Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Validation of an Algorithm to Identify Venous Thromboembolism in Health Insurance Claims Data Among Patients with Rheumatoid Arthritis

Sangmi Kim[1], Carolyn Martin[2], John White[2], Maureen Carlyle[2], Bonnie Bui[2], Shiyao Gao[1], Claudia A Salinas[1]

[1]Global Patient Safety Pharmacoepidemiology, Eli Lilly and Company, Indianapolis, IN, USA; [2]Health Economics and Outcomes Research, Optum Life Sciences, Eden Prairie, MN, USA

Correspondence: Claudia A Salinas, Global Patient Safety Pharmacoepidemiology, Eli Lilly and Company, Lilly Corporate Center, 893 Delaware St, Indianapolis, IN, 46285, USA, Tel +1 317 433 9188, Email claudia.salinas@lilly.com

Clinical Epidemiology downloaded from https://www.dovepress.com/
For personal use only.

Purpose: Health insurance claims databases provide an opportunity to study uncommon events, such as venous thromboembolism (VTE), in large patient populations. This study evaluated case definitions for identifying VTE among patients treated for rheumatoid arthritis (RA) using *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM) codes in claims data.

Patients and Methods: Study participants were insured adults who received treatment for and had a diagnosis of RA between 2016 and 2020. After a 6-month covariate assessment window, patients were observed for ≥1 month until health plan disenrollment, occurrence of a presumptive VTE, or end of the study (12/31/2020). Presumptive VTEs were identified using predefined algorithms based on ICD-10-CM diagnosis codes, anticoagulant use, and care setting. Medical charts were abstracted to confirm the VTE diagnosis. Performance of primary and secondary (less stringent) algorithms was assessed by calculating the positive predictive value (PPV; primary and secondary objectives). Additionally, a linked electronic health record (EHR) claims database and abstracted provider notes were used as a novel alternative source to validate claims-based outcome definitions (exploratory objective).

Results: A total of 155 charts identified with the primary VTE algorithm were abstracted. The majority of patients were female (73.5%), with mean (standard deviation) age 66.4 (10.7) years and Medicare insurance (80.6%). Obesity (46.8%), ever smoking (55.8%), and prior evidence of VTE (28.4%) were commonly reported in medical charts. The PPV for the primary VTE algorithm was 75.5% (117/155; 95% confidence interval [CI], 68.7%, 82.3%). A less stringent secondary algorithm had a PPV of 52.6% (40/76; 95% CI, 41.4%, 63.9%). Using an alternative EHR-linked claims database, the primary VTE algorithm PPV was lower, potentially due to the unavailability of relevant records for validation.

Conclusion: Administrative claims data can be used to identify VTE among patients with RA in observational studies.

Keywords: algorithm validation, rheumatoid arthritis, venous thromboembolism, positive predictive value, administrative claims data, ICD-10-CM

Introduction

Individuals with rheumatoid arthritis (RA) are at increased risk of venous thromboembolism (VTE).[1–4] Both deep vein thrombosis (DVT) and pulmonary embolism (PE), the 2 components of VTE, are common medical concerns associated with significant morbidity and mortality.[5] The increased risk of VTE in RA is likely due to a higher prevalence of risk factors for VTE among these patients, such as older age, smoking, and obesity.[6] However, RA disease activity and severity may also contribute to the increased risk of VTE.[7,8]

Health insurance claims databases provide an opportunity to study uncommon systemic complications like VTE in large populations of individuals treated for RA in the real world. These studies rely on the use of administrative codes and claims-based algorithms to identify outcomes such as VTE.[9–11] In general, case definition algorithms are developed based on available claims data for a clinical condition of interest and an understanding of clinical workflow to identify

Received: 22 December 2022
Accepted: 16 May 2023
Published: 1 June 2023

© 2023 Kim et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v3.0) License (http://creativecommons.org/licenses/by-nc/3.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

cases. These algorithms may include a combination of codes for diagnosis (eg, codes for DVT and PE), laboratory tests or procedures, drugs prescribed for therapies (eg, anticoagulants), or patient-reported symptoms.[12] Interpretation of real-world evidence using algorithms in claims databases relies heavily on the properties and performance of these case definitions in validation studies and their applicability to the target database and study populations.[10,11] Validation studies that evaluate the accuracy of an algorithm to identify cases against a reference standard are thus essential to determine the validity of findings derived from using the algorithm.[12]

A systematic review reported a range of positive predictive values (PPVs) for VTE defined using *International Classification of Diseases, Ninth Revision* (ICD-9) diagnosis codes to appropriately identify VTEs validated against medical charts.[13] Although some studies report that ICD-9 codes do not accurately identify VTE,[14] PPV for a case definition based on *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) was high when VTE type (PE or DVT), code position, healthcare setting (inpatient or outpatient), and evidence of anticoagulation were considered.[15] Only a few studies to date have used *International Classification of Diseases, Tenth Revision* (ICD-10) codes to identify VTE in patients with RA[16] or in internal medicine inpatients,[17] and none have reported on a case definition in patients with RA based on *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM), warranting further investigation.

This study aimed to evaluate the performance of a VTE case definition by confirming presumptive VTEs identified in administrative claims data against medical chart review among a US population of insured adults diagnosed with and treated for RA. The aim was to support a comparative safety study of VTE in patients with RA treated with a Janus kinase inhibitor (JAKi, baricitinib) versus tumor necrosis factor inhibitors (TNFi) using administrative claims databases.[18] The primary objective was to estimate the PPV of the primary VTE case definition. A secondary objective was to estimate the PPV of a less stringent secondary VTE case definition. An exploratory objective was to evaluate the utility of claims data linked to electronic health record (EHR) provider clinical notes and review of these notes to identify and confirm presumptive VTEs.

