Drug, Healthcare and Patient Safety

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Validity of Routine Health Data To Identify Safety Outcomes of Interest For Covid-19 Vaccines and Therapeutics in the Context of the Emerging Pandemic: A Comprehensive Literature Review

Kirsty Andresen[1,2], Marina Hinojosa-Campos[3], Bélène Podmore[1–3], Myriam Drysdale[4], Nawab Qizilbash[1–3], Marianne Cunnington[4,5]

[1]OXON Epidemiology, London, UK; [2]London School of Hygiene and Tropical Medicine, London, UK; [3]OXON Epidemiology, Madrid, Spain; [4]GlaxoSmithKline, Middlesex, UK; [5]Analysis Group, London, UK

Correspondence: Kirsty Andresen, Email Kirsty.andresen@oxonepi.com

Drug, Healthcare and Patient Safety downloaded from https://www.dovepress.com/
For personal use only.

Introduction: Regulatory guidance encourages transparent reporting of information on the quality and validity of electronic health record data being used to generate real-world benefit-risk evidence for vaccines and therapeutics. We aimed to provide an overview of the availability of validated diagnostic algorithms for selected safety endpoints for Coronavirus disease 2019 (COVID-19) vaccines and therapeutics in the context of the emerging pandemic prior to December 2020.

Methods: We reviewed the literature up to December 2020 to identify validation studies for various safety events of interest, including myocardial infarction, arrhythmia, myocarditis, acute cardiac injury, vasculitis/vasculopathy, venous thromboembolism, stroke, respiratory distress syndrome (RDS), pneumonitis, cytokine release syndrome (CRS), multiple organ dysfunction syndrome, and renal failure. We included studies published between 2015 and 2020 that were considered high quality assessed with QUADAS and that reported positive predictive values (PPVs).

Results: Out of 43 identified studies, we found that diagnostic algorithms for cardiovascular outcomes were supported by the highest number of validation studies (n=17). Accurate algorithms are available for myocardial infarction (median PPV 80%; IQR 22%), arrhythmia (PPV range >70%), venous thromboembolism (median PPV: 73%) and ischaemic stroke (PPV range ≥85%). We found a lack of validation studies for less common respiratory and cardiac safety outcomes of interest (eg, pneumonitis and myocarditis), as well as for COVID-specific complications (CRS, RDS).

Conclusion: There is a need for better understanding of barriers to conducting validation studies, including data governance restrictions. Regulatory guidance should promote embedding validation within real-world EHR research used for decision-making.

Keywords: validation, routine health data, Covid-19, safety, vaccines, outcomes

Background

In December 2020, within the context of a rapidly evolving pandemic where effective treatments were not yet available, there was a need for rapid generation of safety and effectiveness data in the real-world to supplement what was learnt during clinical trials.[1] Routine health data, that is, data captured during routine clinical care, such as electronic medical records (EMR) or healthcare insurance claims data,[2] are useful resources that are perfectly placed to quickly and efficiently provide information about important safety endpoints of interest in large patient populations.[3] Furthermore, monitoring the prevalence of these endpoints in the general population can serve as an important baseline comparator against the observed rates occurring in the vaccinated/treated population.[4]

However, routine health databases are created for healthcare planning, monitoring and in some cases for insurance claims reimbursement coordination, not research.[3] It is necessary to understand the validity of diagnostic algorithms that use

Received: 1 April 2023
Accepted: 15 August 2023
Published: 3 January 2024

© 2024 Andresen 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).

diagnostic codes to accurately identify key endpoints of interest as this will impact the ability of epidemiological research to robustly detect safety signals that emerge for these therapies and vaccines within routinely collected health data.

We therefore conducted a literature review to provide an overview of the availability of validated diagnostic algorithms for selected safety endpoints for COVID-19 vaccines and therapeutics in routine health data in North America and Western Europe in order to understand the robustness of real-world evidence that may be generated in non-trial settings in the context of an emerging pandemic.

Methods

Search Strategy

The literature search was conducted in Medline and EMBASE up until the 1st December 2020. Additional searches were conducted by hand-searching reference lists of key articles. Outcomes were selected from the Safety Platform for Emergency Vaccines (SPEAC) as potential Adverse Events of Special Interest (AESI) for COVID-19 vaccines.[5] The selected outcomes were cardiovascular outcomes (myocardial infarction (MI), arrhythmia, myocarditis, acute cardiac injury (ACI), vasculitis/vasculopathy), venous thromboembolism (VT), cerebrovascular outcomes (stroke), respiratory outcomes (respiratory distress syndrome (RDS), pneumonitis) and renal failure. Furthermore, we also included systemic outcomes (cytokine release syndrome (CRS) and multiple organ dysfunction syndrome [MODS]) in this study as they are considered key endpoints in assessing the effectiveness of therapeutics for severe COVID-19. We combined terms for the outcomes of interest with routine health data and validation terms. See the Supplementary Information for the full search strategy.

Screening, Selection Criteria, Data Extraction and Analysis

The studies found in the chosen databases were imported into Endnote. The titles and abstracts were screened by two investigators, any discrepancies were resolved by a third reviewer. We included validation studies reporting positive predictive values (PPVs) published in the last five years (2015–2020) and conducted in human adult populations within North America (USA, Canada) and Europe (France, Germany, Italy, Spain, the Netherlands, Denmark, Finland, Norway, Sweden and the UK). These geographies were chosen to reflect the availability of the majority of population-based healthcare data sources. We restricted the time period as we wanted to capture studies that better reflected the coding practices prior to the introduction of COVID-19 vaccines and therapeutics. However, to be exhaustive, if no studies were identified within the specified five years, the time period was expanded to fifteen years (2005–2020); this was the case for ACI, myocarditis, vasculitis, RDS, pneumonitis, CRS, and MODS. Only articles in the English language were considered given practical limitations. Information on the article characteristics (author, title, journal, year); study population (sample size, selection criteria, study setting), data source, index test algorithm definition, reference standard definition and PPV were extracted from the included articles into an Excel workbook and confirmed by the other investigator. No meta-analysis was conducted due to the heterogeneity of the included studies.

Quality Assessment

Full-text articles were reviewed by the same investigators. Quality was scored using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool.[6] The QUADAS tool was developed as a quality assessment tool to be used in systematic reviews of diagnostic accuracy studies. QUADAS is an evidence-based tool which consists of 14 items phrased as questions, each scored as “yes” (1 point), “no” (0 points) or “unclear” (0 points). These questions aimed to assess various aspects of the studies such as the selection of the study population, the adequacy of the reference test and index test, and the timing of administration of the tests: We evaluated the representativeness of the patient spectrum and the clarity of the selection criteria; the suitability of the reference standard for accurately classifying the target condition and verification; the replicability, independence and interpretation of the index test; Lastly, the adequacy of the time period between the reference standard and index test. Any disagreements were discussed and resolved by a third reviewer. Low quality defined as a QUADAS score <7 was considered a further reason for exclusion.

