Adv Ther (2025) 42:2754–2766
https://doi.org/10.1007/s12325-025-03174-y

ORIGINAL RESEARCH

Validation of an Administrative Claims-based Line of Therapy Algorithm for Women with Ovarian Cancer Using Medical Chart Review

Daniel Simmons · John White · Valery Walker · Stephanie V. Blank · Jiefen Munley · Kimmie McLaurin

Received: November 11, 2024 / Accepted: March 10, 2025 / Published online: April 7, 2025

ABSTRACT

***Introduction*:** New maintenance therapies to treat advanced ovarian cancer have added complexity to identifying lines of therapy (LOTs) for real-world evidence (RWE) studies. This study evaluated the performance of a claims-based algorithm that identifies LOTs among patients with ovarian cancer using medical chart review validation.

***Methods*:** The algorithm was developed previously utilizing the Optum Research Database (ORD), a US database that contains administrative claims data. To validate the algorithm, LOT results generated using claims data vs chart data were compared at the patient level by calculating the percent agreement between total number of active and maintenance LOTs, type of therapy (neoadjuvant vs adjuvant classification), and type of regimen (individual drugs). Patients with a diagnosis of ovarian cancer who initiated chemotherapy between December 1, 2014, and September 15, 2017, were included in the study. We report descriptive statistics, the percentage correspondence between medical records and claims data, and kappa statistics to measure the magnitude of agreement.

***Results*:** A total of 294 patients were included in the analysis; 164 received only chemotherapy and no maintenance, 77 received bevacizumab, and 53 patients received poly (ADP-ribose) polymerase inhibitors (PARPi). Mean age was 64.9 years, and 47.3% had stage III cancer. The algorithm demonstrated substantial agreement between claims and medical records for total number of lines of active and maintenance therapy (weighted kappa 0.65 and 0.62 $p<0.0001$). There was moderate-to-substantial agreement for neoadjuvant and adjuvant therapy (kappa 0.56 and 0.62 $p<0.0001$). The algorithm performed best at identifying early treatment with a regimen match of 82% and 88% agreement for first-line active and first-line maintenance, respectively.

Prior presentation: Parts of this work were presented in poster presentations at ISPOR 2022 (Washington, DC; May 15–18, 2022) and at the Society of Gynecologic Oncology Annual Meeting on Womens' Cancer (Phoenix, AZ; March 18–21, 2022).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12325-025-03174-y.

D. Simmons · J. Munley · K. McLaurin
AstraZeneca, Gaithersburg, MD, USA

J. White · V. Walker (✉)
Optum, 1 Optum Circle, Eden Prairie, MN 55344, USA
e-mail: valery.walker@optum.com

S. V. Blank
Icahn School of Medicine at Mount Sinai, New York, NY, USA

Conclusion: We validated an administrative claims-based algorithm that demonstrates strong concordance with medical records for identifying LOT among patients with ovarian cancer. The algorithm can be applied in future studies to analyze treatment patterns and outcomes.

PLAIN LANGUAGE SUMMARY

We created a method to identify treatments prescribed to patients with ovarian cancer in health insurance claims databases, and then reviewed patients' medical records to evaluate the level of accuracy. On a scale from 0 to 1, agreement for the number of treatments was 0.62–0.65, which represents substantial agreement. The name of the medication was also matched 82–88% of the time. This method using claims databases provides accurate information that can be utilized in the future.

Keywords: Algorithm; Ovarian cancer; Validation; Line of therapy; Administrative claims data

Key Summary Points

Why carry out this study?
With the advent of new maintenance therapies to treat advanced ovarian cancer, better methods of identifying lines of therapy (LOTs) for real-world evidence (RWE) studies are needed
The objective of this study was to evaluate the performance of a claims-based algorithm that identifies LOTs among patients with ovarian cancer using medical chart review validation
What was learned from the study?
Substantial agreement between the claims data and medical record dataset was observed for total number of lines of active and maintenance therapy (weighted kappa 0.65 and 0.62, $p<0.0001$)
There was moderate-to-substantial agreement for neoadjuvant and adjuvant therapy (kappa 0.56 and 0.62, $p<0.0001$)
The algorithm performed best at identifying early treatment with a regimen match of 82% and 88% agreement for first-line active and first-line maintenance, respectively

INTRODUCTION

Ovarian cancer is one of the leading causes of cancer mortality among women in developed countries, with an estimated 12,810 attributable US deaths predicted to occur in 2022 [1]. Because of the nonspecific nature and lack of awareness of ovarian cancer symptoms, most patients present with advanced disease at the time of diagnosis [2]. For more than 2 decades, primary management of ovarian cancer has been comprised of cytoreductive surgery accompanied by platinum and taxane-based chemotherapy administered before or after surgery (neoadjuvant and adjuvant chemotherapy, respectively) [3]. Most patients with ovarian cancer respond well to chemotherapy treatment [4]; however, the disease has a high recurrence rate, with up to 80% of patients experiencing relapse within 5 years, following treatment with platinum-based chemotherapy [5]. Recently, treatment options for ovarian cancer have expanded to include targeted therapies such as angiogenesis inhibitors and poly (ADP-ribose) polymerase inhibitors (PARPi), both of which have been

consistently shown to improve progression-free survival in clinical trials [3]. The success of these new drug classes has effected an important shift in the ovarian cancer treatment landscape, either alone or in combination as a front-line maintenance therapy option for many patients [6].

