

## Agreement of acute serious events recorded across datasets using linked Australian general practice, hospital, emergency department and death data: implications for research and surveillance

Sarah Ahmed<sup>1,\*</sup>, Allan Pollack<sup>1</sup>, Alys Havard<sup>2,3</sup>, Sallie-Anne Pearson<sup>3</sup>, and Kendal Chidwick<sup>1</sup>

### Submission History

Submitted:	12/10/2022
Accepted:	01/12/2022
Published:	24/01/2023

<sup>1</sup>NPS MedicineWise, c/- Wexed Advisors, Level 17, 68 Pitt street, NSW 2000, Sydney, Australia

<sup>2</sup>National Drug and Alcohol Research Centre, UNSW Sydney, NSW 2052, Sydney, Australia

<sup>3</sup>Medicines Intelligence Research Program, School of Population Health, Faculty of Medicine and Health, UNSW Sydney, NSW 2052, Sydney, Australia

### Abstract

#### Introduction

Understanding the level of recording of acute serious events in general practice electronic health records (EHRs) is critical for making decisions about the suitability of general practice datasets to address research questions and requirements for linking general practice EHRs with other datasets.

#### Objectives

To examine data source agreement of five serious acute events (myocardial infarction, stroke, venous thromboembolism (VTE), pancreatitis and suicide) recorded in general practice EHRs compared with hospital, emergency department (ED) and mortality data.

#### Methods

Data from 61 general practices routinely contributing data to the MedicinesInsight database was linked with New South Wales administrative hospital, ED and mortality data. The study population comprised patients with at least three clinical encounters at participating general practices between 2019 and 2020 and at least one record in hospital, ED or mortality data between 2010 and 2020. Agreement was assessed between MedicinesInsight diagnostic algorithms for the five events of interest and coded diagnoses in the administrative data. Dates of concordant events were compared.

#### Results

The study included 274,420 general practice patients with at least one record in the administrative data between 2010 and 2020. Across the five acute events, specificity and NPV were excellent (>98%) but sensitivity (13%–51%) and PPV (30%–75%) were low. Sensitivity and PPV were highest for VTE (50.9%) and acute pancreatitis (75.2%), respectively. The majority (roughly 70–80%) of true positive cases were recorded in the EHR within 30 days of administrative records.

#### Conclusion

Large proportions of events identified from administrative data were not detected by diagnostic algorithms applied to general practice EHRs within the specific time period. EHR data extraction and study design only partly explain the low sensitivities/PPVs. Our findings support the use of Australian general practice EHRs linked to hospital, ED and mortality data for robust research on the selected serious acute conditions.

#### Keywords

primary care; validation; electronic health records; MedicinesInsight; Australia; acute serious outcomes; linked data; linkage

\*Corresponding Author:

Email Address: [gomagoti@gmail.com](mailto:gomagoti@gmail.com) (Sarah Ahmed)

## Introduction

### Background

Electronic health record (EHR) systems are widely used in general practice to support clinical management [1–3]. They are a valuable source of information for surveillance, research and evaluation, providing large sample sizes with comprehensive clinical, sociodemographic and treatment information about patients, much of which is unavailable from other data sources [4, 5]. For these reasons the use of data from general practice EHRs for research and policy decisions is growing in Australia [6, 7]. It is important to explore the utility and limitations of general practice data as collections continue to develop [8].

Linking general practice with other routinely collected data, such as hospital admissions, emergency department (ED) and deaths, improves its utility for research, monitoring and surveillance. Limited such linkage has occurred to date in Australia. Technical and governance issues, while not insurmountable, take time and considerable resources to overcome [9]. A good understanding of the contribution of different datasets is important prior to linking them.

In Australia, for research involving conditions diagnosed, treated and managed largely within general practice, EHRs alone may be a good data source [10–12]. However, patients are often managed across multiple settings with important events and outcomes recorded across disparate sources. The extent to which Australian EHRs accurately reflect acute serious events is unclear. For example, acute cardiovascular events may require urgent care in hospitals or result in deaths, with missing or delayed general practice records. Primary care data from other countries has reasonable recording of acute events [13–16]. However, health care systems differ, and this question hasn't been examined specifically for Australia.

