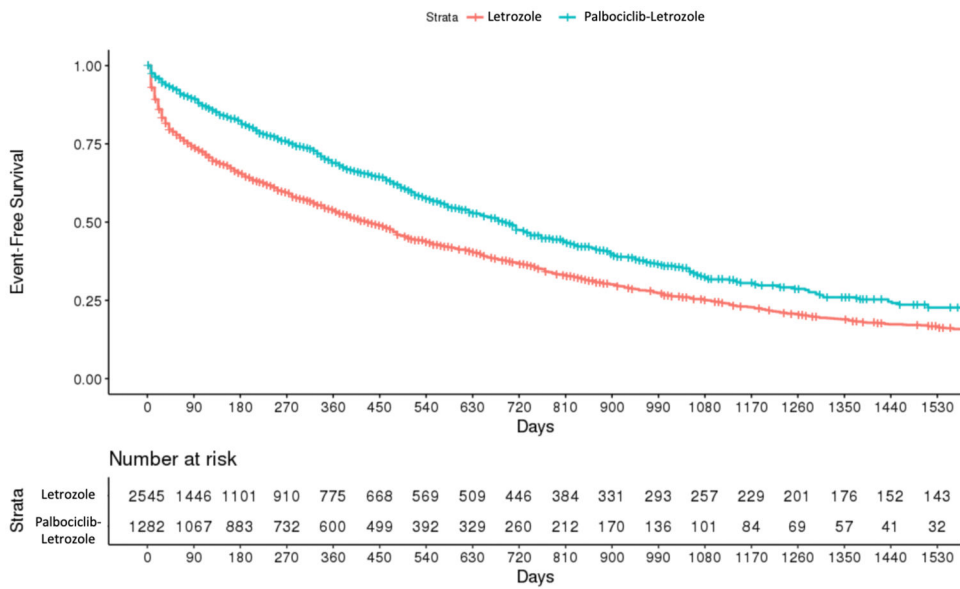


FIGURE 1. Kaplan–Meier estimates of event-free survival. Estimates of event-free survival were computed using the first imputed dataset for computational simplicity. Estimates between the 50 imputed datasets did not appreciably vary

TABLE 1

Patient demographic and clinical characteristics of study cohort

Characteristic	Emulation study (complete cases) ^d		Emulation study (imputed data) ^b		PALOMA-2 trial	
	Palbociclib-Letrozole (n = 1299)	Letrozole only (n = 2537)	Palbociclib-Letrozole (n = 1299)	Letrozole only (n = 2537)	Palbociclib-Letrozole (n = 444)	Letrozole only (n = 222)
Age ^c						
Median (range)—year	66 (25–85)	68 (26–85)	66 (25–85)	68 (26–85)	62 (30–89)	61 (28–88)
<65 year—no. (%)	584 (45.0)	970 (38.2)	584 (45.0)	970 (38.2)	263 (59.2)	141 (63.5)
65 year—no. (%)	715 (55.0)	1567 (61.8)	715 (55.0)	1567 (61.8)	181 (40.8)	81 (36.5)
ECOG performance status or Karnofsky equivalent—no. (%)						
0	521 (56.6)	1219 (64.4)	729 (56.1)	1612 (63.5)	257 (57.9)	102 (45.9)
1	346 (37.6)	577 (30.5)	487 (37.5)	781 (30.8)	178 (40.1)	117 (52.7)
2	54 (5.9)	98 (5.2)	83 (6.4)	144 (5.7)	9 (2.0)	3 (1.4)
Data missing	378 (29.1) ^d	643 (25.3) ^d	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disease stage at initial diagnosis—no. (%)						
I	169 (13.8)	551 (26.2)	181 (13.9)	658 (25.9)	51 (11.5)	30 (13.5)
II	236 (19.3)	469 (22.3)	251 (19.3)	567 (22.3)	137 (30.9)	68 (30.6)
III	121 (9.9)	226 (10.8)	130 (10.0)	276 (10.9)	72 (16.2)	39 (17.6)
IV	695 (56.9)	855 (40.7)	738 (56.8)	1035 (40.8)	138 (31.1)	72 (32.4)
Data missing	78 (6.0) ^d	436 (17.2) ^d	0 (0.0)	0 (0.0)	46 (10.4)	13 (5.9)
Recurrence type—no. (%)						
Distant or other	843 (64.9)	1616 (63.7)	843 (64.9)	1616 (63.7)	305 (68.7)	151 (68.0)
Newly diagnosed	456 (35.1)	921 (36.3)	456 (35.1)	921 (36.3)	139 (31.3)	71 (32.0)
Disease-free interval—no. (%) ^e						
Newly metastatic disease	1002 (77.1)	1972 (77.7)	1002 (77.1)	1972 (77.7)	167 (37.6)	81 (36.5)
12 mo	105 (8.1)	349 (13.8)	105 (8.1)	349 (13.8)	99 (22.3)	48 (21.6)
>12 mo	192 (14.8)	216 (8.5)	192 (14.8)	216 (8.5)	178 (40.1)	93 (41.9)
Disease site—no. (%)						
Visceral	251 (19.3)	376 (14.8)	251 (19.3)	376 (14.8)	214 (48.2)	110 (49.5)