While the introduction of new ovarian cancer therapies has improved patient outcomes, it has also added considerable complexity to the clinical decision-making process. Surgery plus adjuvant or neoadjuvant platinum combination therapy remains the most common front-line treatment for ovarian cancer [7, 8]; however, less is known about treatment patterns for subsequent lines of therapy or how clinical endpoints are affected by different regimen combinations [9, 10]. In light of the rapid pace at which treatment options for ovarian cancer are evolving, data on the use and effectiveness of new regimens have been identified as a research priority [10]. Head-to-head controlled clinical trials are considered the gold standard for evaluating treatment efficacy [11], but real-world evidence may provide additional insights on populations not included in trials especially with the ever-increasing number of possible regimen combinations that may not have been compared in randomized controlled trials (RTCs) [12]. Moreover, results obtained from clinical trials may differ from those seen in real-world practice, given that clinical trial populations are highly selected and tend to be younger, healthier, and more likely to have better outcomes than real-world populations [13, 14].

Because of these limitations, real-world evidence is becoming increasingly important to help bridge evidence gaps that clinical trials are less able to address [15–17]. One commonly used source of real-world data today is administrative claims, which are created to document the interactions of insurance plan members with healthcare systems and typically include information such as patient demographics, dates and sites of service, diagnosis and procedure codes, pharmacy fills, and cost information [17, 18]. Administrative claims offer a rich source of longitudinal health data with large sample sizes; however, an important limitation is that, because they are generated for medical billing purposes and not for research, they often lack detailed clinical information [17]. Regarding oncology research in particular, characteristics such as cancer stage, biomarker testing results, and mainly lines of treatment are difficult to glean from administrative claims alone.

In recent years, administrative claims-based algorithms for identifying ovarian cancer cases have been used to describe patient characteristics and healthcare resource utilization in this population [7, 19, 20], but to date, no validated algorithm has been available for identifying and characterizing the treatment lines to which patients were exposed—a critical tool for meaningful evaluation of the real-world outcomes associated with different regimens. To address this need, the objective of the present study was to evaluate the performance of a claims-based algorithm that we previously developed to identify line of therapy (LOT) among patients with ovarian cancer using medical chart review validation [21]. This algorithm can be applied by researchers and payers to analyze ovarian cancer treatment patterns and patient outcomes using administrative claims data.

METHODS

Study Design and Data Sources

This was a retrospective observational study conducted using administrative claims data linked with abstracted medical chart data to create an individual-level analytic database to support the study objectives.

We previously developed an algorithm [21] utilizing the Optum Research Database (ORD), a deidentified database containing medical and pharmacy claims data with linked enrollment information for individuals enrolled in health plans across the US. It is generally similar to the US insured population regarding age, race, sex, and geographic distribution. The ORD includes data from approximately 8% of US commercial health plan enrollees and 18% of Medicare Advantage enrollees. Medical claims in the ORD include diagnosis and procedure codes from the

International Classification of Diseases, Ninth and Tenth Revisions, Clinical Modification (ICD-9-CM and ICD-10-CM); Current Procedural Terminology or Healthcare Common Procedure Coding System codes; site of service codes; paid amounts; and other information. Pharmacy claims in the ORD include drug name, National Drug Code, dosage form, drug strength, fill date, and financial information for outpatient pharmacy services.

For each patient, one healthcare provider was identified from the administrative claims data and subsequently targeted for medical chart abstraction using a hierarchical process. Specifically, eligible providers were required to be oncologists, with priority given to those with the highest number of medical claims (with a diagnosis of epithelial ovarian, fallopian tube, or primary peritoneal cancer) for the given patient, with a claim date most proximal to the date of the first chemotherapy treatment. Upon procurement, some charts may not have been eligible for abstraction. To be included, charts were required to indicate diagnosis of and treatment for ovarian cancer and to have at least 365 days of data available to abstract. Medical chart data were collected from a convenience sample of patients with providers willing to participate in the chart review process.

The study protocol was reviewed and approved by the New England Institutional Review Board (15 June 2020, #20,201,707). The study obtained a waiver of informed consent and a waiver of authorization under 45 CFR 164.512(i).

Study Population and Cohort Assignment

The study population included adult female commercial and Medicare Advantage health plan enrollees who had a diagnosis of ovarian cancer (at least 2 nondiagnostic medical claims ≥ 30 days apart for epithelial ovarian, fallopian tube, or primary peritoneal cancer [ICD-9-CM 183.0x, 183.2x; ICD-10-CM C56*, C570*]) and initiated a systemic chemotherapy (Supplemental Table S1) during the period of December 1, 2014, through September 15, 2017 (initial identification period, Supplemental Fig. S1). This initial identification period was used to identify initial LOTs for each patient. The date of the first chemotherapy treatment for ovarian cancer during the initial identification period was designated as the index date. Subsequent LOTs were identified during the remainder of the full identification period (September 15, 2017, through November 30, 2019; Supplemental Fig. S1). Included patients were also required to have continuous health plan enrollment for the 6 months prior to the index date and at least 6 months after the index date; patient identifying information available to support medical chart abstraction; and medical charts eligible for abstraction. Patients were excluded if they had any of the following: a claim for pregnancy during the full identification period; a claim for systemic chemotherapy (Supplemental Table S1) during the baseline period; at least 2 non-diagnostic medical claims with a diagnosis for any cancer excluding epithelial ovarian, fallopian tube, primary peritoneal, genitourinary, external genitalia, or intra-abdominal at least 30 days apart during the baseline period; medical chart evidence of any distal cancers that were not a metastasis of the ovarian cancer; or claims or medical chart evidence of clinical trial participation involving an active intervention such as medication or surgery.