Results

Selected Studies

After the removal of duplicates, 5,862 titles and abstracts were screened and 148 were selected for full-text review, of these, 101 studies were excluded. The reasons for exclusion are presented in Figure 1. A total of eight of the 47 studies assessed against the QUADAS tool did not meet the quality criteria and were excluded, resulting in 38 studies extracted across all the events of interest in 2020. The majority of studies were conducted in the US (n=18), followed by Canada (n=7) and the UK (n=4). Only one included study was published prior to 2015.[7]

The most studied outcomes were MI (n=13), VT (n=10), stroke (n=9) and arrhythmias (n=5); a summary of the PPVs for these endpoints are detailed in the forthcoming sections. Only one validation study was identified for each of pneumonitis,[7] myocarditis[8] and RDS.[13] When using diagnostic codes in the primary diagnosis field to define these conditions, the accuracy was middling to low at 72%, 64%, and 46%, respectively (see Table 1 for myocarditis and Table 2 for RDS and pneumonitis).

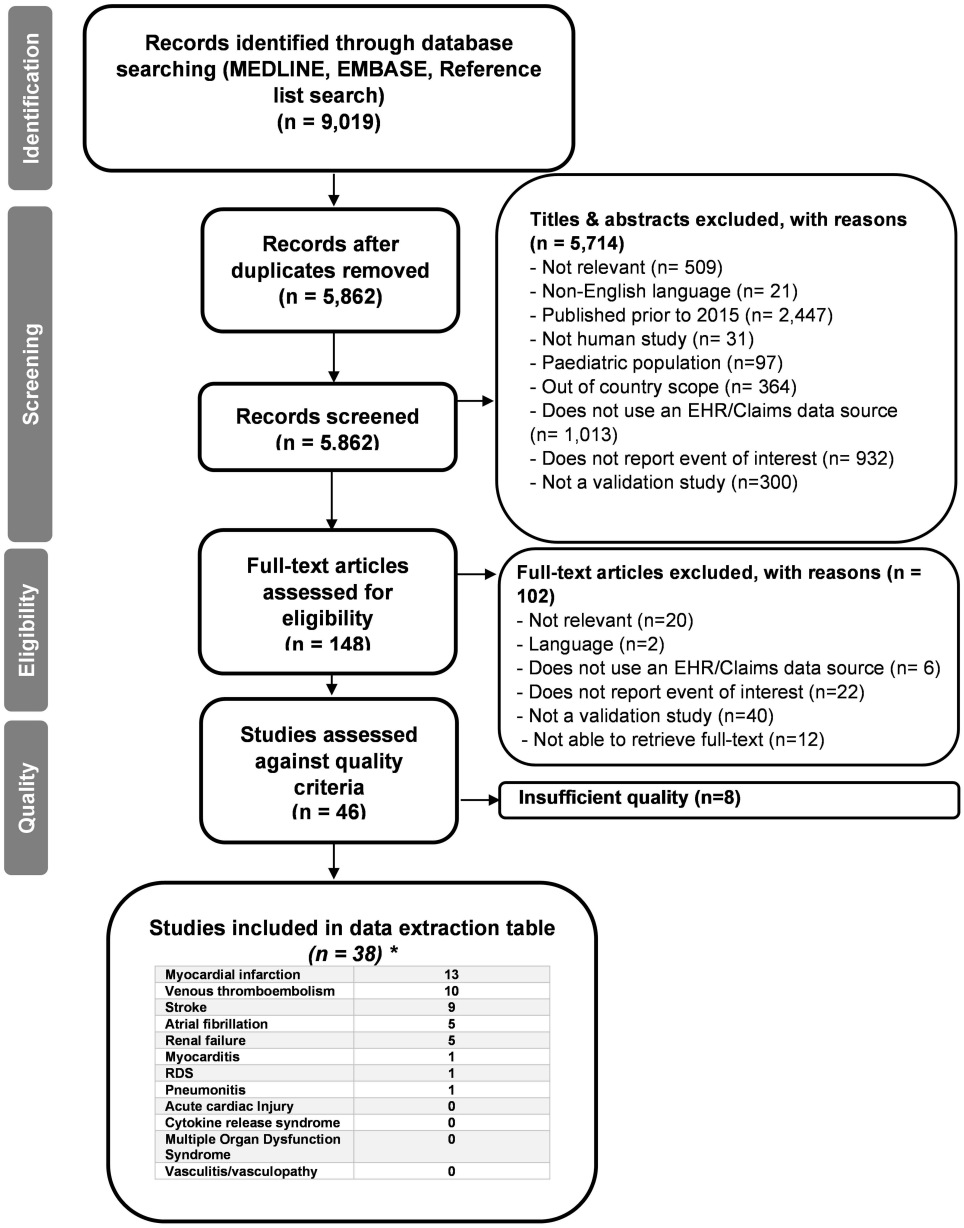

Figure 1 Study flowchart.

Notes: Individual outcomes count sums to more than the total as five studies (Sundbøll et al,[8] Ammann et al,[9] Arana et al,[10] Dalsgaard et al[11] and Psaty et al[12]) provide information for multiple outcomes.

https://doi.org/10.2147/DHPS.S415292 DovePress

Table 1 Characteristics of Selected Studies for Cardiovascular Outcomes

OUTCOME	AUTHOR (YEAR)	DATABASE (COUNTRY)	SETTING (HEALTHCARE SETTING & TYPE OF DATA SOURCE)	CODING DICTIONARY	ALGORITHM DEFINITION	GOLD STANDARD	PPV
Myocardial infarction	Ammann et al (2018)[9]	Sentinel Distributed Database (USA)	Claims (secondary care)	ICD-9-CM	Codes in any diagnostic field	Medical record review	MI: 75%
Myocardial infarction	Arana et al (2020)[10]	Clinical Practice Research Datalink (UK)	EHR data (primary care)	Read codes	Read codes for myocardial infarction	GP questionnaire	MI: 98%
Atrial fibrillation	Ashburner et al (2017)[14]	Primary Care Practice Based Research Network at Massachusetts General Hospital (USA)	EHR (primary care)	ICD-9 & ICD-10	At least one code plus at least one problem list entry or two codes within the previous 3 years.	Medical record review	MI: 96%
Myocardial infarction	Brouwer et al (2015)[15]	North Carolina Medicaid (USA)	Claims (secondary care)	ICD-9-CM	Codes in any diagnostic field	Registry data	MI: 44%
Myocardial infarction	Bush et al (2018)[16]	Medicare and Medicaid Services (USA)	EHR data (primary care)	ICD-9	Codes in any diagnostic field	Registry data	MI: 67%
Myocardial infarction	Colantonio et al (2019)[17]	Medicare (USA)	Claims (secondary care)	ICD-9	Codes in a primary diagnosis field; Codes in any diagnosis field	Registry data	MI (primary position):90%; MI (any field):84%
Myocardial infarction	Cozzolino et al (2019)[18]	Centralised administrative database of the Umbria Region (Italy)	EHR data (secondary care)	ICD-9	Codes in a primary diagnosis field	Medical record data	MI: 95%
Myocardial infarction	Dalsgaard et al (2019)[11]	Danish National Patient Register (Denmark)	EHR (secondary care)	ICD-10	Codes in a primary diagnosis field	Medical record review	MI: 75%
Myocardial infarction	Di Chiara et al (2019)[19]	Clinical-administrative databases of the Health Information System of Friuli Venezia Giulia (Italy)	EHR (secondary care)	ICD-9	Codes in a primary diagnosis field	Troponin measure	MI:96%
Atrial fibrillation	Ding et al (2019)[20]	MIMIC-III database (USA)	EHR (secondary care)	ICD-9	Codes in any diagnostic field	Rhythm assessment	AF: 79%
Myocardial infarction	Floyd et al (2016)[21]	Veterans' Health Administration (USA)	Claims (secondary care)	ICD-9	Codes in any diagnostic field	Medical record review	MI:80%
Myocardial infarction	Govatsmark et al (2020)[22]	Norwegian Myocardial Infarction Register (Norway); Norwegian Patient Register	EHR (secondary care)	ICD-10	Codes in a primary diagnosis field	Medical record review	MI (MI register):98%; MI (NPR) data):95%