Materials and Methods

Data Source and Study Design

This outcome validation study used administrative healthcare medical and pharmacy claims data, including enrollment information with patient-level linked clinical chart and provider notes data from 2 sources. The Optum Research Database (ORD) contains de-identified claims data for more than 73 million individuals (1993 to present) enrolled in commercial and Medicare Advantage Part D health plans across the United States. For each patient included in the ORD analysis, a provider was identified based on claims data and invited to provide the individual's medical chart for a date range of interest (chart procurement) to a professional medical abstraction firm for VTE diagnosis confirmation. Optum's Market Clarity (MC) database is a de-identified claims database linked with the Optum's clinical EHR database, which includes approximately 30 million individuals with complete commercial, Medicare, or Medicaid health plan eligibility. Providers can include free-text notes in the clinical EHR. When available, these notes were abstracted.

Study subjects were adults with evidence of RA diagnosis and treatment with an RA therapy of interest who had continuous medical and pharmacy coverage during a predefined identification window (Table 1). The first claim for the RA therapy represented the study entry date (Day 0 or index date) (Figure 1). The exclusion window encompassed a 6-month pre-index (including index date) claims-based covariate assessment window (baseline) and ≥31 days post-index. The variable post-index follow-up window continued until health plan disenrollment, presumptive claims-based VTE identification (qualifying event), or study end, with the final 31 days reserved for observation of anticoagulant use when required by the VTE case definition applied. The first possible qualifying event date was the index date plus 1 day. The qualifying event date defined a 3-month charts or notes abstraction window (from 1-month pre-event through 2 months post-event) used to confirm VTE diagnosis and collect clinical and behavioral variables not available in claims data.

Table 1 Summary Description of Study Cohorts

	ORD Cohort 1	ORD Cohort 2	MC Cohort
Data sources	ORD linked patient-level charts	ORD linked patient-level charts	MC database matched-EHR provider clinical notes
Insurance type	Commercial and Medicare	Commercial and Medicare	Commercial, Medicare, and Medicaid
Identification window	July 1, 2016, to October 30, 2019	May 1, 2016, to November 30, 2020	May 1, 2016, to August 31, 2020
Index RA therapy	TNFi[a] or JAKi[b]	TNFi[a], JAKi[b], or non-TNFi[c]	TNFi[a], JAKi[b], or non-TNFi[c]
Specific exclusion criteria	Not new RA therapy user[d] Not eligible for re-identification[e]	Not eligible for re-identification[e]	No matched-EHR provider notes[f]
VTE algorithm applied	Primary, secondary	Primary	Primary
Study objective	Primary[g], secondary	Primary[g]	Exploratory

Notes: [a]TNFi therapy included adalimumab, certolizumab pegol, etanercept, golimumab, or infliximab. [b]JAKi therapy included baricitinib, tofacitinib, or upadacitinib. [c]Non-TNFi therapy included abatacept, anakinra, tocilizumab, or sarilumab. [d]Patients with ≥1 claim for a TNFi therapy or a JAKi therapy during the 6-month CAW were excluded only from ORD Cohort 1. [e]Patients without patient and provider identifiable information available to support medical chart identification and abstraction. [f]Patients with provider(s) who had contributed free-text clinical notes within the 3-month CAW in the Optum-matched EHR database that could be extracted and redacted for abstraction. [g]Primary study objective was applied to patients identified from ORD Cohort 1 and ORD Cohort 2 who had a presumptive VTE defined using the primary case definition (primary objective sample).

Abbreviations: CAW, covariate assessment window; EHR, electronic health record; JAKi, Janus kinase inhibitor; MC, Market Clarity; ORD, Optum Research Database; RA, rheumatoid arthritis; TNFi, tumor necrosis factor inhibitor; VTE, venous thromboembolism.

Study Cohorts and Case Identification

To evaluate the primary VTE case definition, the ORD was searched for individuals who had ≥1 claim for a TNFi or a JAKi during an identification window of July 1, 2016, to October 30, 2019 (Table 1). Study subjects were also required to have ≥1 RA diagnosis claim (excluding those for diagnostic services; listed in Supplementary Table 1) during the 6-month baseline through index plus 30 days post-index; be ≥18 years old at index date; have continuous medical and pharmacy coverage without any missing data during the 6-month baseline through ≥31 days post-index; and be eligible for chart procurement (available patient- and provider-identifiable information). Patients were excluded if they had ≥1 claim for the index RA therapy in the 6-month pre-index baseline (evidence of not being new user of index RA therapy), ≥1 claim for anticoagulant therapy that preceded and overlapped with qualifying VTE date (evidence of previous VTE), and ≥1 claim for cancer treatment on index date (evidence of active cancer). Patients meeting these eligibility criteria comprised ORD Cohort 1 (Table 1). However, owing to the limited number of presumptive VTE cases identified in ORD Cohort 1, eligibility criteria were expanded to increase the sample size by extending the identification window (May 1, 2016, to November 30, 2020), adding non-TNFi therapies (abatacept, anakinra, tocilizumab, or sarilumab) for identification, and removing the new user of RA therapy requirement. This new, broader cohort became ORD Cohort 2 (Table 1). The combination of presumptive cases identified in ORD Cohorts 1 and 2 was used to evaluate the primary VTE case definition (primary objective sample). ORD Cohort 1 was also used to evaluate a secondary, less stringent VTE case definition (secondary objective sample) (Table 1).

The MC database was used as an alternate and exploratory source of patients to evaluate the primary VTE case definition. The MC Cohort had a sample identification period of May 1, 2016, to August 31, 2020, and the same eligibility criteria as for ORD Cohort 2 (exploratory objective sample) (Table 1).