This study evaluated five serious health events – acute pancreatitis, myocardial infarction (MI), stroke, suicide and venous thromboembolism (VTE) – as recorded in a general practice EHR dataset (MedicineInsight) compared with administrative hospital, ED and mortality datasets. We also assessed timing of events which is important for some time-sensitive research questions. The selected events are acute and primarily managed in hospitals or EDs, with general practice presentations for ongoing care or discharge. They were selected as important events of interest for future research but may also serve as indicators of the quality of recording of acute conditions with similar clinical care pathways. Although useful as an external reference standard for the purposes of this study, hospital, ED and mortality data are not a “gold standard” for overall prevalence as they are subject to errors and incompleteness, such as missing events that occur outside the hospital system.

## Methods

### Study design

This was an observational study comparing the level of agreement for recording five acute serious events in general practice EHRs, hospital and mortality datasets. We compared routinely used MedicineInsight algorithms flagging these

events with information recorded in linked hospital, ED and mortality data.

### Data sources

MedicineInsight is a large-scale database established by NPS MedicineWise in 2011, containing de-identified EHRs from over 600 participating general practices across Australia [5]. It uses third-party data extraction tools [17, 18] which de-identify, extract and securely transmit data from the Best Practice (BP)<sup>™</sup> or Medical Director (MD)<sup>™</sup> clinical information systems, for harmonisation, cleaning and storage [5]. Extracted data include demographic and clinical entries by healthcare professionals. Identifying data and fields that may contain them including name, date of birth, address, progress notes and correspondence are not extracted. Certain variables are derived such as condition flags [5, 19]. Monthly extractions result in an updated longitudinal database where patients within each practice can be tracked over time. Previous research has examined the algorithms used to create medical conditions flags and recording of death in MedicineInsight [10, 20]. However, recording of acute serious conditions in MedicineInsight have not been validated.

The New South Wales (NSW) Admitted Patient Data Collection (APDC) is a compilation of episode-level records from all admitted patient services provided by NSW public and private hospitals, public psychiatric hospitals, public multi-purpose services, and private day procedures centres [21]. The variables used in this study include the dates of admission and separation, diagnoses coded by trained clinical information managers (International Classification of Diseases Australian Modification [ICD-10-AM]), procedures, and separation mode (discharge, transfer or death).

The NSW Emergency Department Data Collection (EDDC) provides information about patient presentations to the emergency departments (ED) of NSW public hospitals [21]. Information from private hospital and some smaller public hospital EDs were not available for linkage. The variables used in this study include admission date, and diagnosis codes (ICD10-AM, ICD10 and SNOMED-CT) recorded by medical, nursing or clerical personnel at the point of care. These personnel are not trained in clinical coding, noting symptoms are often selected as diagnoses.

The NSW Registry of Birth Deaths and Marriages (RBDM) death registrations contain fact of death information, date of birth, age at death, date of death and year of death registration. The Australian Coordinating Registry Cause of Death Unit Record File (ACR CODURF) provides both fact and cause of death information. Cause of death data is coded using the ICD-10 International Version, not ICD-10-AM. The variables used in this study include dates of birth, death and death registration as well as variables related to cause and place of death. Both RBDM and CODURF include deaths occurring in NSW and do not include deaths of NSW residents who die interstate [21].

The Centre for Health Record Linkage (CHeReL) [22] is a dedicated data linkage unit managed by the NSW Ministry of Health. CHeReL used a privacy preserving record linkage (PPRL) methodology [23] to link records dated between 2010 and 2020 from the NSW APDC, EDDC, RBDM and ACR CODURF [21] to MedicineInsight. CHeReL assigned Project

Person Numbers (PPNs) which identify individuals linked across datasets and are used to merge content data. Content data from each dataset was transferred to the Secured Unified Research Environment (SURE) [24] for storage and analysis. The study population was subset from this linked dataset, with analyses limited to encounters recorded during the study period.