Patients were assigned to study cohorts based on claims for ovarian cancer treatment during the full identification period. Patients with a claim for PARPi (olaparib, niraparib, rucaparib) during the full identification period were assigned to the PARPi cohort; these patients may also have had a claim for bevacizumab during the identification period. Patients with a claim for bevacizumab but no claim for PARPi during the full identification period were assigned to the bevacizumab cohort. Finally, patients with no claims for PARPi or bevacizumab during the full identification period were assigned to the chemotherapy-only cohort.

Study Variables

Claims-based study variables were assessed during the 6 months prior to the index date and for at least 6 months post-index (claims baseline

and follow-up periods, respectively; Supplemental Fig. S1). Chart-based study variables were assessed during the 3 months prior to the index date and for at least 12 months post-index (chart baseline and follow-up periods, respectively; Supplemental Fig. S1). For both claims and chart data, the duration of the variable follow-up period was a minimum of 6 months, a mean of 22 months, and a maximum of 33 months post-index. Baseline variables assessed from claims data included age, sex, insurance type, geographic region, baseline Charlson comorbidity index score [22], and comorbid conditions identified using Clinical Classifications Software from the Agency for Healthcare Research and Quality [23]. Cancer stage at index date and cancer history and surgical history during the 3-month baseline period were assessed from chart data. LOTs during follow-up were identified separately for claims and chart data, as described below.

Line of Therapy Identification

The claims algorithm was designed to identify LOTs based on the timing of events, therapies, and treatment gaps in a three-step process: (1) identify discrete treatment episodes of systemic therapy using treatment start dates and stop dates calculated from days' supply; (2) identify neoadjuvant and adjuvant therapy among patients with evidence of surgery; (3) identify active and maintenance treatment. Detailed descriptions of each algorithm step are presented in Fig. 1. After treatment episodes had been identified using the three-step algorithm, all adjuvant and active treatment LOTs were categorized as Active LOT 1, Active LOT 2, etc., for up to five LOTs per patient. Maintenance lines following an active LOT were classified as Maintenance LOT X, with X corresponding to the number of the active LOT.

LOTs from chart data were identified similarly, with the following differences:

- Treatment start and stop dates were identified from medical charts. For dates that had a missing value for day but known month and year, the day was assigned as the 15th of the month.
- Identification of active and maintenance treatment was based on mention of the words "active" and "maintenance" in medical charts to describe treatment status or from other chart information sufficient to deduce treatment status if "active" or "maintenance" was not mentioned.

Statistical Analysis

Baseline patient characteristics were analyzed descriptively and stratified by treatment cohort. To validate the algorithm, LOT results generated using claims data vs chart data were compared at the patient level by calculating the percent agreement between several characteristics, including total number of active and maintenance LOTs, type of therapy (neoadjuvant vs adjuvant classification), and type of regimen (individual drugs). The magnitude of agreement between total LOT results and type of therapy for claims vs charts was assessed using Cohen's kappa statistic (κ) or weighted k (to account for the presence of off-diagonal counts [24]), both of which were interpreted as follows: ≤ 0, no agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, almost perfect agreement. P-values ≤ 0.05 were considered to indicate statistical significance. All analyses were performed using SAS 9.4 (SAS Institute, Cary, NC, USA).

RESULTS

Study Sample

Among 3833 patients in the ORD with an ovarian cancer diagnosis and systemic chemotherapy during the initial identification period, 820 met the remaining study criteria and were eligible for study participation (Supplemental Fig. S2). From these, 451 charts were procured. A total of 157 charts were found to be ineligible for abstraction upon screening (< 365 days of chart data, $n = 81$; no ovarian cancer diagnosis, $n = 34$; no evidence