Myocardial infarction	Psaty et al (2016)[12]	Centre for Medicare and Medicaid Services (USA)	Claims (secondary care)	ICD-9	Codes in a primary diagnosis field; Codes in any diagnosis field	Registry data	MI (primary position):91%; MI (any field):70%
Myocardial infarction; Atrial fibrillation; Bradycardia; Ventricular tachycardia; Myocarditis	Sundbøll et al (2016)[8]	Danish National Patient Register (Denmark)	EHR (secondary care)	ICD-8 & ICD-10	Codes in a primary diagnosis field	Medical record review	MI: 97%; AF: 95%; Bradycardia: 87%; Ventricular tachycardia: 80%; Myocarditis: 64%
Atrial fibrillation	Tu et al (2016)[23]	Health administrative data holdings for the province of Ontario held at ICES (Canada)	EHR (secondary care)	ICD-9 & ICD-10	Codes in any diagnostic field	Medical record review	AF:71%
Atrial fibrillation	Wei et al (2016)[24]	Vanderbilt University Medical Centre (USA)	EHR (secondary care)	ICD-9	At least one code; ≥2 codes	Medical record review	AF (≥2 codes): 88%; AF (1 code): 72%
Myocardial infarction	Youngson et al (2016)[25]	The Discharge Abstract Database (Canada)	EHR (secondary care)	ICD-10	Codes in a primary diagnosis field	Registry data	Not reported

https://doi.org/10.2147/DHPS.S415292

Table 2 Characteristics of Selected Studies for Pulmonary Conditions

OUTCOME	AUTHOR (YEAR)	DATABASE (COUNTRY)	SETTING (HEALTHCARE SETTING & TYPE OF DATA SOURCE)	CODING DICTIONARY	ALGORITHM DEFINITION	GOLD STANDARD	PPV
RDS	Kerchberger et al (2020)[26]	Validating biomarkers in Acute Lung Injury for Diagnosis (VALID) study (USA)	Electronic medical records (EMR) data (secondary care)	Unknown	Diagnostic billing and procedural codes. Structured data elements indicating presence of an ARDS risk factor, acute hypoxemic respiratory failure with mechanical ventilation, and oxygenation and expert interpretation of chest radiograph	Physician review	RDS:46%
Pneumonitis	Juurlink et al (2006)[7]	Discharge abstract database (Canada)	EMR data (secondary care)	ICD-10	Codes in the primary diagnosis position/ Codes in any diagnostic field	Medical chart review	Pneumonitis: 72% (primary position); 67% (any diagnostic field)

https://doi.org/10.2147/DHPS.S411292 DovePress

To note, the PPV for pneumonitis was reduced when using codes in any diagnosis field. Meanwhile, no validation studies were retrieved for ACI, vasculitis, CRS or MODS.

The quality assessment of the 38 articles is available in Table 3. The majority of studies did not report if the results of the index test and reference test were interpreted without prior knowledge of the results of the other test, thus were categorised as "unclear" (Q10 and Q11).

Cardiovascular Outcomes

Myocardial Infarction

Thirteen validation studies were found for MI (see Table 1 for study details) with the majority (62% [8/13]) conducted in EMR data sources. Validation studies were more frequently conducted in data sources covering the secondary care (hospital) setting; only one study[10] was conducted in primary care. The most common reference standard used was medical record review (n=6 studies),[8,9,11,18,21,22] though other methods included validation against registry data (n=4),[12,15–17,] GP questionnaires (n=1)[10] and troponin measurements (n=1).[19] The most common International Statistical Classification of Diseases and Related Health Problems (ICD) codes reported were ICD-9 codes 410.x and ICD-10 codes I21.x–I.24.x; all studies reported at least one of these codes. Some studies excluded 410.x2 (ICD-9) and I22.x–I24.x (ICD-10) which denote codes for MI in a subsequent episode of care, complications of MI or other ischaemic heart diseases.[9,16,17,19] The exclusion did not seem to have a significant impact on the PPVs.

MI was accurately determined using primary care data in the UK (PPV=98%).[10] In the hospital setting, algorithms that used codes in the primary or secondary diagnosis field[8,11,12,17–19,22,25] reported consistently higher PPV's (PPV range 75%–98%) than those that used codes in any field (PPV range 44%–84%).[9,12,15–17,21]

Arrhythmia

Five selected studies validated algorithms for identifying atrial fibrillation (AF),[8,14,20,23,24] with one study additionally providing PPVs for bradycardia and ventricular tachycardia of 80% and 87%, respectively[8] (see Table 1 for details). All studies were conducted in EMR databases and all but one[14] were conducted in secondary care settings. For AF, the PPVs of all studies were above 70%. Generally, algorithms that used more than one instance of a code for AF had the highest PPVs. The highest PPV was recorded in primary care data (PPV=96%) in the Primary Care Practice Based Research Network at Massachusetts General Hospital (USA). AF was assigned with at least one ICD-9/10 code and one problem list entry term, or two ICD-9/10 codes within three years.[14] The highest PPV in secondary care was reported in the Danish National Patient Register (DNPR) using codes in the primary or secondary field (PPV=95%).[8]

Coagulopathy Outcomes

Venous Thromboembolism

Of the 10 selected studies, only one study used healthcare insurance claims data,[29] the remainder used EMR data (see Table 4 for study details). All studies (n=10) used review of medical records as the reference standard. All but one study assessed the validity of ICD-9/ICD-10 in a hospital or emergency setting, the exception was conducted in primary care data in the UK using Read v.2. codes.[41] All studies based on ICD codes used ICD-9/ICD-10 (415.x/I26.x) codes to identify pulmonary embolism (PE), and 451.x and 453.x/I80.x and I82.x to identify deep-vein thrombosis (DVT). Studies assessing the validity of VT used a combination of PE and DVT codes. To note, there were minor variations in the specific subcodes used between the studies.