## Study population and study period

### General practices

We identified 61 eligible NSW general practice sites participating in MedicinesInsight that used the INCA extraction tool (a prerequisite for enabling PPRL) in the February 2022 database build, from which eligible patients were selected for linkage by CHeReL.

### Patients

The study cohort was extracted from the 'Total MedicinesInsight Linkage Population' as detailed in Figure 1. The study cohort included regular patients with valid age (0 to 112 years) with at least 3 clinical encounters between 1 January 2019 and 31 December 2020 at an eligible MedicinesInsight practice in NSW which met the data quality requirements in February 2022 [5] and with at least one hospital, ED or mortality record between 1 January 2010 and 31 December 2020. Regular patients are defined as those who have at least three consultations in any 2 consecutive years, in accordance with the Royal Australian College of General Practitioner's (RACGP's) definition of 'active' patients [25]. MedicinesInsight is an open cohort and patients in Australia can visit multiple general practices, thus regular patients are often selected for analyses because they are more likely than infrequent attenders to be receiving most of their care at the MedicinesInsight practice, thereby enabling sufficient opportunities for diagnoses and risk factors etc. to be recorded.

The study period encompassed 1 January 2019 to 31 December 2020. Defining start and end of patient follow-up using data from Australian EHRs is challenging, therefore we assume regular patients were attending the MedicinesInsight practice for the entire study period.

## Outcome definitions

### Ascertaining serious acute events

As in most primary health care EHR databases, MedicinesInsight contains diagnostic algorithms [1] that use information from various fields to identify whether patients have specific conditions. These algorithms have been developed by NPS MedicineWise, the custodian of MedicinesInsight, to create efficiencies for users of the data and promote consistency between studies. These algorithms identify conditions using information from three diagnostic EHR fields (diagnosis, reason for visit and reason for prescription) containing either coded terms that the user selects from a drop-down list in the EHR software, or free text. The algorithms identify patients as having the specific condition if a coded term or text string from the pre-defined list has ever been recorded for that patient in any one of the three fields. The pre-defined list is compiled by

trained clinical coders and is based on available Pyefinch (used in BP) and Docle (used in MD) codes, as well as commonly accepted clinical definitions and abbreviations. Docle and Pyefinch are Australian general practice coding systems which consist of clinical terminologies for diseases, clinical findings and therapies [26]. For records identified by a free text string alone, the context in which it is recorded is reviewed by clinical coders at the time of developing the algorithm and periodically thereafter, and irrelevant instances removed. Other fields such as prescriptions or pathology are not searched. Data (including diagnoses) recorded in the unstructured area of the EHR, called 'progress notes', are not collected because they may contain identifiable information [5].

A detailed description of the MedicinesInsight algorithms for acute pancreatitis, MI, stroke (including transient ischaemic attack), suicide (attempted or completed) and VTE (including deep vein thrombosis and pulmonary embolism) is included in Supplementary Appendix 1. Cases in MedicinesInsight were patients flagged by MedicinesInsight diagnostic algorithms as having an event recorded during the two-year study time period (2019 to 2020). Multiple events during the study period were only counted once and the date of the earliest event within the study period was taken as the 'MedicinesInsight index date'.

The external reference standard was cases identified in either APDC, EDDC, RBDM or CODURF data during 2019 and 2020. Primary diagnoses codes were considered in the EDDC while both primary and additional diagnoses codes were considered in the APDC and CODURF. The ICD-10-AM or SNOMED-CT codes used to define each outcome are provided in Supplementary Appendix 2. For each serious acute outcome, multiple events within or across linked datasets during the study period were only counted once and the earliest event within the study period was used to define the 'reference index date'(RID).

## Assessment of agreement regarding fact of event

For each acute serious event we calculated percentage of agreement, sensitivity, specificity, negative and positive predictive values (NPV/PPV) [10, 27] of the MedicinesInsight algorithms compared with an external reference standard-during 2019 and 2020. The external reference standard was a composite of APDC, EDDC, RBDM and CODURF. An outcome recorded in any of these datasets was considered to have truly occurred.