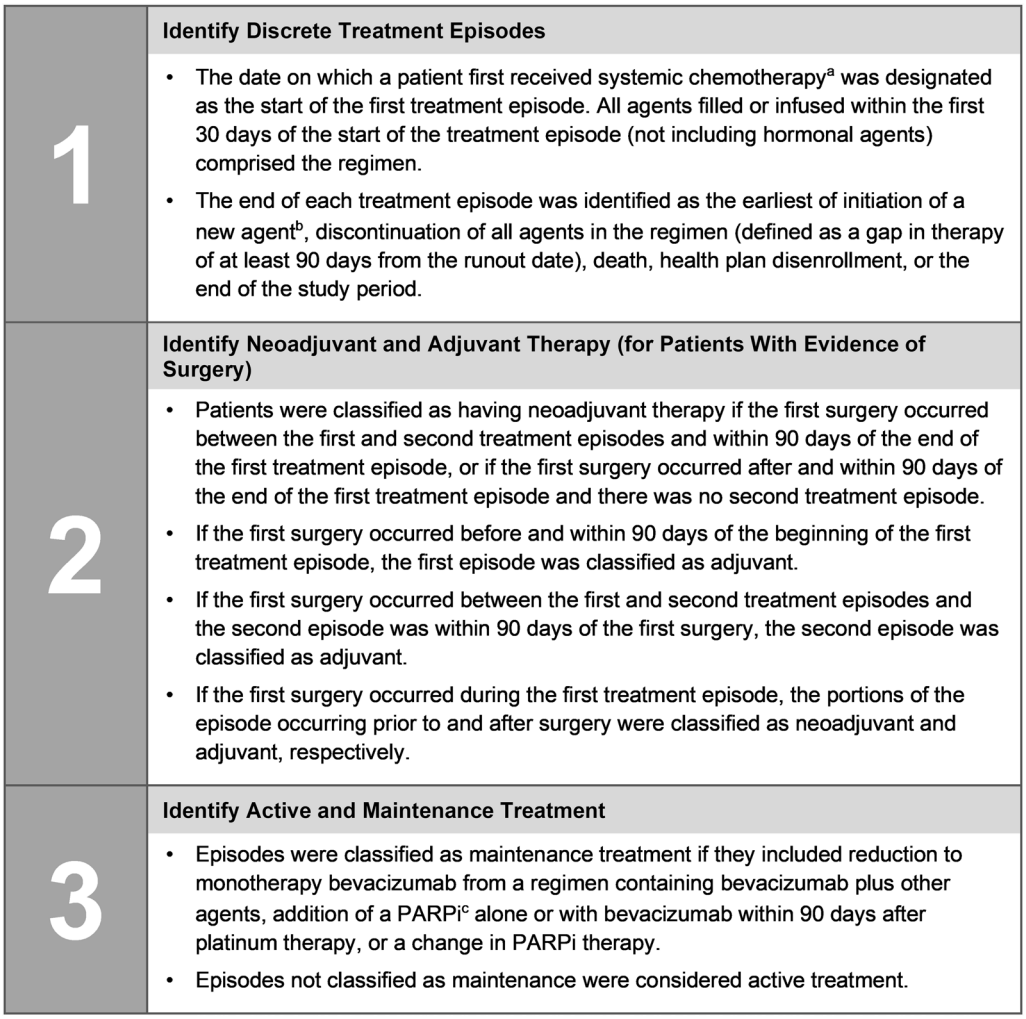

Fig. 1 Three-step line of therapy algorithm. After treatment episodes are identified using the algorithm, adjuvant and active treatment episodes (lines of therapy; LOTs) are categorized as Active LOT 1, Active LOT 2, etc. Maintenance lines following an active LOT are classified as Maintenance LOT X, with X corresponding to the number of the active LOT. [a]Systemic chemotherapies are listed in Supplemental Table S1. Leucovorin was not identified as a chemotherapy, even though it may be given in conjunction with chemotherapy. [b]Switching between platinum agents (cisplatin, oxaliplatin, and carboplatin) or switching between docetaxel and paclitaxel was not considered to constitute the end of a treatment episode. [c]Poly (ADP-ribose) polymerase inhibitor (olaparib, niraparib, or rucaparib)

of ovarian cancer treatment, $n=20$; other reasons, $n=22$). Having no ovarian cancer diagnosis or evidence of ovarian cancer treatment does not mean that the patient did not have these; rather, they were not documented during the observation period available in the chart. In these cases, the patient's chart data were not included in the validation study. The remaining 294 charts were abstracted and constituted the study population (PARPi, $n=53$; bevacizumab, $n=77$; chemotherapy only, $n=164$).

Characteristics among the 294 patients whose charts were abstracted (mean [standard deviation (SD)] age 64.9 [12.3] years) are presented in Table 1. Overall comorbidity burden was high, with 69.7% of patients having a Charlson score of ≥ 5. Common comorbidities included gastrointestinal disorders (77.6%), secondary