A total of eight studies provided algorithms for identifying PE.[8,27–30,33,38,39] Algorithms using codes in the primary or secondary field reported PPV's >85%[8,30,38,39] while those using codes in any field reported lower PPV values (PPV's <85%),[27–30] with one exception (PPV=97%).[33] The latter may have been driven by the requirement for a procedure code for additional diagnostic imaging in the emergency department.[33] The lowest PPV was observed using codes in any diagnostic field in an outpatient setting (PPV=28%).[30]

DVT and VT algorithms had lower PPV's than PE. The highest PPV for DVT was observed in the Danish National Patient Register (PPV=86%).[8] The highest PPV for identifying VT in North America was observed in the Cardiovascular Research Network Venous Thromboembolism (CVRN VTE) which used data from four integrated healthcare delivery

Table 3 QUADAS Quality Assessment

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	SCORE
Al-Ani et al (2015)[27]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	13
Alotaibi et al (2015)[28]	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	11
Ammann et al (2018)[29]	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	10
Arana et al (2020)[10]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	12
Ashburner et al (2017)[14]	No	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	11
Brouwer et al (2015)[15]	No	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	10
Bush et al (2018)[16]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Unclear	Yes	Yes	Yes	12
Colantonio et al (2019)[17]	No	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	10
Cozzolino et al (2019)[18]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	12
Dalsgaard et al (2019)[11]	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	10
Di Chiara et al (2019)[19]	Yes	Yes	Yes	Yes	No	Unclear	Yes	Yes	Yes	Unclear	No	Yes	Yes	Yes	10
Ding et al (2019)[20]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	13
Fang et al (2017)[30]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	12
Floyd et al (2016)[21]	No	Unclear	Yes	Yes	Yes	No	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	9
Govatsmark et al (2020)[22]	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	Unclear	Unclear	Yes	Yes	Yes	9
Hall et al (2016)[31]	No	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Yes	Yes	Yes	9
Juurlink et al (2006)[7]	Yes	No	No	Yes	No	Yes	Yes	No	Yes	Unclear	Unclear	Yes	Yes	Yes	8
Kerchberger et al (2020)[26]	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Unclear	Unclear	Yes	Yes	Yes	8
Kivimaki et al (2017)[32]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	12
Klil-Drori et al (2019)[33]	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12
Koola et al (2018)[34]	No	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	No	Yes	Yes	Yes	10
Logan et al (2019)[35]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	11
Lowenstern et al (2019)[36]	No	Yes	Yes	Yes	Unclear	Unclear	Yes	No	Yes	Yes	Unclear	Yes	Yes	Yes	8
Luhdorf et al (2017)[37]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes	10

https://doi.org/10.2147/DHPS.S415292 DovePress

Ohman et al (2018)[38]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	13
Prat et al (2018)[39]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	12
Psaty et al (2016)[12]	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	No	Unclear	Yes	Yes	Yes	Yes	10
Rebholz et al (2016)[40]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	13
Ruigómez et al (2020)[41]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	12
Sanfilippo et al (2015)[42]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	13
Strom et al (2019)[43]	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	10
Sundbøll et al (2016)[8]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	13
Tu et al (2016)[23]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	12
Van Walraven et al (2018)[44]	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	11
Varmdal et al (2015)[45]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	12
Wei et al (2016)[24]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	11
Xie et al (2018)[46]	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	11
Youngson et al (2016)[25]	No	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Yes	Yes	Yes	9

https://doi.org/10.2147/DHPS.S415292

Table 4 Characteristics of Selected Studies for Coagulopathy Outcomes

OUTCOME	AUTHOR (YEAR)	DATABASE (COUNTRY)	SETTING (HEALTHCARE SETTING & TYPE OF DATA SOURCE)	CODING DICTIONARY	ALGORITHM DEFINITION	GOLD STANDARD	PPV
DVT; PE	Al-Ani et al (2015)[27]	Canadian Institute for Health Information National Ambulatory Care Reporting System (Canada)	EHR data (secondary care)	ICD-10	Codes in any diagnostic field	Medical record review	DVT: 42%PE: 56%
DVT; PE; VT	Alotaibi et al (2015)[28]	University of Alberta Hospital radiology database (Canada)	EHR data (secondary care)	ICD-9-CM & ICD-10	Codes in any diagnostic field	Medical record review	DVT: 78%PE: 71%VT: 73%
PE	Amman et al (2018)[29]	Sentinel Distributed Database (USA)	Claims (secondary care)	ICD-9-CM	Codes in any diagnostic field	Medical record review	PE: 61%
PE; VT	Fang et al (2017)[30]	Cardiovascular Research Network Venous Thromboembolism study (USA)	EHR data (secondary care)	ICD-9	Codes in the primary diagnosis position in the inpatient setting; Code in any diagnostic field in the outpatient setting	Medical record review	PE: 89% (primary diagnosis); 28% (outpatient)VT: 79% (primary diagnosis); 31% (outpatient)
PE	Klil-Drori et al (2019)[33]	Ambulatory Care Database of the Alberta Health Services Calgary Zone (Canada)	EHR data (secondary care)	ICD-10	Codes in any diagnostic field	Medical record review	PE: 97%
DVT; PE; VT	Ohman et al (2018)[38]	Swedish National Patient Register (Sweden)	EHR data (secondary care)	ICD-9 & ICD-10	Codes in primary or secondary diagnosis position	Medical record review	DVT: 54%PE: 86%VT: 71%
PE	Prat et al (2019)[39]	Programme de Medicalisation des d'Information (France)	EHR data (secondary care)	ICD-10	Codes in the primary diagnosis position	Medical record review	PE: 99%
VT	Sanfilippo et al (2015)[42]	Veterans Affairs Central Cancer Registry (USA)	EHR data (secondary care)	ICD-9	Codes in any diagnostic field; Codes in any diagnostic field plus procedure codes for treatment or death; Codes in any diagnostic field plus procedure codes for treatment or death plus evidence of a diagnostic study	Medical record review	VT: 72% (only ICD codes); VT: 91% (ICD codes + treatment/death); VT: 92% (ICD codes + treatment/death + diagnostic study).
DVT; PE; VT	Sundbøll et al (2016)[8]	Danish National Patient Register (Denmark)	EHR data (secondary care)	ICD-8 & ICD-10	Codes in the primary or secondary diagnostic position	Medical record review	DVT: 86%PE: 90%VT: 88%
VT	Ruigómez et al (2020)[41]	The Health improvement network (UK)	EHR data (primary care)	Read codes	Coded entry for VT	Medical record review	VT: 40%

https://doi.org/10.2147/DHPS.S415292 DovePress

systems in Canada (PPV=79%).[30] Both studies identified codes in a primary diagnosis field. The PPVs increased when using additional procedure codes that provided evidence of treatment, death or evidence of a diagnostic procedure to identify VT (PPV=92%).[42] Low accuracy for the identification of VT in a primary care was reported, with a PPV of 40%.[41] For both DVT and VT, there were no discernible trends in PPVs between algorithms, which used codes in the primary/secondary diagnosis field versus any diagnosis field.