## Assessment of agreement regarding timing of event

For the 'true positive' cases, we calculated differences between the MedicinesInsight index date and the reference index date. Results were presented in the following mutually exclusive categories where the MedicinesInsight index date was: 0–30 days after the reference index date, 1–30 days before the reference index date, and more than 30 days before or after the reference index date. The median, quartile 1 and quartile 3 differences between the MedicinesInsight and reference index dates are also presented (per event, rounded to whole days).

Median and quartiles are presented as the distribution of this difference is not normal.

Analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) via the SURE platform. As the data are clustered within practices, variance was adjusted to account for correlation between observations within clusters, and confidence intervals adjusted accordingly, using SURVEYFREQ and SURVEYMEANS procedures in SAS. For confidentiality, small cell numbers (1 to 4) were suppressed, or results aggregated to a higher level.

## Results

### Study Population (Figure 1, Table 2)

The study cohort comprised 274,420 regular patients from 61 general practices with at least one record in a NSW linked dataset between 2010 and 2020 (Table 1). The study cohort was similar to national estimates of Australian patients attending general practice during 2019-20 although it overrepresented older patients (11.2% aged 70–79), inner regional areas (25.3%) and the middle socioeconomic quintile (25.8%) [28]. National comparisons indicate 8.3% of patients aged 70–79, 12.3% from inner regional areas and 19.7% in the middle socioeconomic quintile [28].

During the 2-year study period (Table 2): for MI, there were 1,203 (0.4%) patients in MedicinesInsight with this condition vs 3,178 (1.2%) patients in the linked reference data; for stroke, 1,394 (0.5%) vs 2,359 (0.9%); for VTE, 1,390 (0.5%) vs 954 (0.3%); for acute pancreatitis, 105 (0.04%) vs 550 (0.2%); and for attempted or completed suicide, 583 (0.2%) vs 1,366 (0.5%).

The selection process for the 'Total MedicinesInsight Linkage Population' and study cohort is described in Figure 1.

### Agreement of acute serious events recorded in MedicinesInsight compared with NSW hospitalisation and mortality datasets (Table 3)

In general, sensitivity and PPV of MedicinesInsight algorithms were low across the selected acute serious events (Table 2). Sensitivity – or the number of patients that were identified by the MedicinesInsight algorithm as having the specific acute serious event, as a proportion of patients who truly had the specific event (according to linked datasets) – was highest for VTE (50.9%), followed by stroke (27.8%) and MI (24.3%). PPV – or the proportion of patients that truly had the specific event, from all patients identified by the MedicinesInsight algorithm as having the specific event – was highest for acute pancreatitis (75.2%), followed by MI (64.3%) and stroke (47.0%).

PoA, specificity and NPV were high across events because the majority of regular general practice patients during 2019-20 did not experience these acute serious events (Table 2).

### The timing of selected events between datasets (Table 3)

The majority of 'true positive' cases were first recorded in the EHR 0 to 30 days after they were recorded in hospital, ED

or mortality datasets (or after the reference index date (RID)). Of 773 'true positive' MI cases in MedicinesInsight 68.8% were recorded 0–30 days after the RID, 12.8% were recorded 1–30 days before the RID and 18.4% were recorded more than 30 days before or after the RID. Results were similar for acute pancreatitis and VTE, however a higher proportion of suicide (39.0%) and stroke (31.0%) cases were recorded more than 30 days before or after the RID. The lowest median differences between MedicinesInsight and reference index dates were for myocardial infarctions and strokes (median of 0 days difference for each). The difference in index dates was most broadly distributed for suicides, with an interquartile range of 37 days.

