10-18-2023

Exploring Breast Cancer Systemic Drug Therapy Patterns in Real-World Data

Julia O'Rourke

Jeff Warnick

John Doole

Luc De Keyser

Zuzanna Drebert

See next page for additional authors

Authors

Julia O'Rourke, Jeff Warnick, John Doole, Luc De Keyser, Zuzanna Drebert, Olivia Wan, Courtney N Thompson, Jack W. London, Karen Fairchild, and Matvey B. Palchuk

Exploring Breast Cancer Systemic Drug Therapy Patterns in Real-World Data

Julia O'Rourke, PhD, MS, MMSc[1]; Jeff Warnick, BS[1]; John Doole, PharmD[1]; Luc De Keyser, MD[1]; Zuzanna Drebert, PhD, MEng[1]; Olivia Wan, BS[1]; Courtney N. Thompson, MSHS, CCRP[1]; Jack W. London, PhD[2]; Karen Fairchild, BS[1]; and Matvey B. Palchuk, MD, MS[1]

DOI https://doi.org/10.1200/CCI.23.00061

ABSTRACT

PURPOSE To explore medications and their administration patterns in real-world patients with breast cancer.

METHODS A retrospective study was performed using TriNetX, a federated network of deidentified, Health Insurance Portability and Accountability Act–compliant data from 21 health care organizations across North America. Patients diagnosed with breast cancer between January 1, 2013, and May 31, 2022, were included. We investigated a rule-based and unsupervised learning algorithm to extract medications and their administration patterns. To group similar administration patterns, we used three features in k-means clustering: total number of administrations, median number of days between administrations, and standard deviation of the days between administrations. We explored the first three lines of therapy for patients classified into six groups on the basis of their stage at diagnosis (early as stages I-III *v* late as stage IV) and the sensitivity of the tumor's receptors to targeted therapies: hormone receptor–positive/human epidermal growth factor 2–negative (HR+/*ERBB2*–), *ERBB2*-positive (*ERBB2*+/HR±), or triple-negative (TN; HR–/*ERBB2*–). To add credence to the derived regimens, we compared them to the National Comprehensive Cancer Network (NCCN): Breast Cancer (version 2.2023) recommendations.

RESULTS In early-stage HR+/*ERBB2*– and TN groups, the most common regimens were (1) cyclophosphamide and docetaxel, administered once every 3 weeks for three to six cycles and (2) cyclophosphamide and doxorubicin, administered once every 2 weeks for four cycles, followed by paclitaxel administered once every week for 12 cycles. In the early-stage *ERBB2*+/HR± group, most patients were administered carboplatin and docetaxel with or without pertuzumab and with trastuzumab (for six or more cycles). Medications most commonly administered in our data set (7,798 patients) agreed with recommendations from the NCCN in terms of medications (regimens), number of administrations (cycles), and days between administrations (cycle length).

CONCLUSION Although there is a general agreement with the NCCN Guidelines, real-world medication data exhibit variability in the medications and their administration patterns.

ACCOMPANYING CONTENT

Data Supplement

Accepted July 7, 2023
Published October 18, 2023

JCO Clin Cancer Inform
7:e2300061

INTRODUCTION

Real-world data (RWD) are data collected during routine health care delivery.[1] RWD include medical data that providers collect in electronic health records (EHRs), such as diagnoses, laboratory tests, prescribed and administered medications, genetic and other tests, medical procedures, pathology, imaging, and so on.

Medication information from RWD is valuable in research, clinical,[2] and regulatory areas.[3-5] For example, during the clinical trial design process, researchers can use medication exposure to investigate the effect of the inclusion and exclusion criteria on the number of patients eligible for the trial. Researchers might use EHR to find patients who meet the study criteria during clinical trial patient recruitment. Generally, clinical uses of RWD include understanding treatment patterns and deviations from the guidelines, identifying and managing adverse events, recognizing unmet needs in disease management, and improving clinical care by understanding the effectiveness and safety of drugs in routine clinical care. RWD is also now used to support

CONTEXT

Key Objective
To explore medications and their administration patterns in real-world patients with breast cancer.

Knowledge Generated
Real-world drug administration patterns for the large cohort of patients with breast cancer.

Relevance
Can help in learning about real-world drug administration and improve clinical trial design.

clinical[2] and regulatory decisions, such as new and supplementary drug approvals, to support label revisions[3,4] and postapproval safety-related monitoring.[5]

Oncology treatment can be complex, consisting of one or more lines of therapy (LOTs), each corresponding to a chemotherapy regimen, a systemic anticancer therapy, or other therapy.[6,7] These therapies can include multiple medications and require specific administration patterns as defined in guidelines. These administration patterns dictate the frequency of medication administration, dosage, days between each administration, and how each medication should be administered with respect to the others (ie, at the same time, in a sequence, etc). Furthermore, recommended therapy regimens vary by each cancer type and can deviate from guidelines in the way they are administered to patients (patient treatments can be affected by a patient's drug tolerance, overall patient health status, and patient behavioral and social determinants). Since RWD sources are a byproduct of the health care processes, EHR systems store medication administrations alongside the administration dates; these data do not readily translate into therapy regimens. Understanding RWD-based regimen information is important, and as RWD often differs from guidelines, deriving it from the data itself without any previous assumptions or filters is especially important. This study aims to develop and describe the regimen and LOT algorithm, on the basis of unsupervised learning approach, and apply to patients with breast cancer, grouped by disease stage and the three most prominent biomarkers.