Cerebrovascular Conditions

Nine studies were selected that validated stroke (see Table 5 for study details). Of these, two studies validated ischaemic stroke[10,43] and the remaining general stroke.[11,12,31,32,37,45,46] Three studies were conducted in claims databases.[12,43,46] The validity of identifying ischaemic stroke with codes in any diagnosis field had a high PPV in both primary care and secondary care settings, with PPV's ≥85%.[10,43] The PPVs for general stroke varied, depending on the definition ranging from 44% to 99%.[11,12,31,32,37,45,46] ICD-9/ICD-10 codes for general stroke included 430.x, and 431.x/I60.x and I61.x (haemorrhagic stroke); 433.x and 434.x/I63.x (ischaemic stroke), and 436.x/I64.x (ill-defined stroke). The lowest PPVs were observed in studies where clinical trial or registry data were used as the reference standard to define general stroke: The PPVs for studies using medical record review ranged from 99% to 70% (median 80%),[11,37,45] while those using clinical trial or registry data as the reference ranged from 80% to 40% (median 72%).[12,31,32,43,46]

Renal Conditions

Five studies validated renal failure (see Table 6 for study details). The reference standards used were creatinine measures (n=2), medical record review (n=2) and GP questionnaire (n=1). The selected codes varied according to the definition of renal failure. An algorithm identifying renal failure using unspecific codes (ICD-9 code 586.x and ICD-10 code N19.0) had the highest PPV of 70%.[40] The PPV's for identifying hepatorenal syndrome (a form of kidney impairment that occurs in individuals with severe liver disease) were also high at 79%.[34] The PPV's for identifying renal failure (defined using both AKI and chronic kidney disease codes) in claims data were very low at 13%.[36]

Discussion

We identified 38 studies that validated at least one of our outcomes of interest. Validated algorithms are available to accurately identify cardiovascular and thrombotic events in routine health data such as PE, MI, AF, DVT and stroke where interest transcends multiple therapeutic research areas. The PPVs for renal failure were highly variable and depended on the definitions used. Validation studies for other events of interest, including those more specific to COVID-19 research, were less readily available (myocarditis, pneumonitis, RDS) or not available (ACI, MODS, CRS).

We updated the literature search for ACI, myocarditis, vasculitis, RDS, pneumonitis, CRS, and MODS in July 2023, to understand if any of the research gaps had been filled once the vaccines and therapeutics had been more widely dispersed among the population. We found that since our initial review diagnostic algorithms for myocarditis had been validated in EMR databases in Sweden[47] in the context of Covid-19 and found to be of acceptable quality (PPV Sweden: 96%). No further validation studies of diagnostic algorithms in EMR were found for the other safety outcomes of interest.

For events with validation studies available, some patterns emerged: algorithms for immediate life-threatening conditions such as MI or PE benefit from using codes within the primary and secondary diagnosis fields of the discharge summary in hospital settings. These fields are used to define the condition of admission during the relevant episode of healthcare. Meanwhile, conditions which often have a chronic presentation, such as AF benefit from using more than one instance of a code. DVT and stroke frequently occur as complications of care for other conditions,[48] as such, the use of any diagnosis field increases the sensitivity of the algorithm if it is not the primary reason for admission, is a comorbidity or event occurs during the hospitalisation.

Secondary care settings are optimal for identifying the outcomes in this study. Primary care data can be used to accurately identify MI, stroke and atrial fibrillation, though low accuracy was reported when identifying VT among anticoagulant users.[41] VT is increasingly managed in an outpatient setting so may not be included in primary or inpatient secondary care datasets. However, there may be variation by country; hence, knowledge of a countries' healthcare system and differences in the patient diagnostic and treatment journey are crucial when selecting data sources and defining outcomes of study.

https://doi.org/10.2147/DHPS.S415292 DovePress

Table 5 Characteristics of Selected Studies for Cerebrovascular Outcomes

OUTCOME	AUTHOR (YEAR)	DATABASE (COUNTRY)	SETTING (HEALTHCARE SETTING & TYPE OF DATA SOURCE)	CODING DICTIONARY	ALGORITHM DEFINITION	GOLD STANDARD	PPV
Ischaemic stroke	Arana et al (2020)[10]	Clinical Practice Research Datalink (UK)	EHR data (primary care)	Read codes	Codes for ischaemic stroke	GP questionnaire	Ischaemic stroke: 95%
Stroke	Dalsgaard et al (2019)[11]	Danish National Patient Register (Denmark)	EHR data (secondary care)	ICD-10	Codes in a primary diagnosis position	Medical record review	Stroke: 70%
Stroke	Hall et al (2016)[31]	National Ambulatory Care Reporting System (Canada)	EHR data (secondary care)	ICD-10-CA	Codes in any diagnostic field	Registry data	Stroke: 69%
Stroke	Kivimaki et al (2017)[32]	Hospital Episode Statistics (UK)	EHR data (secondary care)	ICD-10	Codes in a primary or secondary diagnosis position	Registry data	Stroke: 72%
Stroke	Luhdorf et al (2017)[37]	Danish National Patient Register (Denmark)	EHR data (secondary care)	ICD-8 & ICD-10	Codes in a primary or secondary diagnosis position	Medical record review	Stoke: 70%
Stroke	Psaty et al (2016)[12]	Centre for Medicare and Medicaid Services (USA)	Claims (secondary care)	ICD-9 CM	Codes in the primary diagnosis position/ Codes in any diagnostic field	Registry data	Stroke (primary): 80%; Stroke (any): 44%
Ischaemic stroke; transient ischaemic attack	Strom et al (2019)[43]	Medicare Provider Analysis and Review (USA)	Claims (secondary care)	ICD-9 & ICD-10	Codes in any diagnostic field	Clinical trial data	Ischaemic stroke:99% (ICD-10), 87% (ICD-9); Transient ischaemic attack: 75% (ICD-10), 40% (ICD-9)
Stroke	Varmdal et al (2015)[45]	National stroke Register & Norwegian National Patient Register (Norway)	EHR data (secondary care)	ICD-10	Codes in any diagnosis position in the Stroke register; Codes in any diagnosis position in the Norwegian national patient register; Codes in the primary diagnosis fields in the Norwegian national patient register	Medical record review	Stroke (stroke registry): 99%; Stroke (any diagnosis): 94%; Stroke (main diagnosis): 80%
Stroke	Xie et al (2018)[46]	Medicare (USA)	Claims (secondary care)	ICD-9	Codes in any diagnostic field	Registry data	Stroke: 77%