## Discussion

To our knowledge this is the first Australian study comparing the recording of these acute health events in general practice EHRs with linked hospital and mortality datasets. Outside Australia, studies validating algorithms within primary care EHRs report varying PPVs and sensitivities, with measures of agreement being generally lower for acute compared to chronic outcomes [29, 30]. Lower PPVs and sensitivities are also reported where reliance is on free text data (as opposed to where coding systems such as ICD-10 are used) [29].

Our study found Australian general practice EHR diagnostic algorithms for selected acute serious events do not correlate well with combined hospital and death records. For example, we found low sensitivities for VTE (50.9%), stroke (27.8%) and MI (24.3%) indicating a large proportion of patients experiencing these events were incorrectly classified as not having these events, either because the event was not recorded in the practice EHR within the specific time period or was not ascertained by the diagnostic algorithm. The use of these algorithms alone will lead to undercounting of patients with these acute conditions and is insufficient for most observational studies.

The proportion of patients identified by MedicinesInsight algorithms as having the specific acute event who truly had the event (according to linked data) was highest for acute pancreatitis (PPV 75.2%), followed by MI (64.3%) and stroke (47.0%). The high specificity and good PPV for acute pancreatitis indicate that this diagnostic algorithm in particular returns relatively few false positives and may therefore be useful for identifying cohorts of patients who truly have the specific event, but linkage with hospital data is required to find most cases. General practice EHRs contain detailed patient information on potential confounding factors unavailable in other administrative datasets. Using linked general practice, hospital, ED and mortality data is therefore recommended in research on acute serious events to accurately ascertain cases, build appropriate patient cohorts and more adequately control for confounding. Our findings suggest that general practice records will complement hospital, ED and mortality data, and that both are required for robust ascertainment of acute serious events. Overall, our findings support the use of general practice EHRs for research, monitoring or surveillance of the selected acute events, only when linked to hospital, ED and mortality data.

These results were not unexpected in the context of the Australian health setting, despite international validation



Figure 1: Selection of the MedicinesInsight linkage population and the study cohort