Table 1 Baseline patient characteristics by treatment cohort

Characteristic	Total, $N = 294$	PARPi, $n = 53$	Bevacizumab, $n = 77$	Chemo. only, $n = 164$
Age, years, mean (SD)	64.9 (12.3)	66.0 (10.8)	66.9 (9.5)	63.6 (13.7)
Region, *n* (%)				
Northeast	38 (12.9)	5 (9.4)	7 (9.1)	26 (15.9)
Midwest	102 (34.7)	19 (35.9)	30 (39.0)	53 (32.3)
South	129 (43.9)	24 (45.3)	34 (44.2)	71 (43.3)
West	25 (8.5)	5 (9.4)	6 (7.8)	14 (8.5)
Insurance type, *n* (%)				
Commercial	111 (37.8)	14 (26.4)	28 (36.4)	69 (42.1)
Medicare	183 (62.2)	39 (73.6)	49 (63.6)	95 (57.9)
Charlson comorbidity score category, *n* (%)				
0	8 (2.7)	2 (3.8)	1 (1.3)	5 (3.1)
1–2	53 (18)	7 (13.2)	11 (14.3)	35 (21.3)
3–4	28 (9.5)	5 (9.4)	11 (14.3)	12 (7.3)
5 +	205 (69.7)	39 (73.6)	54 (70.1)	112 (68.3)
Top 10 AHRQ comorbidities[a], *n* (%)				
Other gastrointestinal disorders[b]	228 (77.6)	43 (81.1)	48 (62.3)	137 (83.5)
Secondary malignancies	200 (68.0)	42 (79.3)	53 (68.8)	105 (64.0)
Diseases of female genital organs	172 (58.5)	25 (47.2)	31 (40.3)	116 (70.7)
Hypertension	165 (56.1)	27 (50.9)	40 (52.0)	98 (59.8)
Other lower respiratory disease[c]	151 (51.4)	26 (49.1)	41 (53.3)	84 (51.2)
Diseases of the urinary system	151 (51.4)	25 (47.2)	34 (44.2)	92 (56.1)
Diseases of the heart	132 (44.9)	23 (43.4)	36 (46.8)	73 (44.5)
Disorders of lipid metabolism	126 (42.9)	20 (37.7)	33 (42.9)	73 (44.5)
Other nutritional, endocrine, metabolic disorders[d]	122 (41.5)	23 (43.4)	29 (37.7)	70 (42.7)
Other nervous system disorders[e]	108 (36.7)	21 (39.6)	26 (33.8)	61 (37.2)
Cancer stage on index date, *n* (%)				
Stage I	34 (11.6)	0 (0.0)	2 (2.6)	32 (19.5)
Stage II	21 (7.1)	1 (1.9)	0 (0.0)	20 (12.2)
Stage III	139 (47.3)	36 (67.9)	38 (49.4)	65 (39.6)
Stage IV	50 (17.0)	10 (18.9)	20 (26.0)	20 (12.2)

Table 1 continued

Characteristic	Total, $N = 294$	PARPi, $n = 53$	Bevacizumab, $n = 77$	Chemo. only, $n = 164$
Not available	50 (17.0)	6 (11.3)	17 (22.1)	27 (16.5)
Cancer and surgical history during 3-month baseline period, n (%)				
Ovarian cancer–related surgery	135 (45.9)	18 (34.0)	18 (23.4)	99 (60.4)
Recurrent ovarian cancer	61 (20.8)	14 (26.4)	27 (35.1)	20 (12.2)
History of other cancer[f]	56 (19.1)	6 (11.3)	13 (16.9)	37 (22.6)

AHRQ Agency for Healthcare Research and Quality, *chemo.* chemotherapy, *PARPi* poly (ADP-ribose) polymerase inhibitor, *SD* standard deviation

[a]Excluding cancer of ovary and other female genital organs

[b]Includes constipation, dysphagia, and other/unspecified gastrointestinal disorders

[c]Includes postinflammatory pulmonary fibrosis, painful respiration, and other/unspecified lower respiratory disease

[d]Includes disorders of mineral metabolism; obesity; and other/unspecified metabolic, nutritional, and endocrine disorders

[e]Includes disorders of the peripheral nervous system and other central nervous system symptoms/disorders

[f]Includes genitourinary, external genitalia, and intra-abdominal cancers; patients with any other cancer type during the baseline period were previously removed from the analysis

malignancies (68.0%), hypertension (56.1%), other lower respiratory disease (51.4%), and heart disease (44.9%). Nearly half of patients (47.3%) had stage III cancer at the index date, 45.9% had undergone ovarian cancer-related surgery during the 3-month baseline period, 20.8% had recurrent ovarian cancer, and 19.1% had a history of other genitourinary, external genital, or intra-abdominal cancer.

Algorithm Performance

The claims algorithm demonstrated substantial agreement with medical records for identifying total number of LOTs. Percent agreement and weighted kappa statistics for identifying the total number of active and maintenance LOTs are presented in Table 2; weighted kappas for active and maintenance LOTs were 0.65 and 0.62, respectively ($p < 0.001$ for both). The algorithm demonstrated moderate-to-substantial agreement with medical records for identifying the type of therapy. Kappa statistics for identifying neoadjuvant and adjuvant therapy among patients with surgery were 0.56 and 0.62, respectively ($p < 0.001$ for both) (Table 3).

Active and maintenance treatment regimens identified from claims were the same as those identified from charts for 50–86% and 88–100% of patients, respectively, depending on LOT number (Fig. 2). Algorithm performance for regimen match was best for identifying early vs later regimens; for example, percent agreement for active LOTs 1 and 2 was 82% and 86%, respectively, compared with 78% and 75% agreement for active LOTs 3 and 4, respectively, and 50% for LOT 5.

DISCUSSION

Real-world evidence is a valuable tool for evaluating ovarian cancer outcomes in an increasingly complex treatment landscape; however, the utility of administrative claims for this purpose has been limited by the unavailability of certain clinical information in claims data. We developed an algorithm to identify and

Table 2 Algorithm performance: total active and maintenance LOTs identified by both claims algorithm and chart abstraction