Descriptive Characteristics

Patient demographic and clinical characteristics were collected from claims data (age, sex, and Medicare coverage at index date; and Charlson comorbidity score during baseline) and abstracted medical charts or provider notes (comorbidities for RA and VTE, including atrial fibrillation, chronic obstructive pulmonary disease [COPD], hypertension, incident cancer, obesity, smoking status, and evidence of prior VTE during the abstraction window). Total follow-up time from

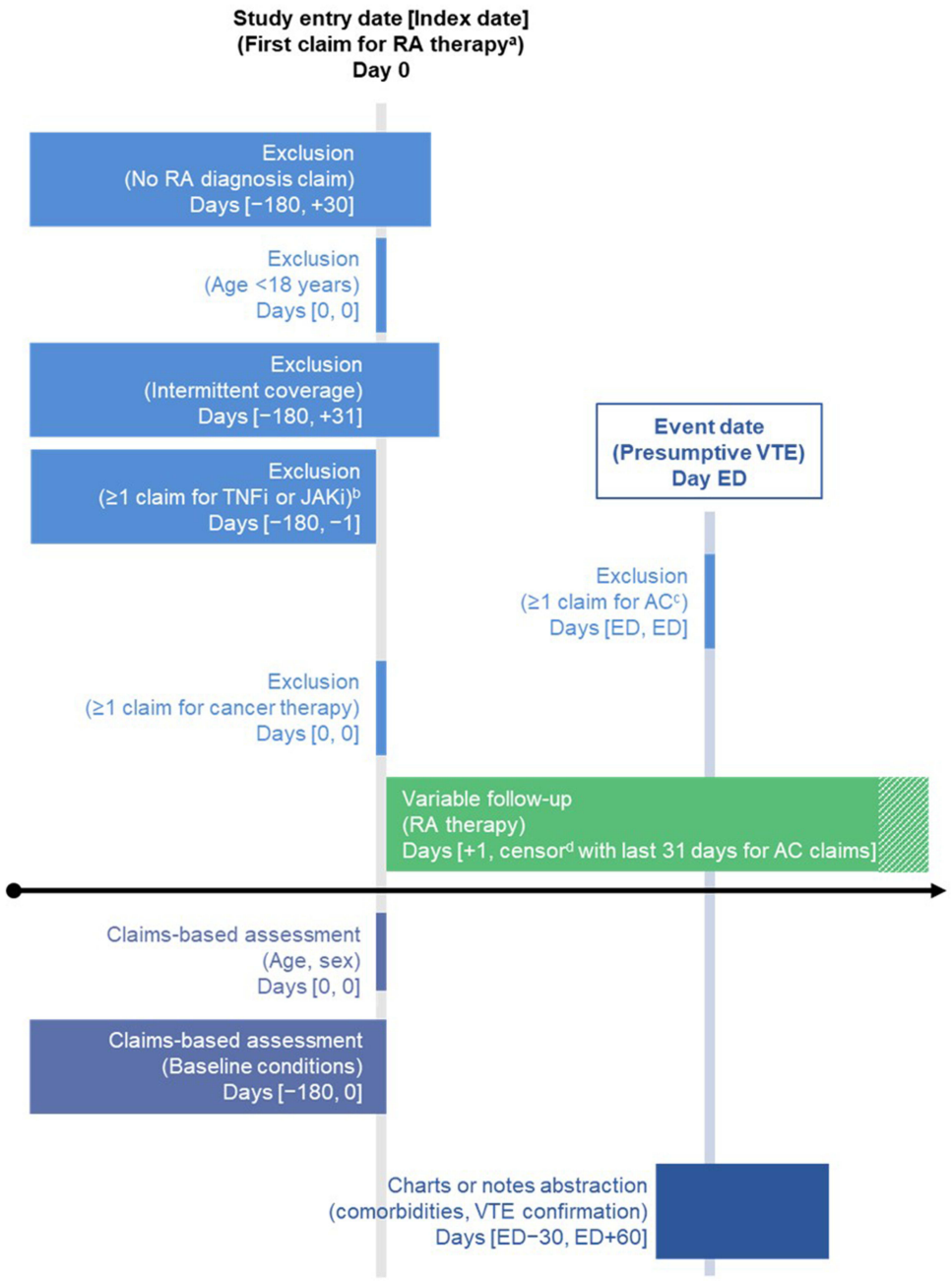

Figure 1 Study design.

Notes: Exclusion criteria are applied to each patient and arranged in event time anchored at study entry [Day 0] and qualifying event [Day ED]. [a]The index date was the date of either the first observed claim for RA therapy or the first claim for a JAKi (even if a claim for a TNFi was observed first), prioritizing baricitinib, followed by tofacitinib and then upadacitinib. [b]Exclusion of patients with ≥1 claim for a TNFi or a JAKi during the 6-month exclusion window was only applied to Optum Research Database Cohort 1. [c]Exclusion of patients with ≥1 claim for AC therapy prescribed before the VTE qualifying event and with a runout date that overlaps with the VTE qualifying ED. [d]Censor occurred at either disenrollment, date of presumptive VTE [Day ED], or end of study period (December 31, 2020), leaving the last 31 days for evaluation of claims for AC therapy (part of presumptive VTE identification).

Abbreviations: AC, anticoagulant; ED, event date; JAKi, Janus kinase inhibitor; RA, rheumatoid arthritis; TNFi, tumor necrosis factor inhibitor; VTE, venous thromboembolism.

1 day post-index to either death, disenrollment, or end of study and time to presumptive VTE from index date to event date were also reported. RA therapies during the variable post-index follow-up were described from claims data.