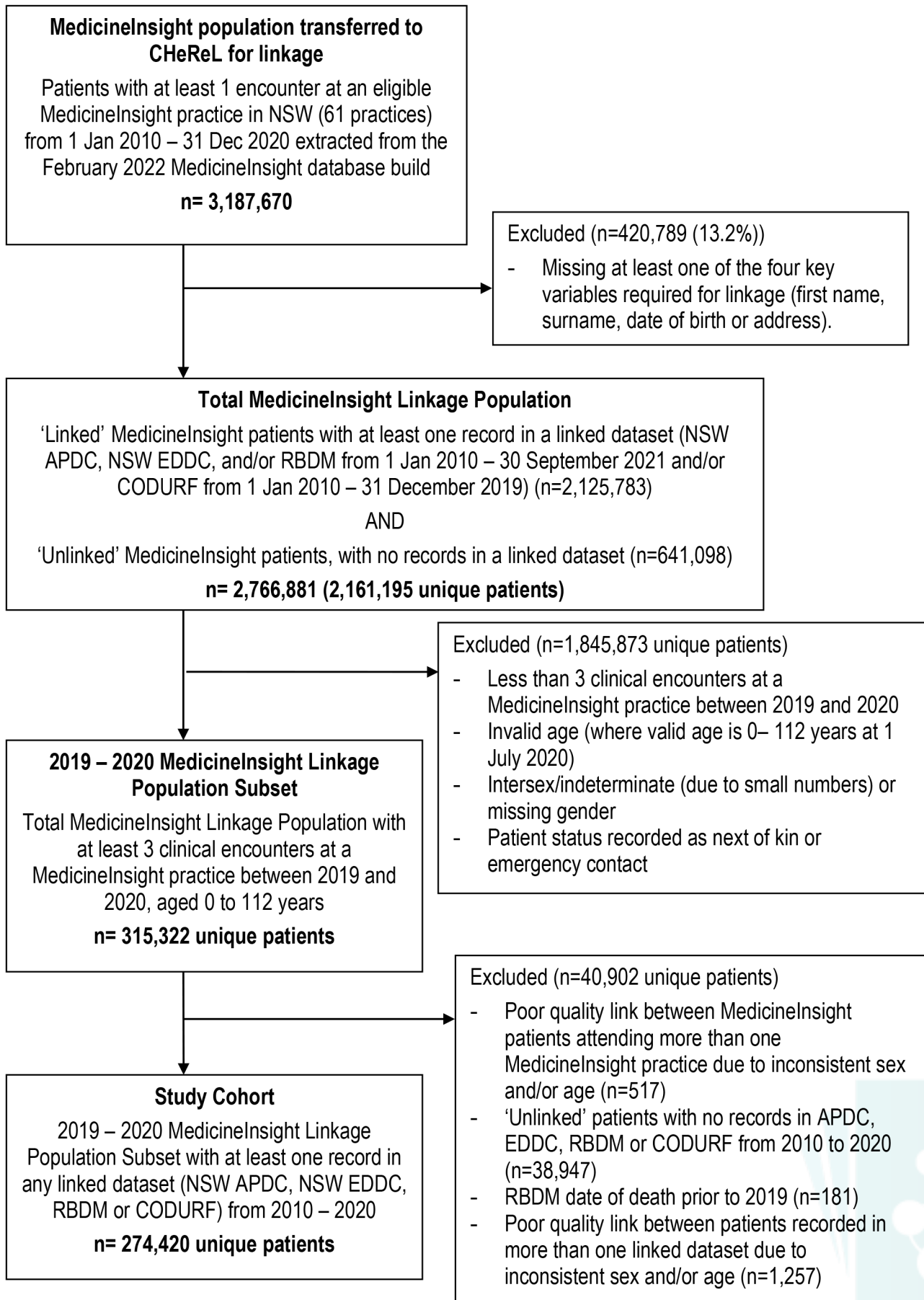


Table 1: Characteristics of the study population

	Study population N (%)	Australian national data (MBS) <sup>a</sup> 2019–20 [28] N (%)
Total persons	274,420 (100%)	22,178,760 (100%)
Female	154,622 (56.3%)	11,595,257 (52.3%)
Mean age in years (SD)	43.4 (25.6)	
Median age in years (Q1, Q3)	43.0 (21.9, 64.2)	–
Age group (in years)		
0–9	37,525 (13.7%)	2,763,081 (12.5%)
10–19	23,822 (8.7%)	2,419,160 (10.9%)
20–29	26,817 (9.8%)	2,646,230 (11.9%)
30–39	35,999 (13.1%)	3,117,218 (14.1%)
40–49	32,094 (11.7%)	2,914,753 (13.1%)
50–59	31,891 (11.6%)	2,843,363 (12.8%)
60–69	33,685 (12.3%)	2,558,260 (11.5%)
70–79	30,651 (11.2%)	1,841,556 (8.3%)
80–89	16,704 (6.1%)	864,260 (3.9%)
90+	5,232 (1.9%)	210,879 (1.0%)
Region		
Major city	184,776 (67.3%)	15,888,344 (71.6%)
Inner regional	69,311 (25.3%)	2,737,905 (12.3%)
Outer regional	20,041 (7.3%)	2,707,665 (12.2%)
Remote/ very remote	292 (0.1%)	844,227 (3.8%)
Socioeconomic status (SEIFA IRSAD quintile)		
1 (most disadvantaged)	30,992 (11.3%)	3,467,086 (15.6%)
2	58,864 (21.5%)	3,563,822 (16.1%)
3	70,848 (25.8%)	4,378,392 (19.7%)
4	54,851 (20.0%)	4,626,996 (20.9%)
5 (most advantaged)	58,865 (21.5%)	6,135,506 (27.7%)

Notes: The Medicare Benefits Schedule (MBS) data collection contains information on services that qualify for a benefit under the Health Insurance Act 1973 and for which a claim has been processed. MBS data includes patients with at least 1 GP visit during 2019–20, Study population includes regular patients as defined in methods.