Total LOTs identified by claims algorithm	Total LOTs identified by chart abstraction						K[a]	Weighted κ[a]	P-value[b]
Active ($N=293$)[c]	None ($n=17$)	1 ($n=178$)	2 ($n=52$)	3 ($n=31$)	4 ($n=8$)	5 ($n=7$)			
None	**1 (0%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0.57	0.65	< 0.001
1	13 (4%)	**153 (52%)**	8 (3%)	0 (0%)	3 (1%)	1 (0%)			
2	2 (1%)	18 (6%)	**40 (14%)**	5 (2%)	0 (0%)	0 (0%)			
3	1 (0%)	4 (1%)	3 (1%)	**22 (8%)**	3 (1%)	0 (0%)			
4	0 (0%)	2 (1%)	1 (0%)	3 (1%)	**1 (0%)**	2 (1%)			
5	0 (0%)	1 (0%)	0 (0%)	1 (0%)	1 (0%)	**4 (1%)**			
Maintenance ($N=294$)	None ($n=262$)	1 ($n=19$)	2 ($n=11$)	3 ($n=2$)					
None	**247 (84%)**	3 (1%)	2 (1%)	1 (0%)	–	–	0.62	0.62	< 0.001
1	8 (3%)	**13 (4%)**	1 (0%)	0 (0%)	–	–			
2	5 (2%)	3 (1%)	**8 (3%)**	0 (0%)	–	–			
3	2 (1%)	0 (0%)	0 (0%)	**1 (0%)**	–	–			

LOT line of therapy

Data are presented as *n* (%) where the % is calculated from the total number of patients (active $N=293$, maintenance $N=294$). Bolded cells in the diagonal indicate exact matches between claims and chart

[a]The kappa statistic (κ) is interpreted as follows: ≤ 0, no agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, almost perfect agreement

[b]*P*-values test whether the estimated kappa is not due to chance. They do not test the strength of agreement. Also, *p*-values are sensitive to sample size

[c]One patient had > 5 total active LOTs and is not included in this analysis

characterize ovarian cancer LOTs from insurance claims data by identifying discrete treatment episodes, distinguishing neoadjuvant from adjuvant therapy, and categorizing LOTs as active or maintenance treatment. Validation of this algorithm against abstracted medical chart data revealed substantial concordance with medical records for both the total number of LOTs identified and the type of therapy. There was a better match for earlier LOTs, in particular LOT 1. After first-line therapy, there is more heterogeneity in treatments and patients may switch more often. We also had smaller numbers in the later LOTs. For active LOTs 1 through 3—the LOTs that included the highest patient numbers—regimens identified via the claims algorithm matched those identified from medical chart data for 78–86% of patients. Similarly, maintenance treatment regimens identified via claims matched those identified via chart for 88% of patients in LOT 1 and 89% of patients in LOT 2.

While validated algorithms to identify incident or advanced-stage ovarian cancer from administrative claims data have been reported previously [25, 26], this study is the first to our knowledge to present a validated claims-based algorithm for characterizing ovarian cancer LOTs, treatment type, and treatment regimens. A similar algorithm was recently utilized by

Table 3 Algorithm performance: identification of neoadjuvant and adjuvant therapy

LOTs identified in claims	LOTs identified by chart abstraction		κ[a]	*P*-value[b]
Neoadjuvant, *n* (%)[c]	Yes	No		
Yes	26 (9%)	2 (1%)	0.56	< 0.001
No	30 (10%)	236 (80%)		
Adjuvant, *n* (%)[c]	Yes	No		
Yes	160 (54%)	48 (16%)	0.62	< 0.001
No	5 (2%)	81 (28%)		

LOT line of therapy

[a]The kappa statistic (κ) is interpreted as follows: ≤ 0, no agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, almost perfect agreement

[b]*P*-values test whether the estimated kappa is not due to chance. They do not test the strength of agreement. Also, *p*-values are sensitive to sample size

[c]Percentages are calculated among the total study population ($N = 294$)

Simmons et al. (2022) to describe healthcare resource utilization and costs associated with progression of ovarian cancer [21]. In the Simmons et al. study, the start of a new LOT was used as a proxy for disease progression; however, this approach required data review by clinical experts to verify that active and maintenance treatment lines had been accurately identified [21]. The validated algorithm presented here will allow researchers to identify LOT among patients with ovarian cancer from administrative claims data with less need for medical review, facilitating the conduct of population-level studies based on claims databases in the future. In addition, the present algorithm was shown to discriminate neoadjuvant from adjuvant therapy, which previous claims algorithms have not incorporated [7, 21]. However, the claims algorithm may overestimate adjuvant and underestimate neoadjuvant. Of the LOTs identified by the chart as neoadjuvant (56), less than half [26] were identified by the algorithm, and of the 208 charts identified as adjuvant by the algorithm, approximately one-third (48) were not identified by charts. In the future, it may be possible to adjust this algorithm to better identify this classification. While the current algorithm has limitations, it does provide a method to identify lines of therapy in this patient population.

Given the increasing importance of real-world evidence for informing treatment decisions in oncology [17, 18], this LOT algorithm represents a valuable tool for leveraging administrative claims to conduct robust analyses of ovarian cancer treatments in clinical practice. Data from such studies could be leveraged to determine current treatment patterns and treatment sequencing for ovarian cancer both overall and among particular patient groups—for example, patients who had a short time to progression after first-line treatment or who are on more-advanced lines of therapy—to help identify treatment pathways that are associated with better clinical and economic outcomes.

Study Strengths and Limitations

Strengths of this study include a sample drawn from the ORD, which represents both commercial and Medicare Advantage enrollees and has a broad geographic distribution across the US, and the inclusion of treatment and healthcare utilization information from all of a patient's healthcare providers to capture the full impact of care for ovarian cancer.