Patients with breast cancer diagnoses across all stages were selected from 21 US-based health care organizations (HCOs). Treatment guidelines for patients with breast cancer are defined on the basis of the disease stage, biomarkers (ie, estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 expression), disease biology, and physician preferences.[8] The first therapy is generally prescribed based on the initial cancer diagnosis. Subsequent LOTs are prescribed on the basis of the response to the first-line therapy, such as disease recurrence, disease progression, and any side effects. As a result, treatment regimens administered to patients with breast cancer can vary significantly, constructing a viable use case for this therapy regimen extraction algorithm.

METHODS

Data Source

A retrospective study was performed using data from the TriNetX Network, a federated network of deidentified, Health Insurance Portability and Accountability Act–compliant, real-time health data, including diagnoses, procedures, medications, laboratory values, and genomic information curated from multiple EHRs from approximately 30 million patients from 21 HCOs across North America as of May 2022. The primary data sources in this study were the cancer registries and EHR systems. Because this study used only deidentified patient records and did not involve collecting, using, or transmitting individually identifiable data, this study was exempted from institutional review board approval (Fig 1).

Inclusion Criteria

Patients with breast cancer were identified using the International Classification of Diseases, Tenth and Ninth Revision, Clinical Modification (ICD-10-CM and ICD-9-CM, respectively) diagnostic codes (codes ICD-10 C50[9] and ICD-9 174[10]). Only patients with an initial date of diagnosis after 2012 and who had at least one administration of a prespecified oncology medication were included. Patients also had to have a record of their cancer stage and biomarker test. Patients were categorized into three groups on the basis of the sensitivity of their tumor's receptors to targeted therapies, listed in order of specificity: group one consisted of patients whose tumors were positive for human epidermal growth factor 2, irrespective of hormone receptor status (*ERBB2*+/HR±); group 2 included patients with tumors that are HR-positive but *ERBB2*-negative (HR+/*ERBB2*−); and group three encompassed patients with tumors that are triple-negative (TN), HR-negative, and *ERBB2*-negative. We

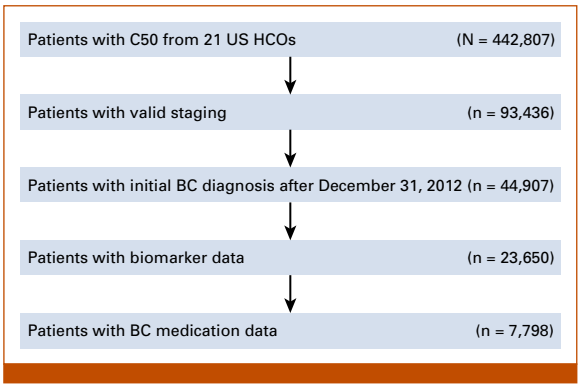

FIG 1. Cohort selection. HCOs, health care organizations.

additionally stratified the analysis on the basis of the stage at diagnosis (early *v* late stage), defining early-stage breast cancer as stages I-III and late-stage breast cancer as stage IV.

Patients with multiple primary cancer diagnoses were removed. All data extraction logic was implemented in the Snowflake database.

Included medications and their procedural codes are listed in the Data Supplement (Table S1). We excluded aromatase inhibitors (AIs) and tamoxifen from the analysis as these drugs are prescribed for extended periods, and data availability is inconsistent across HCOs.

LOT Algorithm

We investigated a method built on a rule-based algorithm and clustering analysis to (1) extract therapy from medication data on 7,798 patients with breast cancer, (2) explore drug administration patterns, and (3) extract and align therapy regimens into LOTs. Several studies have been conducted to determine each regimen and LOT.[11-13] In this study, we used a previously used approach in defining therapy regimens and then extended analyses to explore patterns of drug administration within the regimens using cluster analysis (Fig 2).

Figure 2 outlines a three-step approach used to define each regimen and LOT. We defined a drug administration time period to describe the time interval each medication was administered to a patient (step 1). Drug administration periods provide information on how each medication was administered (ie, cycle frequency, cycle length, cycle length variability, and the overall length of the regimen). K-means clustering[14] using the Euclidian distance algorithm was used to group similar drug administration periods (step 2). Drug administration periods were clustered into 25 clusters using three features: the total number of administrations, the median number of days between administrations, and the standard deviation of the days between administrations. Values greater than the 90th percentile were capped at the value of the 90th percentile. This analysis was performed using R base (v3.6.2) and stats (v3.6.2) packages.