Presumptive VTE

The VTE primary case definition was based on ICD-10-CM diagnosis codes (Supplementary Table 1) for 4 types of VTE (PE, lower extremity DVT, upper extremity DVT, and other venous thromboses [VT]); site of care (hospital, emergency

Table 2 VTE Case Definition Criteria

Primary Algorithm	Secondary Algorithm
PE diagnosis **Hospital or ER** 1. Primary position 2. Secondary position + AC therapy[a]	**PE diagnosis** **Hospital or ER** 1. Primary position 2. Secondary position **Outpatient** 3. Any position + AC therapy[a]
DVT diagnosis **Hospital or ER** 1. Lower in primary position 2. Upper in primary position + AC therapy[a] 3. Lower in secondary position + AC therapy[a] **Outpatient** 4. Lower in any position + AC therapy[a] 5. Upper in any position + AC therapy[a]	**DVT diagnosis** **Hospital or ER** 1. Lower in primary position 2. Lower in secondary position 3. Upper in primary position 4. Upper in secondary position + AC therapy[a] **Outpatient** 5. Lower in any position[a] 6. Upper in any position + AC therapy[a]
Other VT **Any setting** Any position + AC therapy[a]	**Other VT** **Hospital or ER** 1. Primary position 2. Secondary position + AC therapy[a] **Outpatient** Any position + AC therapy[a]

Notes: [a]Anticoagulant therapy included was defined as having a claim for apixaban, dabigatran, dalteparin, edoxaban, enoxaparin, fondaparinux, rivaroxaban, tinzaparin, or warfarin within 31 days of the presumptive VTE.
Abbreviations: AC, anticoagulant; DVT, deep vein thromboembolism; ER, emergency room; PE, pulmonary embolism; VT, venous thrombosis; VTE, venous thromboembolism.

room [ER], or outpatient); and code position on claim (listed first [primary] or not [secondary]).[13] Within the algorithm, a hierarchy was applied by VTE type, code position on claim, and site of care, which in combination determined the need to include a post-VTE claim for an anticoagulant within 31 days (Table 2). The earliest claim for a VTE meeting the primary case definition was defined as the qualifying event. When multiple qualifying claims occurred on the same day, priority was given to PE over other VTE types among hospital claims, and to hospital claims over ER claims. The primary VTE case definition was applied to ORD Cohorts 1 and 2 and the MC Cohort. A secondary, less stringent VTE case definition (Table 2) was also applied to ORD Cohort 1. The secondary VTE algorithm includes qualifying criteria for the primary VTE algorithm. Therefore, some patients qualified for both the primary and secondary algorithms.

Clinical VTE Confirmation Rules

The VTE case confirmation rules were developed in consultation with clinical experts and were specific to the VTE type. A VTE was confirmed if the abstracted charts (ORD 1 or 2) or provider notes (MC) reported on or within 30 days of the presumptive VTE either a death from VTE or a VTE diagnosis with confirmation from imaging procedures (Supplementary Materials). Cases of VTE diagnosis that did not have accompanying diagnostic tests, had inconclusive imaging test results, or unknown VTE types, place of diagnosis, or service underwent further adjudication by clinical experts.

In the ORD cohorts, the primary algorithm hierarchy facilitated provider identification for chart procurement, with higher levels of care setting being more likely to include multidisciplinary treatment teams and imaging facilities. In the MC cohort, relevant EHR provider notes for abstraction were identified using a list of targeted search terms because an initial feasibility assessment showed a widely variable number of notes available per patient, many of which were not

relevant for the study. Further details on provider selection (ORD) and provider notes (MC) identification are specified in the Supplementary Materials.

Statistical Analyses

All study variables, including baseline and outcome measures, were analyzed descriptively. Performance of the VTE algorithm was assessed by calculating the PPV, based on the numbers of true positives (TP, presumptive VTEs confirmed) and false positives (FP, presumptive VTEs not confirmed) applying the formula: PPV = TP/(TP+FP). The 95% confidence interval (CI) around the PPV was obtained using a normal approximation method. Cases with insufficient information to confirm a VTE (by confirmation rules or expert clinical adjudication) were not included in PPV computation. The presumptive cases identified in the combined ORD Cohorts 1 and 2 (ORD Cohort 1+2) were used to calculate the primary case definition PPV (primary objective sample). To correctly estimate the PPV of the secondary algorithm, all patients from ORD Cohort 1 who qualified for the secondary algorithm were included, including those who qualified for both the primary and secondary algorithms (secondary objective sample). Although the MC database was also explored to identify and validate presumptive VTEs per the primary algorithm case definition (exploratory objective sample), patients from this exploratory sample were not combined with those from the primary sample due to limitations unique to the MC database, and the alternate source of information used for VTE confirmation (ie, providers' notes versus medical charts). Patient and clinical characteristics were analyzed descriptively across study samples with numbers and proportions (%) for categorical variables and means and standard deviations for continuous variables. Demographic characteristics of patients with abstracted versus not abstracted charts/provider notes were presented using means and standard deviations for continuous variables and numbers and proportions for categorical variables. All analyses were performed using SAS version 9.4 (SAS Inc., Cary, NC, USA).

Results

Study Subjects

From a total of 77,002 ORD patients receiving a TNFi or a JAKi, 8995 met all the inclusion/exclusion criteria for ORD Cohort 1 and 70 presumptive VTE cases were identified using the primary VTE case definition (Supplementary Figure 1). Of these, 46 charts were procured and 38 (54.3% of 70) were relevant and abstracted. Using expanded identification and selection criteria specific for ORD Cohort 2, a total of 103,258 patients receiving a biologic disease-modifying antirheumatic drug (bDMARD) or JAKi were found, and 17,776 met all inclusion/exclusion criteria, among which the primary algorithm identified 222 presumptive VTE cases; of these, 184 charts were procured and 117 (52.7% of 222) were abstracted. Although the size of ORD Cohort 1 and ORD Cohort 2 varied (based on applied selection criteria), comparable percentages of charts were abstracted from the presumptive VTE cases identified. Patients whose charts were abstracted were similar to those whose charts were not abstracted in terms of age, sex, and insurance coverage type (Supplementary Table 2). The combined 155 patients consisting of 38 from ORD Cohort 1 and 117 from ORD Cohort 2 (ORD Cohort 1+2) formed the performance sample for the primary algorithm (primary objective sample).