Table 2: Measures of agreement for each acute serious outcome between the EHR (MedicineInsight) algorithms and linked (Reference) data

Study population N=274,420	Acute Pancreatitis	Myocardial Infarction	Stroke	Suicide	VTE*
Reference+ EHR+ (True positive)	79	773	655	177	486
Reference- EHR+ (False positive)	26	430	739	406	904
Reference+ EHR- (False negative)	471	2,405	1,704	1,189	468
Reference- EHR- (True negative)	273,844	270,812	271,322	272,648	272,562
Sensitivity (95 % CI)	14.4 (10.5, 18.2)	24.3 (22.0, 26.7)	27.8 (25.6, 30.0)	13.0 (9.8, 16.2)	50.9 (47.0, 55.0)
Specificity (95 % CI)	100.0 (100.0, 100.0)	99.8 (99.8, 99.9)	99.7 (99.7, 99.8)	99.9 (99.8, 99.9)	99.7 (99.6, 99.7)
PPV (95 %CI)	75.2 (67.3, 83.1)	64.3 (59.8, 68.7)	47.0 (43.2, 50.8)	30.4 (27.3, 33.4)	35.0 (31.4, 38.5)
NPV (95 % CI)	99.8 (99.8, 99.9)	99.1 (90.0, 99.3)	99.4 (99.3, 99.5)	99.6 (99.5, 99.6)	99.8 (99.8, 99.9)
PoA (%)	99.8	99.0	99.1	99.4	99.5

\*Venous Thromboembolism.

studies demonstrating good recording of acute serious events in primary care data [31, 32]. Presentations for these conditions occur mainly in acute care settings, with general practice involvement mainly for ongoing care or at discharge. Where patients visit multiple general practices, discharge

information may be sent to a non-MedicineInsight practice, delayed, or not sent at all. Additionally, patients may present to a non-MedicineInsight practice for ongoing care or be managed entirely in secondary care. Furthermore, the information contained within discharge summaries does not automatically

Table 3: Time between recording of acute serious events in EHR (MedicineInsight) and linked (Reference) data for 'true positive' cases during 2019–20

MedicineInsight index date vs reference index date (RID)	Acute pancreatitis		Myocardial infarction		Stroke		Suicide		VTE*	
	n	%	n	%	n	%	n	%	n	%
1–30 days before RID	9	11.4	99	12.8	95	14.5	19	10.7	58	11.9
0–30 days after RID	53	67.1	532	68.8	357	54.5	89	50.3	328	67.5
>30 days before or after RID	17	21.5	142	18.4	203	31.0	69	39.0	100	20.6
Total	79	100.0	773	100.0	655	100.0	177	100.0	486	100.0
Median difference** (Q1, Q3) in days	2 (-1, 11)		0 (-1, 7)		0 (-5, 13)		1 (-21, 16)		2 (-1, 10)	

\*Venous Thromboembolism.

\*\*difference=MedicineInsight index date minus RID.

populate the diagnosis fields within EHRs and information may be recorded in sections of the EHR that are not extracted by MedicineInsight.

Other potential explanations for the low sensitivities and PPVs in this study relate to study methodology, however we expect these only partly account for the magnitude of the discordance observed between EHR and reference hospital, ED and mortality records. Events recorded in MedicineInsight but not reference datasets were defined as false positives in this study but may be an indication of true cases that did not require hospitalisation rather than an indication of incorrect diagnoses. Scenarios leading to the misclassification of MedicineInsight cases as false positives, and consequent underestimation of PPVs, include acute events managed in general practice without hospitalisation (such as VTE) or hospitalisations and deaths for events that were not captured in the linked reference data because they occurred outside NSW. Reference hospital, ED and mortality data are also subject to errors and incompleteness. In some countries a common strategy for researchers using general practice EHRs is to confirm events and dates through questionnaires sent to general practitioners particularly in the subset of patients for whom linkage to hospital and death records is not possible [32]. Conducting a validation study using general practice questionnaires was not feasible for this study; hence linked hospital and mortality datasets were used as the reference standard.