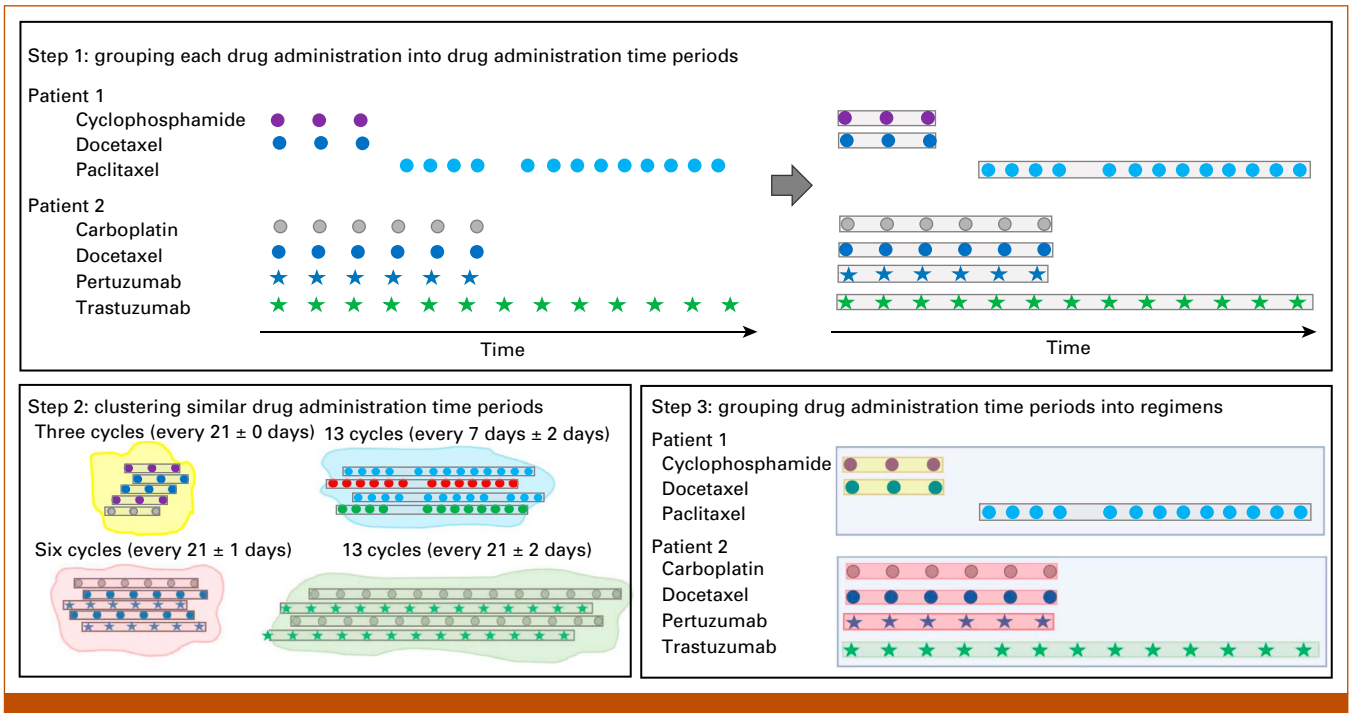

FIG 2. Regimen and LOT algorithm overview. LOT, line of therapy. See the description of Figure 2 in the Data Supplement.

TABLE 1. Demographic and Clinical Characteristics of the Study Cohort

Category	No. of Patients (%)
No. of patients	7,798
Age at diagnosis, median (SD)	57 (12.5)
Female sex	7,738 (99.2)
Race	
White	4,807 (61.6)
Black or African American	2,125 (27.3)
Asian	225 (2.9)
American Indian or Alaska Native	19 (0.2)
Native Hawaiian or other Pacific Islander	11 (0.1)
Unknown	611 (7.8)
Ethnicity	
Not Hispanic or Latino	5,603 (71.9)
Hispanic or Latino	438 (5.6)
Unknown	1,757 (22.5)
Summary stages	
Stage I, count	2,735 (35.1)
HR+/*ERBB2*−	1,488 (54.4)
ERBB2+ (HR±)	762 (27.9)
TN	485 (17.7)
Stage II, count	2,825 (36.2)
HR+/*ERBB2*−	1,536 (54.4)
ERBB2+ (HR±)	669 (23.7)
TN	620 (21.9)
Stage III, count	1,291 (16.6)
HR+/*ERBB2*−	769 (59.6)
ERBB2+ (HR±)	265 (20.5)
TN	257 (19.9)
Stage IV, count	947 (12.1)
HR+/*ERBB2*−	622 (65.7)
ERBB2+ (HR±)	217 (22.9)
TN	108 (11.4)
Stages I, II, and III, count	6,851 (87.9)
HR+/*ERBB2*−	3,793 (55.4)
ERBB2+ (HR±)	1,696 (24.8)
TN	1,362 (19.9)

Abbreviations: *ERBB2*+, human epidermal growth factor 2–positive; *ERBB2*–, human growth factor 2–negative; HR+, hormone receptor–positive; HR–, hormone receptor–negative; SD, standard deviation; TN, triple-negative.

After the clustering analysis, therapy regimens were derived (step 3). Regimens were defined as any group of drug administration periods administered <30 days apart. Medications administered for an extended time (ie, >400 days between the start and end dates) were spliced to isolate shorter regimens (Data Supplement, Fig S1).

In the final step of the analysis, regimens were arranged by the date they were administered and labeled as LOT 1, LOT 2, LOT 3, and so on. The logic for the drug administration periods and grouping these periods into regimens was implemented in the Snowflake database (v7.10.1 Bozeman, MT).

We compared derived regimens with the National Comprehensive Cancer Network (NCCN).[15]

RESULTS

Patients

We started with 442,807 patients from 21 US-based HCOs. We removed patients who had incomplete data, had been diagnosed before January 2013, were not administered any of the medications of interest after January 2013, and had missing stage or biomarker information.