The secondary VTE case definition was applied to the 8995 patients included in the ORD Cohort 1 and identified 126 presumptive cases, of which 85 charts were procured and 76 (60.3% of 126) abstracted (Supplementary Figure 1). These 76 patients were used to evaluate the performance of the secondary algorithm (secondary objective sample). Finally, using similar identification and selection criteria as for ORD Cohort 2, a total of 298,762 patients receiving a bDMARD or JAKi were identified in the MC database, and 28,288 were selected to form the MC Cohort (Supplementary Figure 1). Among these, the primary VTE algorithm identified 253 presumptive cases, of which 73 were found to have relevant unstructured provider notes during the 3-month abstraction window (from 2-month pre- and 1-month post-event), and 64 (25.3% of 253) were eligible for abstraction (exploratory objective sample).

Patient Descriptive Characteristics

Patients from the primary objective sample had a mean (standard deviation [SD]) age of 66.4 (10.7) years, with 62.6% being ≥65 years of age, and a majority being female (73.5%) and having Medicare insurance coverage (80.6%) (Table 3).

Table 3 Patient Demographic and Clinical Characteristics by Study Sample

Sample, Cohort Source	Primary Objective, ORD Cohort 1+2 (n = 155)	Secondary Objective, ORD Cohort 1 (n = 76)	Exploratory Objective, MC Cohort (n = 64)
Claims-based variables			
Age, y	66.4 (10.7)	64.6 (11.4)	60.2 (12.5)
Age ≥65 years	97 (62.6)	38 (50.0)	24 (37.5)
Female	114 (73.5)	52 (68.4)	48 (75.0)
Non-commercial insurance	125 (80.6)	52 (68.4)	32 (50.0)[a,b]
Charlson comorbidity score	2.5 (1.7)	2.6 (2.1)	1.9 (1.2)
Follow-up time, d	912.0 (524.7)	588.3 (341.6)	1134.1 (466.0)
Time to qualifying VTE, d	481.8 (417.6)	279.1 (259.1)	545.9 (454.7)
RA therapy during follow-up			
TNFi: adalimumab	47 (30.3)	26 (34.2)	17 (26.6)
Certolizumab	7 (4.5)	6 (7.9)	2 (3.1)
Etanercept	29 (18.7)	11 (14.5)	18 (28.1)
Golimumab SC	2 (1.3)	2 (2.6)	2 (3.1)
Golimumab IV	11 (7.1)	6 (7.9)	7 (10.9)
infliximab	20 (12.9)	8 (10.5)	14 (21.9)
JAKi: baricitinib	1 (0.6)	0	0
Tofacitinib	32 (20.6)	23 (30.3)	10 (15.6)
Upadacitinib	0	1 (1.3)	0
Non-TNFi: abatacept	20 (17.1)[c]	NR	10 (15.6)
Anakinra	1 (0.9)[c]		0
Sarilumab	3 (2.6)[c]		0
Tocilizumab	28 (23.9)[c]		9 (14.1)
Charts- or notes-based comorbidities			
Evidence of prior VTE	44 (28.4)	21 (27.6)	18 (28.1)
Atrial fibrillation	25 (16.1)	10 (13.2)	6 (9.4)
COPD	42 (27.1)	21 (27.6)	9 (14.1)
Hypertension	112 (72.3)	51 (67.1)	37 (57.8)
Cancer	12 (7.7)	5 (6.6)	6 (9.4)
Obese (BMI ≥30 kg/m^2)[b]	66 (46.8)	35 (50.0)	36 (67.9)
Ever smoker[b]	82 (55.8)	36 (56.3)	29 (54.7)

Notes: Data represent mean (SD) for continuous variables and n (%) for categorical variables. [a]Included 20 (31.25%) covered under Medicare and 12 (18.75%) under Medicaid. [b]Denominator varies accounting for missing data. [c]Claims for non-TNFi were not reported in ORD Cohort 1, so the denominator is 117.
Abbreviations: BMI, body mass index; COPD, chronic obstructive pulmonary disease; IV, intravenous; JAKi, Janus kinase inhibitor; MC, Market Clarity; NR, not reported; ORD, Optum Research Database; RA, rheumatoid arthritis; SC, subcutaneous; SD, standard deviation; TNFi, tumor necrosis factor inhibitor; VTE, venous thromboembolism.

The mean (SD) Charlson comorbidity score was 2.5 (1.7). During follow-up, the most common RA therapy claims were for adalimumab (30.3%), tocilizumab (23.9%), tofacitinib (20.6%), and etanercept (18.7%). The time to presumptive VTE ranged from 2 to 1676 days, with a mean (SD) of 482 (418) days. The most frequent comorbidities observed during chart abstraction were hypertension (72.3%) and COPD (27.1%). Approximately half of patients were obese (46.8%) or ever smokers (55.8%), and 28.4% had a history of prior VTE according to medical charts. Demographic and clinical characteristics were overall similar between patients from ORD Cohort 1 and 2 (Supplementary Table 3). Patients

included in the secondary objective sample had very similar demographic and clinical characteristics to those included in the primary objective sample (Table 3).