This study also has several limitations. First, as this analysis included abstraction of charts

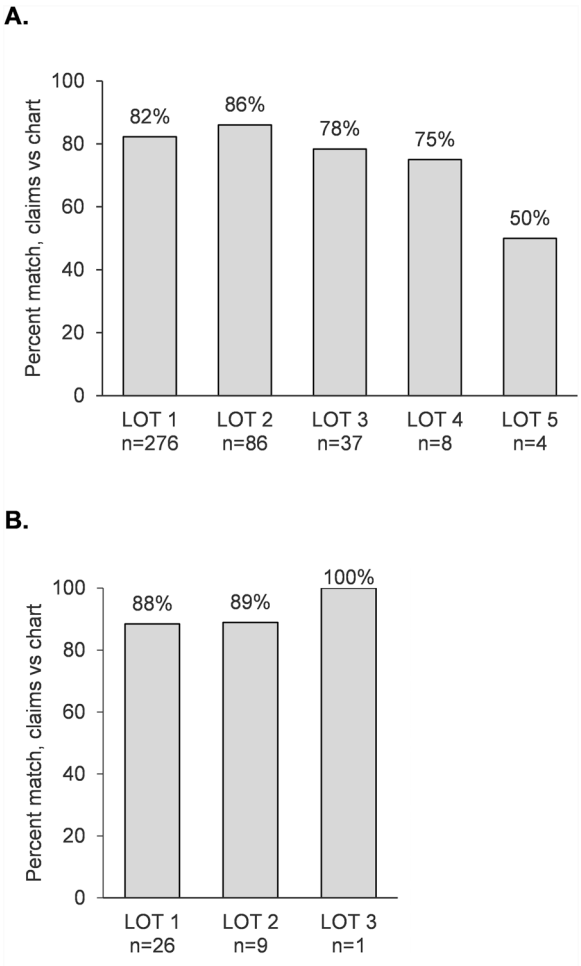

Fig. 2 Algorithm performance: regimen match between claims and chart for **A** active LOTs and **B** maintenance LOTs. Percent match is calculated among the number of patients with a LOT in both the claims and the chart. *LOT* line of therapy

from a single oncology provider per patient, some clinical data may have been missing for patients treated by multiple providers. In addition, only a subset of patients identified in the claims was available for chart review (1378 out of 1773), and only a proportion of patients' providers agreed to participate (451 out of 820 agreed to participate and 294 charts were eligible). Second, although up to 33 months of follow-up data were collected in the chart review, not all charts had this length of follow-up. Third, patient numbers in later LOTs were small. Results may not be generalizable to uninsured populations because the study sample was selected from among patients with commercial or Medicare Advantage insurance. The sample included more patients with Medicare Advantage insurance (62%) and may have been skewed more towards how Medicare Advantage patients are treated, but we do not know the impact of this. Finally, the algorithm has not been tested outside of the ORD. Future work could test this algorithm in different data sources.

CONCLUSION

We developed and validated an administrative claims-based algorithm that demonstrates substantial concordance with medical records for identifying LOT among patients with ovarian cancer. The algorithm can be applied in future studies to analyze treatment patterns and outcomes in this patient population using administrative claims data.

ACKNOWLEDGEMENTS

The authors acknowledge Sharanya Murali and Angela Belland (both of Optum) for study oversight. Medical writing support was provided by Yvette Edmonds and Gretchen Hultman (both of Optum). We thank the physicians who participated in the medical chart review portion of the study.

Medical Writing, Editorial, and Other Assistance. Medical writing assistance was provided by Yvette Edmonds and Gretchen Hultman (both of Optum) and was funded by AstraZeneca and Merck Sharp & Dohme LLC, a subsidiary of Merck & Co, Inc, Rahway, NJ, USA.

Author Contributions. Conceptualization and study design were conducted by Dan Simmons, John White, Valery Walker, Jiefen Munley, Kimmie McLaurin. The study methodology was determined by Dan Simmons, John White, Valery Walker, Jiefen Munley, Kimmie McLaurin.

Acquisition and analysis of the data were conducted by John White and Valery Walker. Dan Simmon, John White, Valery Walker, Jiefen Munley, Kimmie McLaurin, and Stephanie Blank interpreted the findings. All authors read and approved the final manuscript and reviewed and commented on all previous versions.

Funding. This study and the journal's rapid service and open access fees were funded by AstraZeneca and is part of an alliance between AstraZeneca and Merck Sharp & Dohme LLC, a subsidiary of Merck & Co, Inc, Rahway, NJ, USA.

Data Availability. The datasets generated during and/or analyzed during the current study are not publicly available due to the proprietary elements contained in the database, which are owned by Optum. The disclosure of this data to third part clients assumes certain data security and privacy protocols are in place and that the third-party client has executed Optum's standard license agreement, which includes restrictive covenants governing the use of the data.

Declarations

Conflicts of Interest. Daniel Simmons, Jiefen Munley, and Kimmie McLaurin are employees of AstraZeneca. John White and Valery Walker are employees of Optum, which received funding from AstraZeneca to conduct this study. Stephanie Blank has received grant support through the NCI and serves on the Gynecologic Oncology Division of the American Board of Obstetrics and Gynecology.

Ethical Approval. The study protocol was reviewed and approved by the New England Institutional Review Board (15 June 2020, #20201707). The study obtained a waiver of informed consent and a waiver of authorization under 45 CFR 164.512(i).