This study included 7,798 patients diagnosed after 2012. The median age at diagnosis was 57 years, 99% of the cohort were females, and 62% were White (Table 1). Eighty-eight percent (88%) of the cohort were diagnosed with either breast cancer stage I, II, or III, and 12% were diagnosed with stage IV. At each disease stage, HR+ patients were the majority.

Cluster Analysis of Drug Administration Periods

Of the 7,798 patients analyzed, 29,835 periods were identified (Data Supplement, Table S2). The cluster analysis revealed that the most common therapy administration pattern was one cycle (29% of all drug administration periods). The second and third most common administration patterns were four cycles once every 2 weeks and three cycles once every 3 weeks, respectively. The top 10 most common clusters accounted for 76% of all drug administration periods.

Determining the LOT and Therapy Administration Patterns

Using the LOT algorithm, we grouped 29,835 drug administration periods into 15,070 regimens and then time-ordered them to create LOTs for each patient. Sixty-two percent (62%) of drug administration time periods were part of the first LOT, 15% were part of the second LOT, 8% were part of the third LOT, and 5% were part of the fourth LOT.

All patients had one LOT since inclusion criteria required that each patient have at least one drug administration. Thirty-eight percent (38%) of patients had at least two LOTs, 19% had at least three LOTs, 11% had at least four LOTs, and 7% had at least five LOTs (Data Supplement, Table S3).

Figures 3-5 (for early-stage breast cancer; see description and table accompanying Figs 3-5 in the Data Supplement Tables S4-S6) and the Data Supplement (Figs S2-S4; for late-stage breast cancer) show the first three LOTs. We selected to tabulate the most common regimens, and within

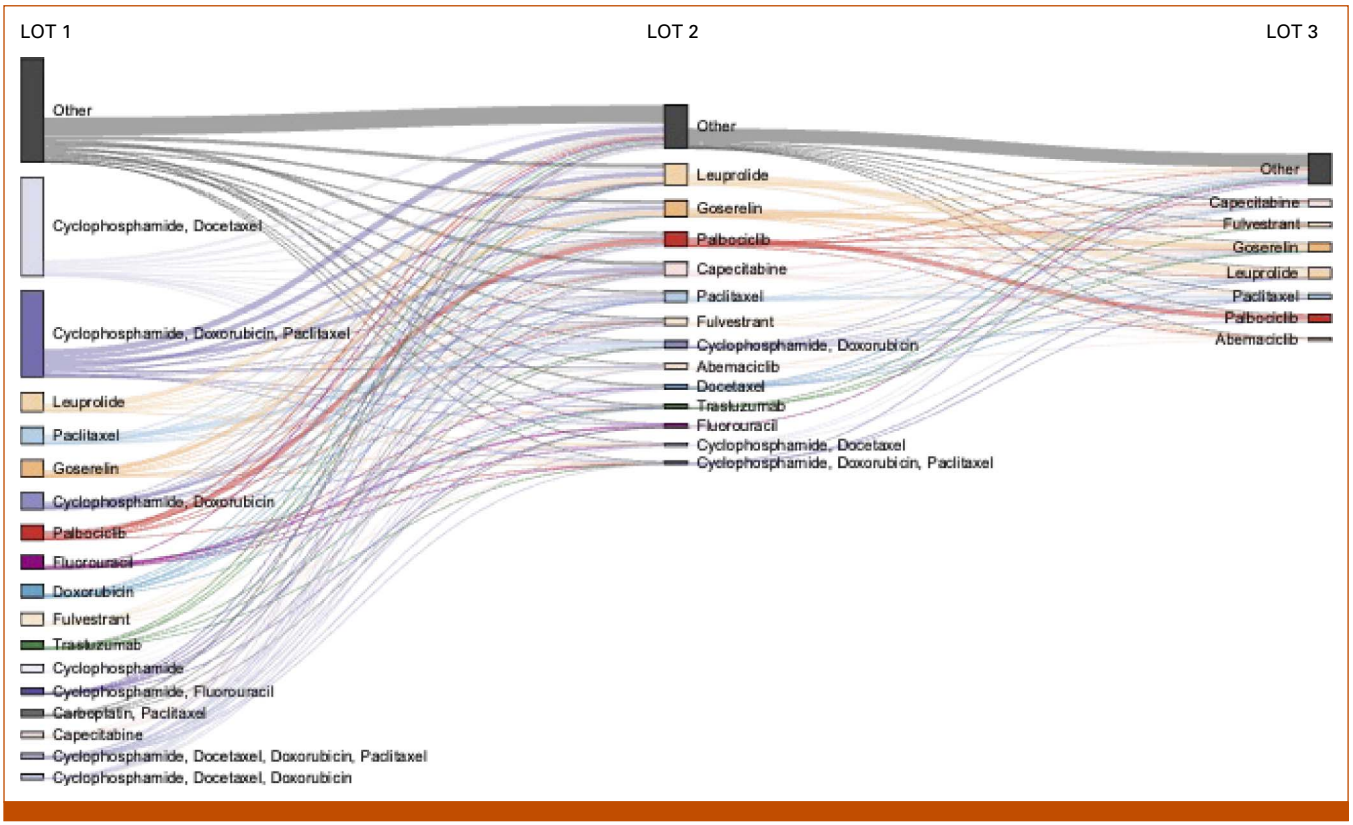

FIG 3. LOT for HR+/*ERBB2*− early-stage patients. *ERBB2*−, human epidermal growth factor 2–negative; HR+, hormone receptor–positive; LOT, line of therapy. See Data Supplement Table S4 for a description and an accompanying table.

each of these regimens, we selected the most common administration patterns.