However, compared with the primary and secondary objective samples (ORD Cohorts), the 64 patients from the exploratory objective sample (MC Cohort) had a younger age, lower mean Charlson comorbidity score, and lower proportion with non-commercial insurance coverage (Table 3). The most common claims for RA therapies during follow-up were etanercept (28.1%), adalimumab (26.6%), and infliximab (21.9%). The time to presumptive VTE ranged from 8 to 1584 days, with a mean (SD) of 546 (455) days. The most frequent comorbid conditions identified in provider notes were hypertension and COPD, although they were reported less frequently than in the other samples; a greater percentage of patients were obese, but similar percentages had ever smoked or had a history of prior VTE compared with the ORD cohorts.

Qualifying VTE and Performance of VTE Case Definitions

Among the 155 patients with abstracted charts included in the primary objective sample, many met several of the 17 qualifying criteria for the primary VTE case definition based on claims data. Of these 155 patients, 117 had confirmed VTE (TP), and the resulting PPV was 75.5% (95% CI: 68.7, 82.3) (Table 4). Similar PPVs were obtained from ORD Cohort 1 alone (76.3%; 95% CI: 62.8, 89.8) and ORD Cohort 2 alone (75.2%; 95% CI: 67.4, 83.0). Based on abstracted medical charts data, 30 patients had more than 1 confirmed VTE. The types of confirmed VTEs were nearly evenly split between lower extremity DVT (n = 75) and PE (n = 67) (Table 5). When patients had 2 confirmed VTEs, the most frequent combination was PE and lower extremity DVT. The most common reasons that claims-based VTEs were not

Table 4 Performance of VTE Case Definitions

Sample	**Primary Objective**			**Secondary Objective**	**Exploratory Objective**
Cohort Source	**ORD Cohort 1+2 (n = 155)**	**ORD Cohort 1 (n = 38)**	**ORD Cohort 2 (n = 117)**	**ORD Cohort 1 (n = 76)**	**MC Cohort (n = 64)**
False positive, n (%)	38 (24.5)	9 (23.7)	29 (24.8)	36 (47.4)	28 (43.8)
True positive, n (%)	117 (75.5)	29 (76.3)	88 (75.2)	40 (52.6)	36 (56.3)
PPV, % (95% CI)	**75.5 (68.7, 82.3)**	**76.3 (62.8, 89.8)**	**75.2 (67.4, 83.0)**	**52.6 (41.4, 63.9)**	**56.3 (44.1, 68.4)**

Abbreviations: CI, confidence interval; MC, Market Clarity; ORD, Optum Research Database; PPV, positive predictive value; VTE, venous thromboembolism.

Table 5 Characteristics of Confirmed VTEs

Sample, Cohort Source	**Primary Objective, ORD Cohort 1+2 (n = 155)**	**Secondary Objective, ORD Cohort 1 (n = 76)**	**Exploratory Objective, MC Cohort (n = 64)**
Total number of patients with confirmed VTEs	117	40	36
Patients with 1 VTE, n (%)	87 (56.1)	26 (34.2)	29 (45.3)
Patients with 2 VTE, n (%)	30 (19.4)	14 (18.4)	7 (10.9)
***VTE type*, n (%)**			
PE	67 (43.2)	27 (35.5)	17 (26.6)
Lower extremity DVT	75 (48.4)	27 (35.5)	19 (29.7)
Upper extremity DVT	1 (0.6)	0	3 (4.7)
Other VT	4 (2.6)	0	4 (6.3)
Both PE and lower extremity DVT	NR	14 (18.4)	NR

Abbreviations: DVT, deep vein thromboembolism; MC, Market Clarity; NR, not reported; ORD, Optum Research Database; PE, pulmonary embolism; VT, venous thrombosis; VTE, venous thromboembolism.

confirmed were final diagnosis ruling out VTE; treatment decisions based on VTE history; lack of or inappropriate imaging test; and miscoded VTE for related condition (such as arterial thrombosis or fat embolus).

Among the 76 abstracted charts for the secondary objective (ORD Cohort 1), 40 had confirmed VTE (TP) resulting in a PPV of 52.6% (95% CI: 41.4, 63.9) (Table 4). The types of VTE confirmed through chart abstraction were evenly split between lower extremity DVT and PE (n = 27 each), with 14 patients having both confirmed lower extremity DVT and PE (Table 5). No cases of upper extremity DVT or other VT were confirmed during chart review.

In the exploratory objective sample, a VTE was confirmed through provider notes review for 36 of 64 patients, resulting in a PPV of 56.3% (95% CI: 44.1, 68.4) (Table 4). Seven patients had 2 confirmed VTEs. Lower extremity DVT and PE were the most common types of confirmed VTEs (Table 5).

Discussion

Although there are no established benchmarks for PPV in real-world evidence studies,[11] the US Food and Drug Administration has recommended a PPV threshold of >70% as a guide when using validation against full-text medical record review.[19] In this study, 75.5% of VTEs identified from a claims-based algorithm were confirmed with abstracted medical chart data using rigorous clinical criteria, highlighting the accuracy of the primary VTE algorithm in its ability to correctly identify VTE in individuals receiving RA therapy. Not surprisingly, the PPV of the less stringent secondary algorithm was lower at 52.6%, which is likely attributable to the inclusion of a wider number of diagnostic and other criteria in the secondary algorithm definition, including relaxing the anticoagulant requirement and allowing outpatient diagnosis for PE. Interestingly, the same primary algorithm validated using an alternative data source (EHR provider notes from the MC database) produced a lower PPV (56.3%) than using the traditional chart reviews from the ORD.