Table 6 Characteristics of Selected Studies for Renal Conditions

OUTCOME	AUTHOR	DATABASE (COUNTRY)	SETTING (HEALTHCARE SETTING & TYPE OF DATA SOURCE)	CODING DICTIONARY	ALGORITHM DEFINITION	GOLD STANDARD	PPV
HRS	Koola et al (2018)[34]	Veterans Information Systems and Technology Architecture/Computerized Patient Record System (ViSTa/CPRS) (USA)	EHR data (All settings)	ICD-9	Code at discharge; Code at any time in the inpatient stage	Medical record review	HRS: 79% (at discharge); HRS: 76% (at any time)
AKI	Logan et al (2019)[35]	NHS Tayside (UK)	EHR data (All settings)	ICD-10	Code in any diagnosis position	Creatinine values	Not reported
Renal failure	Lowenstern et al (2019)[36]	Medicare (USA)	Claims (secondary care)	ICD-9	Code in any diagnosis position	GP criteria	Renal failure: 13%
Renal failure	Rebholz et al (2016)[40]	Atherosclerosis Risk in Communities (ARIC) Study. (USA)	EHR data (secondary care)	ICD-9	Code in any diagnosis position	Medical record review	Renal failure: 70%
Severe renal failure	van Walraven et al (2018)[44]	Not reported (USA)	EHR (not reported)	ICD-10	Code in any diagnosis position	Creatinine values	Renal failure: 60%

https://doi.org/10.2147/DHPS.S415292 DovePress

There were fewer validation studies conducted in healthcare insurance claims data compared to EMR data sources. Furthermore, the use of different reference standards (medical records vs registry/clinical trial data) seemed to influence the study results. Recent FDA guidance focuses primarily on medical record review as a reference standard for the validation of outcomes in real-world evidence.[49] This is in line with the findings of our review, we found studies using clinical trial/registry data had lower PPVs than those using medical record review. The lack of healthcare insurance claims data validation studies may reflect a differential ability to link back to medical records versus EMR data sources. Therefore, to increase outcomes validation information additional strategies may be needed to increase validation opportunities for use with insurance claims data such as increased data linkages to EMR or registry data to allow cross validation.

Additional barriers to the generation of outcomes validation data may link to limitations in diagnostic coding guidance. For example, ICD coding guidelines explicitly recommend that ACI should be coded under the umbrella "ill-defined heart diseases" codes[50] decreasing the probability of high performing diagnostic PPVs based on these diagnostic codes. Similarly, coding guidelines indicate that MODS should be coded using the specific organ failure codes with no code available to specify "MODS" as a condition. The 2016 International Consensus Definitions for Sepsis and Septic Shock includes "life threatening organ dysfunction" as part of the definition of sepsis.[51] However, caution should be exercised as numerous sepsis validation studies agree that this condition is under recorded in administrative data.[52] In these cases where coding guidance is more non-specific, routine healthcare data may not be an optimal approach for the generation of benefit risk data.

Strengths and Limitations

The main strength of this study was the comprehensive collection of the most relevant AESIs for COVID-19 research in line with the SPEAC, or key effectiveness endpoints for therapeutics in the context of severe COVID-19.[5]

We acknowledge some limitations to our literature review. There is no standard term for "routine health data" thus it is not well catalogued in MEDLINE and EMBASE (eg no MeSH term). The time frame used may have caused us to miss relevant references. This may have been especially the case for more established data sources, such as the Clinical Research Practice Datalink (CPRD) or the DNPR which have been collecting data for over 30 years.[53,54] Nonetheless, we have identified various validation studies conducted in these data sources within our specified timeframe.[8,10,11,22,38] We used data to provide a more accurate picture of how the diagnostic codes and algorithms fare in the landscape of an emerging pandemic prior to treatments becoming available to understand the capacity of electronic health records to capture emerging safety signals. Coding practices may have changed over time due to the implementation of new coding classifications such as the change from ICD-9 to ICD-10 in many databases, or due to changes in the delivery and recording of care. An example of the latter is the Quality Outcome Frameworks (QOF), a GP incentive program established in the UK to improve the quality of care. Another limitation is language bias, which may account for the fact that English-speaking countries are overrepresented in the data. Therefore, our findings represent one point in time as well as variations across healthcare systems and geographies for that point in time.

Most studies did not provide patient characteristics, the safety and effectiveness of treatments may be affected by different patient level factors.[55–57] Routine health monitoring is vital for high-risk individuals and as such, the validity of diagnostic algorithms should also be tested in different patient subgroups.

The lack of reporting of disease prevalence of the source population may have influenced the PPVs and hindered the generalisability of the disease algorithms in other populations. Furthermore, specificity and sensitivity are important measures of validation; however, due to the difficulty and resources needed to determine a reference population in routine healthcare data, few studies presented these measures. Also, the use of different gold standards seemed to influence the study results, thus making it difficult to compare validation statistics between diagnostic codes and algorithms using different reference standards. Given the varying PPVs, even for well-studied conditions, assessment is needed of the generalisability of an existing algorithm for a given event in a different data source, the degree of adaptation needed and addressing the influence of misclassification on the results with sensitivity analyses.

Conclusion

Validated diagnostic codes and algorithms are available to identify VT, MI, stroke and atrial fibrillation in routine health data. There is some evidence to support the identification of arrhythmias, pneumonitis, myocarditis and AKI within routine health data albeit with lower accuracy. These primarily reflect secondary healthcare settings. Nonetheless, when designing a study assessing medicinal/biological product benefit risk in the real world a number of factors should be considered when selecting an appropriate set of diagnostic codes given the wide range of PPVs often reported. These include diagnosis code field, minimum number of diagnostic codes required, coding guidance, source of validation data (medical record/registry/clinical trial data) as well as the characteristics of the disease, type of variable (outcome/exposure/covariate) to be identified and the type of data source and healthcare setting.

Emerging scientific areas, requiring development of new diagnostic codes and/or new outcomes of interest, highlight the importance of embedding validation within routine health data studies. Potential barriers to validation work should be better understood and solutions considered as this information is critical to evaluate the robustness of emerging evidence for safety and effectiveness of COVID-19 therapeutics and vaccines using routine health databases. Given increasing emphasis on outcome validation (49), increasing opportunities and methods for validation work is critical. This could be through regulatory frameworks setting clear expectations for validation work as well as facilitating additional data linkage opportunities that may offer an alternative to medical record review.

Ethics and Consent Statements

The authors state no ethical approvals or informed consent was needed for this review.

Acknowledgments

An abstract of this work has been accepted to ISPE's 38th International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE 2022). This study was conducted by OXON Epidemiology, a Contract Research Organisation, funded by GlaxoSmithKline.

Funding

This study was funded by GlaxoSmithKline in collaboration with Vir Biotechnology.

Disclosure

KA, MH-C, BP & NQ were full- or part-time employees of OXON. MC was an employee at GSK at the time the project was executed. MD is a full-time GSK employee and hold share in GSK. The authors report no other conflicts of interest in this work.