Treatments for Early-Stage HR+/*ERBB2*− Breast Cancer

Figure 3 shows that two regimens accounted for 41% of patients in LOT 1: (1) cyclophosphamide and docetaxel (22% of patients in LOT 1; administered as three, five, or six cycles once every 3 weeks of each drug) and (2) cyclophosphamide, doxorubicin, and paclitaxel (19% of patients in LOT 1; administered as four cycles once every 2 weeks and paclitaxel for 12 cycles once every week or four cycles once every 2 weeks).

For early-stage *ERBB2*− breast cancer, the preferred regimens in NCCN Guidelines[15] included (1) cyclophosphamide and doxorubicin (dose-dense AC, administered once every 2 weeks for four cycles) followed by paclitaxel (administered once every 2 weeks for four cycles or once every week for 12 cycles) or (2) cyclophosphamide and docetaxel (TC regimen, administered once every 3 weeks for four cycles).

Another preferred treatment listed in the guidelines aligned with our results is monotherapy with capecitabine; however, as this is an oral medication, our results on the number of cycles did not fully correspond with the guidelines (recommended once every 3 weeks for six to eight cycles).[15]

Treatments for Early-Stage TN Breast Cancer

Figure 4 shows that the top three regimens in LOT 1 were similar to the top regimens in HR+/*ERBB2*− early-stage breast cancer and included (1) cyclophosphamide, doxorubicin, and paclitaxel (24% of patients), (2) cyclophosphamide and docetaxel (15% of patients), and (3) paclitaxel (10% of patients). Most patients had one LOT.

Treatments for Early-Stage *ERBB2*+/HR± Breast Cancer

Patients with *ERBB2*+ breast cancer had more variability in the top regimens in LOT 1 than patients with *ERBB2*− breast cancer. The five most common first LOTs included seven regimens, and each included trastuzumab administered for 13 cycles once every 3 weeks (Fig 5). Most patients had one LOT.

These results aligned with the guidelines.[15] The preferred regimens listed in the guidelines are (1) paclitaxel (administered once every once per week for 12 cycles), followed by trastuzumab administered once per week or once every 3 weeks for 1 year, (2) carboplatin and docetaxel administered once every 3 weeks for six cycles, given with and