The PPV estimates in our study fall within a range of what has been reported in previous chart validation studies using claims-based algorithms for VTE. In a study of adult patients with VTE diagnoses between 2004 and 2010 from the Cardiovascular Research Network Venous Thromboembolism, the overall PPV of any VTE code, based on ICD-9-CM coding, was 51.9%, but varied widely by clinical setting and VTE type.[15] Primary discharge diagnosis codes obtained from a hospital/ER encounter had greater PPV (78.9%) compared with codes in secondary positions (44.4%) or obtained in an outpatient setting (30.9%). These results demonstrated the importance of the diagnosis setting, the position of codes used in inpatient settings, and evidence of anticoagulant therapy, to the PPV of the case definition. Recently, a population-based study of patients with RA from the Swedish National Patient Register (2009–2018) reported a PPV of 87% for incident VTE using their main algorithm, based on ICD-10 coding.[16] Differences between the Swedish and US healthcare systems may contribute to differences in reported PPVs, such as the more granular ICD-10-CM codes used in the US healthcare system than the ICD-10 codes used in Sweden.

Validation of claims-based outcome definitions involves verification of the outcome against a gold standard (eg, medical records) and typically requires a time- and labor-intensive process of procurement of medical charts, manual data extraction, and adjudication. While conducting a traditional outcome validation study based on medical charts, this study also experimented with a novel alternative to this process using Optum's MC database, which contains EHR data with provider notes where available. Because the validation exercises were conducted on samples drawn from two different claims systems (ORD vs multi-payer MC), it has provided an interesting case for comparison of validation methods, based on abstracted charts versus abstracted physician notes. Several reasons may have contributed to the lower PPV seen in the MC Cohort versus the ORD Cohorts. First, specific providers (eg, hematologists) were not identified for procurement and notes abstraction, as with the traditional ORD-chart review approach. Second, VTE diagnoses may have been documented more often in other clinical systems than in the current notes, and the current notes may not have included sufficient information to validate a VTE despite the use of intentionally broad search terms, suggested by clinical experts, to avoid excluding notes with relevant information. Finally, patients from the MC Cohort were younger than those in the ORD Cohorts and could have a lower prevalence of VTE, which could have resulted in a lower PPV.[20] Regardless of the reasons, this finding suggests that abstracted charts may still be the preferred source for validation of claims-based case definitions since they contain the most comprehensive information about a condition of interest. However, the PPV of >50% suggests that abstracted provider notes could be considered an alternative for future development of algorithms when charts are not available.

Limitations

Case confirmation criteria required the chart/notes to indicate a diagnosis of VTE and corresponding diagnostic imaging. These rigorous case confirmation criteria may have omitted some true VTEs that were diagnosed using alternative or less rigorous approaches. Additionally, when VTEs were diagnosed by clinicians without appropriate imaging testing, it is unclear if this represented a lack of diagnostic resources available to the clinician or a lack of awareness of best practice standards. Had these patients received appropriate diagnostic imaging, the final case finding results of the claims-based algorithm may have been improved. Among patients whose charts lacked critical evidence necessary to meet the case confirmation criteria established for the study, procurement of an alternative chart with additional clinical insight could have allowed an adjudication. Direct consultation with the provider could also have been pursued to allow adjudication but could not be implemented for this study without patient consent.

Specific limitations of the MC database for research purposes are inherent to EHR databases, which do not necessarily reflect all treatment a patient with RA or VTE is receiving. Unlike a closed system data source such as enrollment-controlled administrative claims, an open EHR database will only include information provided by contributing providers. Also, not all EHR platforms capture data uniformly (by variable, variable type, or allowed values), and differences may persist despite steps taken by multiple platforms to normalize incoming data streams.

Conclusions

The primary case definition based on ICD-10-CM diagnosis codes, clinical setting, and anticoagulant dispensing achieved an adequate accuracy for identifying VTE in a US population of insured adults diagnosed and treated for RA using medical chart review. The secondary VTE case definition, which relaxed the requirement for anticoagulant and included PE diagnosed in outpatient settings, had a lower PPV. When the primary VTE case definition was validated by comparing presumptive cases in MC data with provider notes from the linked EHR, the accuracy was reduced compared with results in the ORD data. The unavailability of relevant records for validation of presumptive cases may potentially have limited the utility of using such data for case validation studies.

Our results demonstrate reliability of the selected primary algorithm to identify VTE among patients with RA in US claims data and may be generalizable to other observational studies of similar target populations and data sources.

Abbreviations

bDMARD, Biologic disease-modifying antirheumatic drug; CAW, Covariate assessment window; CI, confidence interval; COPD, Chronic obstructive pulmonary disease; DVT, Deep vein thrombosis; EHR, Electronic health record; ER, Emergency room; FP, False positive; ICD-9, *International Classification of Diseases, Ninth Revision*; ICD-9-CM, *International Classification of Diseases, Ninth Revision, Clinical Modification*; ICD-10, *International Classification of Diseases, Tenth Revision*; ICD-10-CM, *International Classification of Diseases, Tenth Revision, Clinical Modification*; JAKi, Janus kinase inhibitor; MC, Market Clarity; ORD, Optum Research Database; PE, Pulmonary embolism; PPV, Positive predictive value; RA, Rheumatoid arthritis; SD, Standard deviation; TNFi, Tumor necrosis factor inhibitor; TP, True positive; US, United States; VT, Venous thrombosis; VTE, Venous thromboembolism.

Data Sharing Statement

The data contained in Optum's database contains proprietary elements owned by Optum and therefore cannot be broadly disclosed or made publicly available at this time. The disclosure of this data to third-party clients assumes certain data security and privacy protocols are in place and that the third-party client has executed our standard license agreement, which includes restrictive covenants governing the use of the data.

Ethics Approval and Informed Consent

All data (claims based and chart/notes based) obtained in this study were accessed in compliance with the Health Insurance Portability and Accountability Act of 1996. This study was reviewed and approved by an institutional review board (WCG IRB).

Acknowledgments

The authors thank Jerry Seare, MD (Medical Director, Optum HEOR), Marcelo Gomes, MD (Cleveland Clinic), and Kim Brown, RN (Cleveland Clinic) for their clinical review.