References

1. Choudhary OP, Singh I. Making sound public health policy decisions for COVID-19 vaccination: vaccine effectiveness, safety, affordability, programmatic logistics and roll-out globally. *J Travel Med*. 2021. doi:10.1093/jtm/taab031
2. Cadarette SM, Wong L. An introduction to health care administrative data. *Can J Hosp Pharm*. 2015;68(3):232–237. doi:10.4212/cjhp.v68i3.1457
3. Sauer CM, Chen L-C, Hyland SL, Girbes A, Elbers P, Celi LA. Leveraging electronic health records for data science: common pitfalls and how to avoid them. *Lancet Digital Health*. 2022;4(12):e893–e898. doi:10.1016/S2589-7500(22)00154-6
4. Willame C, Dodd C, Duran CE, et al. Background rates of 41 adverse events of special interest for COVID-19 vaccines in 10 European healthcare databases - an ACCESS cohort study. Research Support, Non-U.S. Gov't. *Vaccine*. 2023;41(1):251–262. doi:10.1016/j.vaccine.2022.11.031
5. Law B, Sturkenboom M. Priority list of adverse events of special interest: COVID-19. Available from: https://brightoncollaboration.org/priority-list-of-adverse-events-of-special-interest-covid-19/. Accessed 14 December 2023.
6. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25. doi:10.1186/1471-2288-3-25
7. Juurlink D, Preyra C, Croxford R, et al. Canadian Institute for Health InformationDischarge Abstract Database: a Validation Study; 2006. Available from: https://www.ices.on.ca/publications/research-reports/canadian-institute-for-health-information-discharge-abstract-database-a-validation-study/. Accessed 14 December 2023.
8. Sundbøll J, Adelborg K, Munch T, et al. Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study. *BMJ Open*. 2016;6(11):e012832. doi:10.1136/bmjopen-2016-012832

9. Ammann EM, Schweizer ML, Robinson JG, et al. Chart validation of inpatient ICD-9-CM administrative diagnosis codes for acute myocardial infarction (AMI) among intravenous immune globulin (IGIV) users in the Sentinel Distributed Database. *Pharmacoepidemiol Drug Saf*. 2018;27(4):398–404. doi:10.1002/pds.4398
10. Arana A, Margulis AV, Varas-Lorenzo C, et al. Validation of cardiovascular outcomes and risk factors in the Clinical Practice Research Datalink in the United Kingdom. *Pharmacoepidemiol Drug Saf*. 2020;30(2):237–247. doi:10.1002/pds.5150
11. Dalsgaard E-M, Witte DR, Charles M, Jørgensen ME, Lauritzen T, Sandbæk A. Validity of Danish register diagnoses of myocardial infarction and stroke against experts in people with screen-detected diabetes. *BMC Public Health*. 2019;19(1):228. doi:10.1186/s12889-019-6549-z
12. Psaty BM, Delaney JA, Arnold AM, et al. Study of Cardiovascular Health Outcomes in the Era of Claims Data. *Circulation*. 2016;133(2):156–164. doi:10.1161/circulationaha.115.018610
13. Kerchberger VE, Bastarache JA, Shaver CM, et al. Chest radiograph interpretation is critical for identifying acute respiratory distress syndrome patients from electronic health record data. American Journal of Respiratory and Critical Care Medicine. American Thoracic Society International Conference, ATS 2020. United States; 2020;201(1).
14. Ashburner JM, Singer DE, Lubitz SA, Borowsky LH, Atlas SJ. Changes in use of anticoagulation in patients with atrial fibrillation within a primary care network associated with the introduction of direct oral anticoagulants. *Am J Cardiol*. 2017;120(5):786–791. doi:10.1016/j.amjcard.2017.05.055
15. Brouwer ES, Napravnik S, Eron JJ, et al. Validation of medicaid claims-based diagnosis of myocardial infarction using an HIV clinical cohort. *Med Care*. 2015;53(6):e41–8. doi:10.1097/MLR.0b013e318287d6fd
16. Bush M, Sturmer T, Stearns SC, et al. Position matters: validation of medicare hospital claims for myocardial infarction against medical record review in the atherosclerosis risk in communities study. *Pharmacoepidemiol Drug Saf*. 2018;27(10):1085–1091. doi:10.1002/pds.4396
17. Colantonio LD, Levitan EB, Yun H, et al. Use of medicare claims data for the identification of myocardial infarction: the reasons for geographic and racial differences in stroke study. *Med Care*. 2018;56(12):1051–1059. doi:10.1097/MLR.0000000000001004
18. Cozzolino F, Montedori A, Abraha I, et al. A diagnostic accuracy study validating cardiovascular ICD-9-CM codes in healthcare administrative databases. The Umbria Data-Value Project. *PLoS One*. 2019;14(7):e0218919. doi:10.1371/journal.pone.0218919
19. Di Chiara A, Clagnan E, Valent F. Epidemiology and mortality in an Italian region after the adoption of the universal definition of myocardial infarction. Comparative study validation study. *J Cardiovasc Med*. 2020;21(1):34–39. doi:10.2459/JCM.0000000000000893
20. Ding EY, Albuquerque D, Winter M, et al. Novel method of atrial fibrillation case identification and burden estimation using the MIMIC-III electronic health data set. *J Intensive Care Med*. 2019;34(10):851–857. doi:10.1177/0885066619866172
21. Floyd JS, Blondon M, Moore KP, Boyko EJ, Smith NL. Validation of methods for assessing cardiovascular disease using electronic health data in a cohort of Veterans with diabetes. *Pharmacoepidemiol Drug Saf*. 2016;25(4):467–471. doi:10.1002/pds.3921
22. Govatsmark RE, Janszky I, Slordahl SA, et al. Completeness and correctness of myocardial infarction diagnoses in a medical quality register and an administrative health register. *Eur Heart J*. 2017;38(Supplement 1):1026. doi:10.1093/eurheartj/ehx502.P4894
23. Tu K, Nieuwlaat R, Cheng SY, et al. Identifying Patients With Atrial Fibrillation in Administrative Data. *Can J Cardiol*. 2016;32(12):1561–1565. doi:10.1016/j.cjca.2016.06.006
24. Wei W-Q, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inform Assoc*. 2016;23(e1):e20–e27. doi:10.1093/jamia/ocv130
25. Youngson E, Welsh RC, Kaul P, McAlister F, Quan H, Bakal J. Defining and validating comorbidities and procedures in ICD-10 health data in ST-elevation myocardial infarction patients. *Medicine*. 2016;95(32):e4554. doi:10.1097/md.0000000000004554
26. Kerchberger VE, Bastarache JA, Shaver CM, et al. Chest radiograph interpretation is critical for identifying acute respiratory distress syndrome patients from electronic health record data. Conference Abstract. American Journal of Respiratory and Critical Care Medicine Conference: American Thoracic Society International Conference, ATS; 2020;201(1).
27. Al-Ani F, Shariff S, Siqueira L, Seyam A, Lazo-Langner A. Identifying venous thromboembolism and major bleeding in emergency room discharges using administrative data. *Thromb Res*. 2015;136(6):1195–1198. doi:10.1016/j.thromres.2015.10.035
28. Alotaibi GS, Wu C, Senthilselvan A, McMurtry MS. The validity of ICD codes coupled with imaging procedure codes for identifying acute venous thromboembolism using administrative data. *Vasc Med*. 2015;20(4):364–368. doi:10.1177/1358863X15573839
29. Ammann EM, Cuker A, Carnahan RM, et al. Chart validation of inpatient International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) administrative diagnosis codes for venous thromboembolism (VTE) among intravenous immune globulin (IGIV) users in the Sentinel Distributed Database. *Medicine*. 2018;97(8):e9960. doi:10.1097/MD.0000000000009960
30. 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
31. Hall R, Mondor L, Porter J, Fang J, Kapral MK. Accuracy of administrative data for the coding of acute stroke and TIAs. *Can J Neurol Sci*. 2016;43(6):765–773. doi:10.1017/cjn.2016.278
32. Kivimaki M, Batty GD, Singh-Manoux A, Britton A, Brunner EJ, Shipley MJ. Validity of cardiovascular disease event ascertainment using linkage to UK hospital records. *Epidemiology*. 2017;28(5):735–739. doi:10.1097/EDE.0000000000000688
33. Klil-Drori AJ, Prajapati D, Liang Z, et al. External validation of ASPECT (Algorithm for Suspected Pulmonary Embolism Confirmation and Treatment). *Med Care*. 2019;57(8):e47–e52. doi:10.1097/MLR.0000000000001055
34. Koola JD, Davis SE, Al-Nimri O, et al. Development of an automated phenotyping algorithm for hepatorenal syndrome. *J Biomed Inform*. 2018;80:87–95. doi:10.1016/j.jbi.2018.03.001
35. Logan R, Davey P, De Souza N, Baird D, Guthrie B, Bell S. Assessing the accuracy of ICD-10 coding for measuring rates of and mortality from acute kidney injury and the impact of electronic alerts: an observational cohort study. *Clin Kidney J*. 2020;13(6):1083–1090. doi:10.1093/ckj/sfz117
36. Lowenstern A, Lippmann SJ, Brennan JM, et al. Use of medicare claims to identify adverse clinical outcomes after mitral valve repair. *Circ Cardiovasc Interv*. 2019;12(5):e007451. doi:10.1161/CIRCINTERVENTIONS.118.007451
37. Luhdorf P, Overvad K, Schmidt EB, Johnsen SP, Bach FW. Predictive value of stroke discharge diagnoses in the Danish National Patient Register. *Scand J Public Health*. 2017;45(6):630–636. doi:10.1177/1403494817716582
38. Ohman L, Johansson M, Jansson JH, Lind M, Johansson L. Positive predictive value and misclassification of diagnosis of pulmonary embolism and deep vein thrombosis in Swedish patient registries. *Clin Epidemiol*. 2018;10:1215–1221. doi:10.2147/CLEP.S177058
39. Prat M, Derumeaux H, Sailler L, Lapeyre-Mestre M, Moulis G. Positive predictive values of peripheral arterial and venous thrombosis codes in French hospital database. *Fundament Clinic Pharmacol*. 2018;32(1):108–113. doi:10.1111/fcp.12326