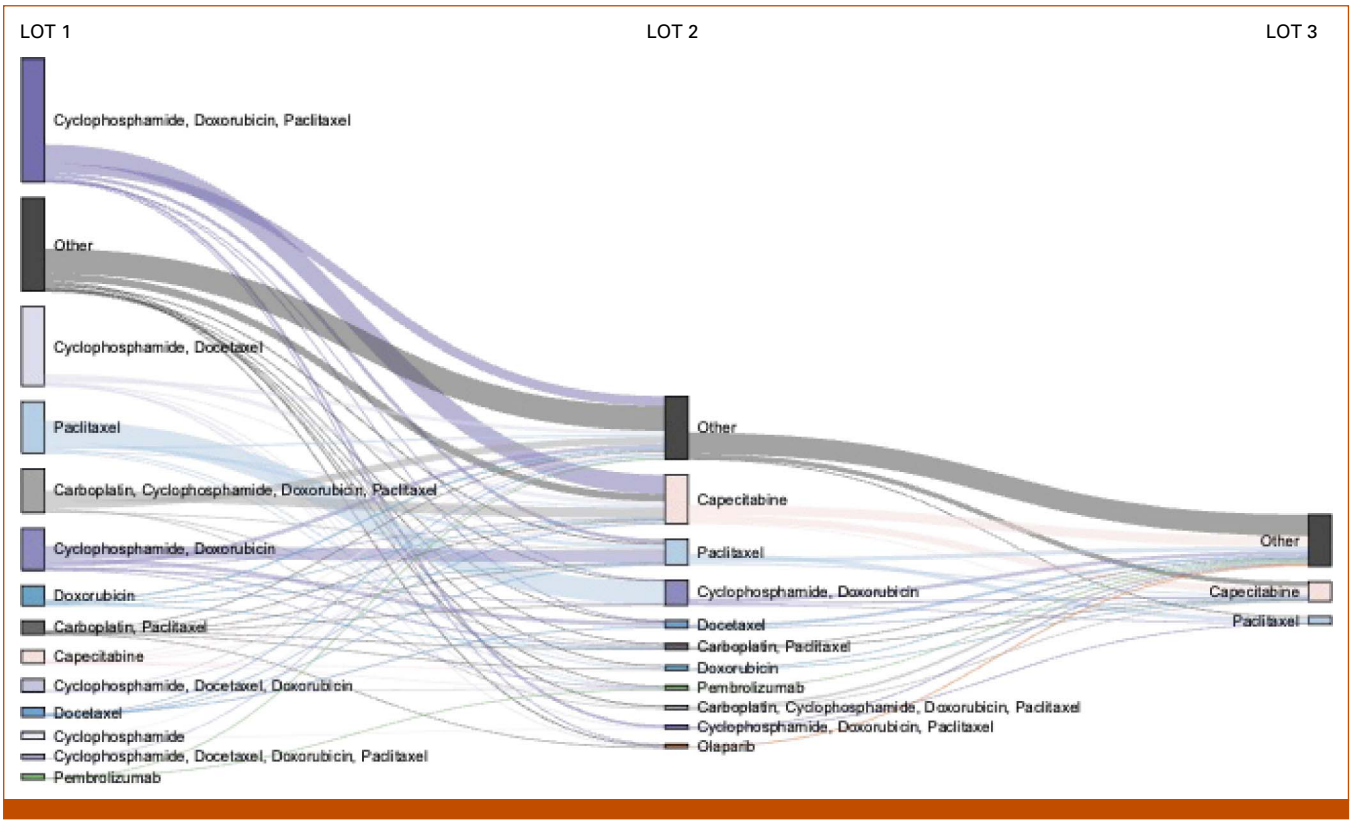

FIG 4. LOT for triple-negative early-stage patients. LOT, line of therapy. See Data Supplement Table S5 for a description and an accompanying table.

followed by trastuzumab administered once every 3 weeks for 1 year (the TCH regimen), and (3) carboplatin with docetaxel administered once every 3 weeks for six cycles given with and followed by trastuzumab and pertuzumab once every 3 weeks for 1 year (the TCH+).[15] We observed that most patients were administered carboplatin and docetaxel with or without pertuzumab and with trastuzumab (for six or more cycles). The paclitaxel with trastuzumab regimen was also common in our data with a cadence similar to the guidelines. Notably, in some situations, our algorithm split the regimens where trastuzumab and/or pertuzumab were given continuously for over 400 days into a separate LOT.

Treatments for Stage IV Breast Cancer

There were 947 (12.1%) stage IV patients in the study. The HR+/*ERBB2*− group had 622 (65.7%) stage IV patients. In this group, palbociclib was the most common medication administered in all three LOTs (32%, 37%, and 34% of patients in LOTs 1, 2, and 3, respectively; Data Supplement, Fig S2).

The TN group had 108 (11.4%) stage IV patients. In this group, paclitaxel was the most common medication in LOT 1 (14% of patients), and capecitabine was common in all three LOTs (9%, 26%, and 13% of patients in LOTs 1, 2, and 3, respectively; Data Supplement, Fig S3).

The *ERBB2*+/HR± group had 217 (22.9%) stage IV patients. Most of the regimens in this group were similar to the regimens in the *ERBB2*+ early-stages group, and most of the regimens included trastuzumab as treatment (Data Supplement, Fig S4).

For late-stage disease, guidelines list multiple varying regimens.[15] Patients with HR+, *ERBB2*− tumors are preferably treated with an AI (eg, anastrozole) or fulvestrant together with a cyclin-dependent kinase 4/6 inhibitor (ie, palbociclib, ribociclib, or abemaciclib). We detected the highest numbers of patients on palbociclib, fulvestrant, capecitabine, or abemaciclib alone. Since we have not included AI in our analysis, we can only speculate that these patients also took an AI orally.

In patients with late-stage *ERBB2*+ breast cancer, our results for the largest group corresponded with the preferred LOT 1 regimens, which include pertuzumab and trastuzumab with docetaxel or paclitaxel cycled once every 3 weeks.[5]

DISCUSSION

Data from 21 US-based HCOs were used to explore therapy administration patterns in patients with breast cancer without any previous assumptions about therapeutic regimens. We started with over 400,000 patients with a breast cancer diagnosis, reduced the cohort to 93,000 patients with