Medical writing support was provided by professional medical writer Catherine Champagne, PhD, CMPP, employee of Kay Square Scientific (Newtown Square, PA, USA), which received funding from Optum.

Funding

This study was funded by Eli Lilly and Company. Optum received funding from Eli Lilly to conduct this study. The funding agreement did not impact the authors' independence in designing the study, collecting the data, interpreting the data, writing the manuscript, and submitting the manuscript for publication.

Disclosure

Sangmi Kim, Shiyao Gao, and Claudia A Salinas are employees and shareholders of Eli Lilly and Company. Carolyn Martin, John White, Maureen Carlyle, and Bonnie Bui were employees of Optum Life Sciences at the time of the study. Eli Lilly and Company contracted with Optum Life Sciences for the purposes of this research. The authors report no other conflicts of interest in this work.

References

1. Bacani AK, Gabriel SE, Crowson CS, Heit JA, Matteson EL. Noncardiac vascular disease in rheumatoid arthritis: increase in venous thromboembolic events? *Arthritis Rheum*. 2012;64(1):53–61. doi:10.1002/art.33322
2. Holmqvist ME, Neovius M, Eriksson J, et al. Risk of venous thromboembolism in patients with rheumatoid arthritis and association with disease duration and hospitalization. *JAMA*. 2012;308(13):1350–1356. doi:10.1001/2012.jama.11741
3. Kim SC, Schneeweiss S, Liu J, Solomon DH. Risk of venous thromboembolism in patients with rheumatoid arthritis. *Arthritis Care Res*. 2013;65(10):1600–1607.
4. Chung WS, Peng CL, Lin CL, et al. Rheumatoid arthritis increases the risk of deep vein thrombosis and pulmonary thromboembolism: a nationwide cohort study. *Ann Rheum Dis*. 2014;73(10):1774–1780. doi:10.1136/annrheumdis-2013-203380
5. Cushman M, Barnes GD, Creager MA, et al. Venous thromboembolism research priorities: a scientific statement from the American Heart Association and the International Society on Thrombosis and Haemostasis. *Circulation*. 2020;142(6):e85–e94. doi:10.1161/CIR.0000000000000818
6. Bell EJ, Lutsey PL, Basu S, et al. Lifetime risk of venous thromboembolism in two cohort studies. *Am J Med*. 2016;129(3):339.e19–e26. doi:10.1016/j.amjmed.2015.10.014
7. Liang H, Danwada R, Guo D, et al. Incidence of inpatient venous thromboembolism in treated patients with rheumatoid arthritis and the association with switching biologic or targeted synthetic disease-modifying antirheumatic drugs (DMARDs) in the real-world setting. *RMD Open*. 2019;5(2):e001013. doi:10.1136/rmdopen-2019-001013
8. Molander V, Bower H, Frisell T, Askling J. Risk of venous thromboembolism in rheumatoid arthritis, and its association with disease activity: a nationwide cohort study from Sweden. *Ann Rheum Dis*. 2021;80(2):169–175. doi:10.1136/annrheumdis-2020-218419
9. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. *J Epidemiol Commun Health*. 2014;68(3):283–287. doi:10.1136/jech-2013-202744
10. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1018–1032. doi:10.1002/pds.4295
11. Beyrer J, Abedtash H, Hornbuckle K, Murray JF. A review of stakeholder recommendations for defining fit-for-purpose real-world evidence algorithms. *J Comp Eff Res*. 2022;11(7):499–511. doi:10.2217/cer-2022-0006
12. Weinstein EJ, Ritchey ME, Lo Re V. Core concepts in pharmacoepidemiology: validation of health outcomes of interest within real-world healthcare databases. *Pharmacoepidemiol Drug Saf*. 2023;32(1):1–8. doi:10.1002/pds.5537
13. Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using administrative and claims data. *Pharmacoepidemiol Drug Saf*. 2012;21(Suppl 1):154–162. doi:10.1002/pds.2341
14. Baumgartner C, Go AS, Fan D, et al. Administrative codes inaccurately identify recurrent venous thromboembolism: the CVRN VTE study. *Thromb Res*. 2020;189:112–118. doi:10.1016/j.thromres.2020.02.023
15. Fang MC, Fan D, Sung SH, et al. Validity of using inpatient and outpatient administrative codes to identify acute venous thromboembolism: the CVRN VTE study. *Med Care*. 2017;55(12):e137–e143. doi:10.1097/MLR.0000000000000524
16. Molander V, Bower H, Askling J. Validation and characterization of venous thromboembolism diagnoses in the Swedish National Patient Register among patients with rheumatoid arthritis. *Scand J Rheumatol*. 2022;1:1–7.
17. Verma AA, Masoom H, Pou-Prom C, et al. Developing and validating natural language processing algorithms for radiology reports compared to ICD-10 codes for identifying venous thromboembolism in hospitalized medical patients. *Thromb Res*. 2022;209:51–58. doi:10.1016/j.thromres.2021.11.020
18. Salinas CA, Louder A, Polinski J, et al. Evaluation of VTE, MACE, and serious infections among patients with RA treated with baricitinib compared to TNFi: a multi-database study of patients in routine care using disease registries and claims databases. *Rheumatol Ther*. 2022;13:1–23. doi:10.1007/s40744-022-00505-1

19. Schumock GT, Lee TA, Pickard AS, et al. Alternative methods for health outcomes of interest validation. Mini-Sentinel Methods; 2013. Available from: https://www.sentinelinitiative.org/sites/default/files/surveillance-tools/validations-literature/Mini-Sentinel-Alternative-Methods-for-Health-Outcomes-of-Interest-Validation_0.pdf. Accessed September 26, 2022.
20. Tenny S, Hoffman MR. Prevalence. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2022.

Clinical Epidemiology Dovepress

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: https://www.dovepress.com/clinical-epidemiology-journal