Characteristic	Emulation study (complete cases) ^a		Emulation study (imputed data) ^b		PALOMA-2 trial	
	Palbociclib-Letrozole (n = 1299)	Letrozole only (n = 2537)	Palbociclib-Letrozole (n = 1299)	Letrozole only (n = 2537)	Palbociclib-Letrozole (n = 444)	Letrozole only (n = 222)
Nonvisceral	432 (33.3)	660 (26.0)	432 (33.3)	660 (26.0)	230 (51.8)	112 (50.5)
Unknown	616 (47.4)	1501 (59.2)	616 (47.4)	1501 (59.2)	103 (23.2)	48 (21.6)
No. of disease sites—no. (%)						
1	332 (61.8)	583 (67.4)	1062 (81.8)	2192 (86.4)	138 (31.1)	66 (29.7)
2	120 (22.3)	186 (21.5)	139 (10.7)	231 (9.1)	117 (26.4)	52 (23.4)
3	42 (7.8)	54 (6.2)	55 (4.2)	71 (2.8)	112 (25.2)	61 (27.5)
4	43 (8.0)	42 (4.9)	43 (3.3)	43 (1.7)	77 (17.3)	43 (19.4)
Data missing	762 (58.7) ^d	1672 (65.9) ^d	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

^a 735 (29%) of letrozole only initiators and 129 (10%) of palbociclib-letrozole initiators in emulation study had missing or incomplete biomarker data.

^b Absolute difference in percent between PALOMA-2 Trial and average of 50 imputed datasets.

^c Ages were not available for subjects 85 years to preserve privacy. Calculations assume these subjects are 85 years old.

^d For missing data categories, percentages are based on total subjects in the treatment arm.

^e Defined as the time interval between last cancer treatment received prior to initial metastasis and initial metastasis in the emulation study.

TABLE 2

Parameter estimates of cox proportional hazards model by method of data analysis

	Hazard ratio	95% confidence interval	Standardized difference^a
PALOMA-2 trial result	0.64	(0.52, 0.78)	
Following multiple imputation (adjusted by stratification)	0.62	(0.56, 0.68)	-0.05
Following multiple imputation (adjusted by IP weighting)	0.66	(0.59, 0.73)	0.05
Complete cases only (adjusted by stratification)	0.48	(0.40, 0.58)	-0.40
Complete cases only (adjusted by IP weighting)	0.51	(0.43, 0.62)	-0.31

Abbreviation: IP, inverse probability of treatment.

^aComparing the log hazard ratios of the PALOMA-2 Trial Result (top row) and the real-world evidence analyses (remaining rows).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 3

Median time-to-next treatment

	Palbociclib + Letrozole	Letrozole + Placebo
PALOMA-2 trial, months	28.0 (95% CI: 23.6–29.6)	17.7 (95% CI: 14.3–21.5)
Real-world evidence study ^a , months	23.1 (95% CI: 20.8–24.7)	14.2 (95% CI: 12.8–15.9)

^aMedian times were calculated using the Kaplan–Meier estimator using the first imputed dataset adjusted by IP weights.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

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
PALOMA-2 trial result	0.64	(0.52, 0.78)	-
Following multiple imputation and restriction to top 25th percentile CR	0.63	(0.52, 0.77)	-0.01
Following multiple imputation and restriction to top 50th percentile CR	0.61	(0.54, 0.69)	-0.05
Following multiple imputation and restriction to top 75th percentile CR	0.60	(0.53, 0.68)	-0.07
Following multiple imputation (not restricted by CR)	0.62	(0.56, 0.68)	-0.05

^aComparing PALOMA-2 trial result (top row) to real-world evidence analyses (remaining rows).