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7–33.

2. National Cancer Institute. SEER cancer stat facts: Ovarian cancer. Bethesda, MD: National Cancer Institute; 2021 [Available from: https://seer.cancer.gov/statfacts/html/ovary.html. Accessed 05 Oct 2022.

3. Kurnit KC, Fleming GF, Lengyel E. Updates and new options in advanced epithelial ovarian cancer treatment. Obstet Gynecol. 2021;137(1):108–21.

4. Sato S, Itamochi H. Neoadjuvant chemotherapy in advanced ovarian cancer: latest results and place in therapy. Ther Adv Med Oncol. 2014;6(6):293–304.

5. du Bois A, Reuss A, Pujade-Lauraine E, Harter P, Ray-Coquard I, Pfisterer J. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire (GINECO). Cancer. 2009;115(6):1234–44.

6. DiSilvestro P, Colombo N, Harter P, Gonzalez-Martin A, Ray-Coquard I, Coleman RL. Maintenance treatment of newly diagnosed advanced ovarian cancer: time for a paradigm shift? Cancers (Basel). 2021;13(22):5756. https://doi.org/10.3390/cancers13225756.

7. Beachler DC, Lamy FX, Russo L, Taylor DH, Dinh J, Yin R, et al. A real-world study on characteristics, treatments and outcomes in US patients with

advanced stage ovarian cancer. J Ovarian Res. 2020;13(1):101.

8. Hall JP, Chang J, Moon R, Higson O, Byrne K, Doherty JP, et al. Real-world treatment patterns in patients with advanced (stage III-IV) ovarian cancer in the USA and Europe. Future Oncol. 2020;16(15):1013–30.

9. Jessmon P, Boulanger T, Zhou W, Patwardhan P. Epidemiology and treatment patterns of epithelial ovarian cancer. Expert Rev Anticancer Ther. 2017;17(5):427–37.

10. Eisenhauer EA. Real-world evidence in the treatment of ovarian cancer. Ann Oncol. 2017;28(suppl_8):viii61–5.

11. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Int J Surg. 2011;9(8):672–7.

12. Lee JM, Minasian L, Kohn EC. New strategies in ovarian cancer treatment. Cancer. 2019;125(Suppl 24):4623–9.

13. Unger JM, Barlow WE, Martin DP, Ramsey SD, Leblanc M, Etzioni R, et al. Comparison of survival outcomes among cancer patients treated in and out of clinical trials. J Natl Cancer Inst. 2014;106(3):dju002.

14. Karanis YB, Bermudez Canta FA, Mitrofan L, Mistry H, Anger C. "Research" vs "real world" patients: the representativeness of clinical trial participants. Ann Oncol. 2016;27(Suppl 6):vi526–44.

15. Miksad RA, Abernethy AP. Harnessing the power of Real-World Evidence (RWE): a checklist to ensure regulatory-grade data quality. Clin Pharmacol Ther. 2018;103(2):202–5.

16. Roberts MH, Ferguson GT. Real-world evidence: bridging gaps in evidence to guide payer decisions. Pharmacoecon Open. 2021;5(1):3–11.

17. Penberthy LT, Rivera DR, Lund JL, Bruno MA, Meyer AM. An overview of real-world data sources for oncology and considerations for research. CA Cancer J Clin. 2022;72(3):287–300.

18. Brixner D, Biskupiak J, Oderda G, Burgoyne D, Malone DC, Arondekar B, et al. Payer perceptions of the use of real-world evidence in oncology-based decision making. J Manag Care Spec Pharm. 2021;27(8):1096–105.

19. Arend RC, O'Malley DM, Banerjee S, McLaurin K, Davidson R, Long GH. Utilization of poly(ADP-Ribose) polymerase inhibitors in ovarian cancer: a retrospective cohort study of US Healthcare Claims Data. Adv Ther. 2022;39(1):328–45.

20. Suidan RS, He W, Sun CC, Zhao H, Rauh-Hain JA, Fleming ND, et al. Total and out-of-pocket costs of different primary management strategies in ovarian cancer. Am J Obstet Gynecol. 2019;221(2):136e1-e9.

21. Simmons D, Blank SV, ElNaggar AC, Chastek B, Bunner SH, McLaurin K. Health care resource utilization and costs associated with disease progression in ovarian cancer. Adv Ther. 2022;39(6):2544–61.

22. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011;173(6):676–82.

23. Clinical Classifications Software (CCS) for ICD-9-CM/ICD-10-CM. Agency for Healthcare Research and Quality, Rockville, MD. Rockville, MD: Agency for Healthcare Research and Quality; [Available from: http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed 05 Oct 2022.

24. Tang W, Hu J, Zhang H, Wu P, He H. Kappa coefficient: a popular measure of rater agreement. Shanghai Arch Psychiatry. 2015;27(1):62–7.

25. Esposito DB, Russo L, Oksen D, Yin R, Desai VCA, Lyons JG, et al. Development of predictive models to identify advanced-stage cancer patients in a US healthcare claims database. Cancer Epidemiol. 2019;61:30–7.

26. Huepenbecker SP, Zhao H, Sun CC, Fu S, He W, Giordano SH, et al. Algorithm to identify incident epithelial ovarian cancer cases using claims data. JCO Clin Cancer Inform. 2022;6: e2100187.