40. Rebholz CM, Coresh J, Ballew SH, et al. Kidney Failure and ESRD in the Atherosclerosis Risk in Communities (ARIC) Study: comparing ascertainment of treated and untreated kidney failure in a cohort study. *Am J Kidney Dis*. 2015;66(2):231–239. doi:10.1053/j.ajkd.2015.01.016
41. Ruigómez A, Brobert G, Vora P, García Rodríguez LA. Validation of venous thromboembolism diagnoses in patients receiving rivaroxaban or warfarin in the health improvement network. *Pharmacoepidemiol Drug Saf*. 2020. doi:10.1002/pds.5146
42. Sanfilippo KM, Wang T-F, Gage BF, Liu W, Carson KR. Improving accuracy of international classification of diseases codes for venous thromboembolism in administrative data. *Thromb Res*. 2015;135(4):616–620. doi:10.1016/j.thromres.2015.01.012
43. Strom JB, Zhao Y, Faridi KF, et al. Comparison of clinical trials and administrative claims to identify stroke among patients undergoing aortic valve replacement. *Circ Cardiovasc Interv*. 2019;12(11):e008231. doi:10.1161/circinterventions.119.008231
44. van Walraven C Improved correction of misclassification bias with bootstrap imputation. *Med Care*. 2018;56(7):e39–e45.
45. Varmdal T, Bakken IJ, Janszky I, et al. Comparison of the validity of stroke diagnoses in a medical quality register and an administrative health register. *Scand J Public Health*. 2016;44(2):143–149. doi:10.1177/1403494815621641
46. Xie F, Colantonio LD, Curtis JR, et al. Development of algorithms for identifying fatal cardiovascular disease in Medicare claims. *Pharmacoepidemiol Drug Saf*. 2018;27(7):740–750. doi:10.1002/pds.4421
47. Gedeborg R, Holm L, Feltelius N, et al. Validation of myocarditis diagnoses in the Swedish patient register for analyses of potential adverse reactions to COVID-19 vaccines. *Ups J Med Sci*. 2023;128. doi:10.48101/ujms.v128.9290
48. Heit JA. Epidemiology of venous thromboembolism. *Nat Rev Cardiol*. 2015;12(8):464–474. doi:10.1038/nrcardio.2015.83
49. FDA (Federal Drugs Agency). Real-World Data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products; 2021. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory. Accessed 14 December 2023.
50. Newhouser K. How to choose the right code for a non-traumatic myocardial injury. Fourth Universal Definition of Myocardial Infarction (UDMI) CDI & Coding Considerations for Myocardial Injury. Available from: https://www.medpartners.com/cdi-coding-considerations-for-myocardial-injury/. Accessed December 04, 2023.
51. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801–810. doi:10.1001/jama.2016.0287
52. Jolley RJ, Sawka KJ, Yergens DW, Quan H, Jetté N, Doig CJ. Validity of administrative data in recording sepsis: a systematic review. *Critical Care*. 2015;19(1):139. doi:10.1186/s13054-015-0847-3
53. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol*. 2015;44 (3):827–836. doi:10.1093/ije/dyv098
54. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490. doi:10.2147/CLEP.S91125
55. Notarte KI, Guerrero-Arguero I, Velasco JV, et al. Characterization of the significant decline in humoral immune response six months post-SARS-CoV-2 mRNA vaccination: a systematic review. *J Med Virol*. 2022;94(7):2939–2961. doi:10.1002/jmv.27688
56. Notarte KI, Ver AT, Velasco JV, et al. Effects of age, sex, serostatus, and underlying comorbidities on humoral response post-SARS-CoV-2 Pfizer-BioNTech mRNA vaccination: a systematic review. *Crit Rev Clin Lab Sci*. 2022;59(6):373–390. doi:10.1080/10408363.2022.2038539
57. Priyanka COP, Choudhary OP. Vaccine efficacy against COVID-19: a foresight on the host-associated factors. *J Formos Med Assoc*. 2021;120 (6):1405–1407. doi:10.1016/j.jfma.2020.11.021

Drug, Healthcare and Patient Safety Dovepress

Publish your work in this journal

Drug, Healthcare and Patient Safety is an international, peer-reviewed open-access journal exploring patient safety issues in the healthcare continuum from diagnostic and screening interventions through to treatment, drug therapy and surgery. The journal is characterized by the rapid reporting of reviews, original research, clinical, epidemiological and post-marketing surveillance studies, risk management, health literacy and educational programs across all areas of healthcare delivery. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/drug-healthcare-and-patient-safety-journal