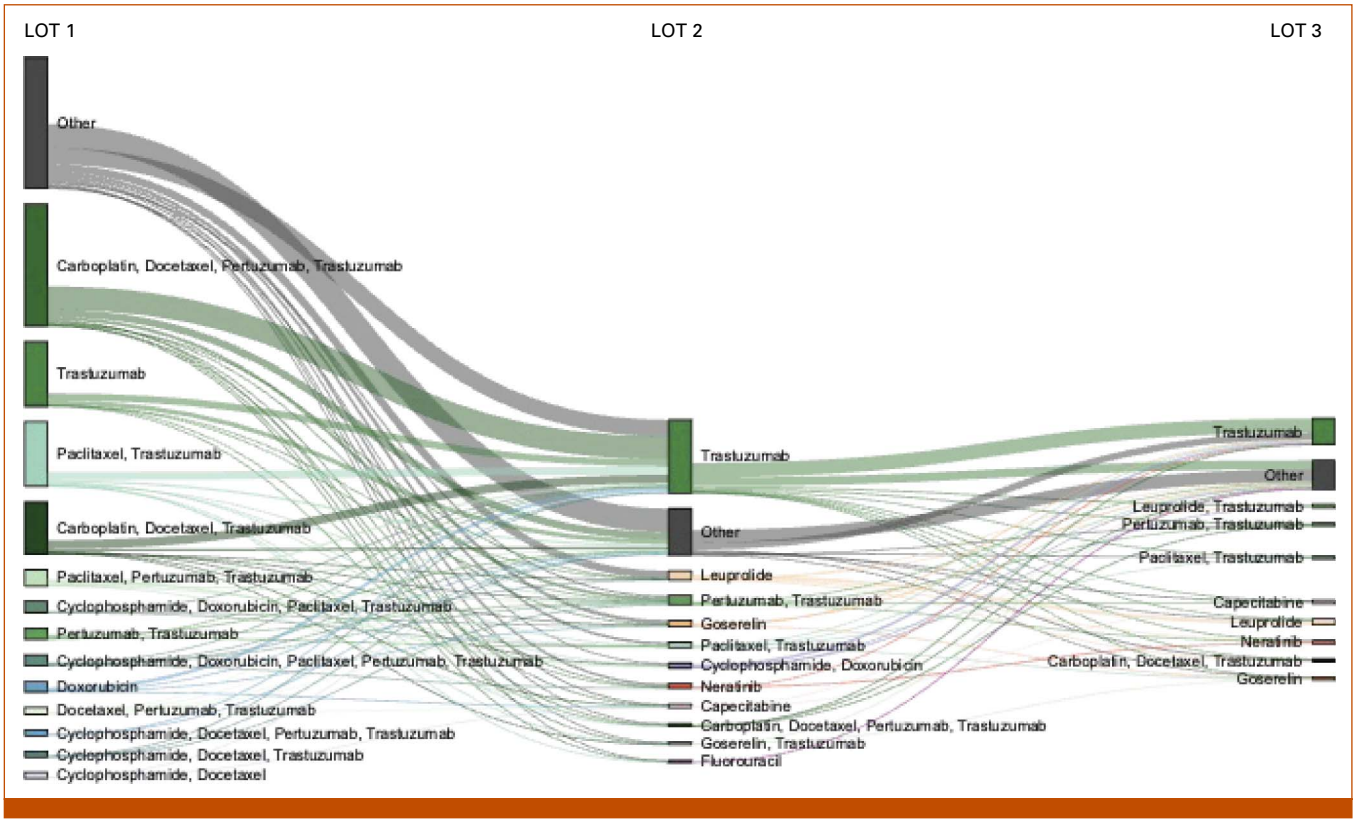

FIG 5. LOT for *ERBB2*+ early-stage patients. *ERBB2*+/HR±, human epidermal growth factor 2–positive; LOT, line of therapy. See Data Supplement Table S6 for a description and an accompanying table.

valid breast cancer data, and narrowed the cohort to 7,800 patients who met the study criteria. Patients in this study were primarily females (99.2%), White (61.6%), and at early-stage diagnosis (87.9% diagnosed at stages 1, 2, or 3). As expected, the prevalence of HR+/*ERBB2*− patients was the highest in this cohort, followed by *ERBB2*+/HR± patients, with TN patients being the least common.[8] This demonstrates that although RWD required a large number of patients to arrive at a cohort that meets the study criteria, the final cohort was representative of the US breast cancer population.

We derived therapy regimen information, including medications and their administration patterns without any previous knowledge of how these medications were recommended to be administered. The NCCN prepared multiple evidence-based guidelines (The NCCN Guidelines) to ensure the quality of cancer care.[7] We compared the most common derived regimens with NCCN Guidelines for Breast Cancer.[15] The derived regimens aligned with these preferred regimens for each patient subgroup in terms of regimen medications, number of cycles, and cycle length.[15]

Although the derived regimens aligned with the NCCN Guidelines, there was variability in each regimen's administration patterns. One of the causes of this variability may have been the source of the medication data. These data were originated from medication orders or administration reports within information systems that support hospital-based pharmacy workflows. In oncology, the difference between what is planned, prescribed, dispensed, or effectively administered can be significant; medication orders could capture just the first prescription for which the administration is to be continued at a different hospital. How drugs are administered (orally *v* intravenous) can affect the way data are recorded in an EHR. In our analysis, oral medications (ie, capecitabine, palbociclib, everolimus, lapatinib, abemaciclib, neratinib, olaparib, alpelisib, and ribociclib) appear as administered in one cycle. Individual oral medication administrations are rarely recorded in EHRs, except when the medications are administered in inpatient settings. Most oral medication data are sourced from the prescription data, recorded only when the prescription is written.

Additionally, orders for some chemotherapies might have changed or even canceled in response to alarming results from laboratory tests performed right before the administration of the treatment. Some sources may also aggregate or even prune administrations or their reporting to what is relevant for refill logistics or direct billing.

In conclusion, this study demonstrated the strength of using structured RWD to analyze large number of patients with significant granularity, including patient characteristics, cancer biomarkers, and details of each medicine (eg, ingredient and its administration date). We used rule-based

algorithms combined with machine learning tools to generate drug treatment insights from RWD, including drug combination, drug administration pattern, and deviation from the recommended guidelines.

Although we found that administration of prescribed regimens was generally consistent with recommended guidelines, there was sufficient variability indicating possible side effects, drug tolerability, or problems with access to care.[7,16] In addition, the uptake of new therapies varies, and RWD can help explore why some treatments are underutilized or used differently from the trial treatment protocols and guidelines. For example, patients with early terminated regimens or patients with more significant variability in cycle length (ie, clusters with high standard deviation) might have clinical (eg, drug tolerance, adverse side effects) or social (eg, access to care) reasons.[17]

There is a strong interest in using RWD from multi-institutional EHRs. Our work demonstrated limitations of RWD: from 442,807 patients, only 7,798 patients met full inclusion criteria. This drop in the count of the eligible patients can be caused by gaps in EHR data as oncology care is complex and can last for years, spanning across multiple institutions. In addition, data extraction and integration process from multiple HCOs into TriNetX network can introduce data loss. Using natural language processing or manual abstraction could potentially overcome these limitations. To improve data completeness and coverage, we shall continue data quality efforts with participating HCOs.

We excluded AIs from the analysis as these medications are prescribed for extended periods of time and data availability is inconsistent. Although our analysis determined gonadotropin-releasing hormone agonists goserelin or leuprolide (recommended cotreatments to AIs and tamoxifen as part of a combination regimen) are used with frequency—we found that 13 percent of patients in the study received these medications—their inclusion without inclusion of AIs in the analysis is a limitation of this work. Expanding this work to include other treatments (such as AIs and tamoxifen) and other groups of patients, such as patients who did not require systemic therapies, could further help in understanding of patient journeys and agreement with recommendations.

AFFILIATIONS

[1]TriNetX, LLC, Cambridge, MA
[2]Department of Cancer Biology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA

CORRESPONDING AUTHOR

Julia O'Rourke, PhD, MS, MMSc, TriNetX, LLC, 125 Cambridgepark Dr, Ste 500, Cambridge, MA 02140; e-mail: Julia.ORourke@trinetx.com.

AUTHOR CONTRIBUTIONS

Conception and design: John Doole, Matvey B. Palchuk
Administrative support: Courtney N. Thompson
Collection and assembly of data: Jeff Warnick, John Doole, Karen Fairchild
Data analysis and interpretation: John Doole, Luc De Keyser, Zuzanna Drebert, Olivia Wan, Courtney N. Thompson, Jack W. London, Karen Fairchild, Matvey B. Palchuk
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.
Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Julia O'Rourke
Employment: TriNetX LLC

Jeff Warnick
Employment: TriNetX
Stock and Other Ownership Interests: TriNetX

Luc De Keyser
Employment: Medpace
Leadership: TriNetX

Zuzanna Drebert
Employment: TriNetX

Courtney N. Thompson
Employment: Avalyn Pharma, Medicago, Leidos Biomedical Research/NCI
Travel, Accommodations, Expenses: Avalyn Pharma

Jack W. London
This author is an Associate Editor for *JCO Clinical Cancer Informatics*. Journal policy recused the author from having any role in the peer review of this manuscript.
Consulting or Advisory Role: TriNetX

Karen Fairchild
Stock and Other Ownership Interests: JNJ

Matvey B. Palchuk
Employment: TriNetX

No other potential conflicts of interest were reported.

REFERENCES

1. Graham S, McDonald L, Wasiak R, et al: Time to really share real-world data? F1000Res 7:1054, 2018
2. Blonde L, Khunti K, Harris SB, et al: Interpretation and impact of real-world clinical data for the practicing clinician. Adv Ther 35:1763-1774, 2018

3. Dang A: Real-world evidence: A primer. Pharmaceut Med 37:25-36, 2023
4. U.S. Food and Drug Administration: Framework for FDA's Real-World Evidence Program, 2018. https://www.fda.gov/media/120060/download
5. Levenson MS: Regulatory-grade clinical trial design using real-world data. Clin Trials 17:377-382, 2020
6. Mestres JA, iMolins AB, Martinez LC, et al: Defining the optimal sequence for the systemic treatment of metastatic breast cancer. Clin Transl Oncol 19:149-161, 2017
7. National Comprehensive Cancer Network: Framework for resource stratification. https://www.nccn.org/global/what-we-do/nccn-framework-for-resource-stratification-of-nccn-guidelines
8. Waks AG, Winer EP: Breast cancer treatment: A review. JAMA 321:288-300, 2019
9. World Health Organization: C50 malignant neoplasm of breast, in International Statistical Classification of Diseases and Related Health Problems (ed 10), 2019. https://icd.who.int/browse10/2019/en#/C50-C50
10. World Health Organization: Basic tabulation list, in International Classification of Diseases: [9th] Ninth Revision, Basic Tabulation List With Alphabetic Index, 1978
11. Meng W, Mosesso KM, Lane KA, et al: An automated line-of-therapy algorithm for adults with metastatic non-small cell lung cancer: Validation study using blinded manual chart review. JMIR Med Inform 9:e29017, 2021
12. Meng W, Ou W, Chandwani S, et al: Temporal phenotyping by mining healthcare data to derive lines of therapy for cancer. J Biomed Inform 100:103335, 2019
13. Jeon H, You SC, Kang SY, et al: Characterizing the anticancer treatment trajectory and pattern in patients receiving chemotherapy for cancer using harmonized observational databases: Retrospective study. JMIR Med Inform 9:e25035, 2021
14. Lloyd S: Least squares quantization in PCM. IEEE Trans Inf Theory 28:129-137, 1982
15. National Comprehensive Cancer Network: Breast cancer (version 2.2023). https://www.nccn.org/guidelines/category_1
16. Rashid N, Koh HA, Baca HC, et al: Clinical impact of chemotherapy-related adverse events in patients with metastatic breast cancer in an integrated health care system. J Manag Care Spec Pharm 21:863-871, 2015
17. Wells JS, Strickland OL, Dalton JA, et al: Adherence to intravenous chemotherapy in African American and white women with early-stage breast cancer. Cancer Nurs 38:89-98, 2015