Diagnoses recorded in EHRs outside of the study timeframe were not included, leading to potential underestimates of sensitivities and PPVs. For example, if a patient was diagnosed with a VTE by their general practice at the end of 2018 and presented to hospital with it in 2019, this study did not count the earlier diagnosis recorded in MedicineInsight as it fell outside study timeframes. Similarly, if a patient presented to hospital with an MI in 2020 but the general practice was only notified in 2021, this study did not count the later diagnosis recorded in MedicineInsight. In both examples the MedicineInsight case is misclassified as false negative, leading to underestimates of the true sensitivity and PPV. Our analysis was limited to events identified by diagnostic algorithms (Supplementary Appendix 1) and may undercount events where data is recorded in non-diagnostic fields, reducing the

number of cases ascertained and subsequently leading to lower sensitivities and PPVs. Combining case ascertainment using clinical coding with prescription or other clinical management data improved measures of agreement [30].

Of the 'true positive' cases (those identified in both EHRs and linked data during the study time period) the majority (approximately 70–80%) were recorded in the EHR within 30 days of hospital, ED or mortality record. Cases recorded outside the 30-day window could reflect true positive cases or separate events. Studies demonstrate increased documentation in primary care EHRs within broader timeframes since hospitalisation (eg looking at 60 day windows from an event) but the proportions of records with exact date matches remain relatively low [33].

## Strengths and limitations

This is the second Australian study [20], to our knowledge, to use full-scale record linkage to compare acute events between general practice EHRs and population-based data collections. Coded diagnoses from NSW hospital admissions, emergency department and cause of death data was used as the reference standard against which accuracy of the diagnostic algorithms for the acute serious conditions was benchmarked. The majority of validation studies of primary care EHR use EHR reviews as the reference standard [29, 34]. However, as EHR reviews are time consuming and labour intensive, the sample size is often small which limits the generalisability of results [10]. Linking to external data sources as the reference standard enables a much larger sample size and improves study power and representativeness. The limitation of this approach is that the recording of diagnoses in the linked datasets may be inaccurate or incomplete, accordingly inaccurately estimating event prevalence. We did not access private emergency department data and patients with milder symptoms (eg milder strokes) may not present to hospital or may present outside NSW.

MedicineInsight patients are broadly similar to Australian patients who visited a general practice during 2019–20, in terms of age, sex and socioeconomic status [28]. However, this study was limited to NSW and to practices using the INCA extraction tool. Furthermore, to improve data completeness

this study was based on a cohort of patients who regularly seek health care and have a record in one of the linked administrative datasets in the past 11 years. Compared to the general patient population the study cohort was biased towards older patients over 70 years and females. Results may not be generalisable to people who have less morbidity and utilise health care less frequently. As estimates of PPV and NPV depend on the incidence of the specific health condition [34], the PPV estimates returned in this study may be higher, and our NPV estimates may be lower, than those yielded by the diagnostic algorithms in a population with a lower incidence of the condition. A further threat to the generalisability of the results arises from only including practices from NSW in this study, however we expect these results would be similar in other jurisdictions.

Data in this study should not be used to determine true incidence of acute conditions. Firstly, the study cohort is not representative of the general patient population as discussed above. Secondly, the validity of information recorded in both MedicinesInsight and reference datasets is unclear, for example we miscategorise as false positives cases where patients present only to general practices and not hospital. We did not evaluate the veracity of the hospital or mortality data, just its agreement with MedicinesInsight algorithms. Finally, healthcare access patterns across the world were affected by COVID-19 and related measures, potentially impacting results during 2020 [35].

## Conclusions

Our findings provide valuable insights about the identification of acute serious events in Australian general practice with important implications for research. Identifying acute serious events, often managed in non-primary care settings, in Australian general practice EHRs alone will lead to undercounting of patients with these acute events and is insufficient for most observational research. Limitations related to data extraction from general practice EHRs and study design only partly explain observed low sensitivities and PPVs. Our findings suggest that general practice records complement hospital and mortality data by improving case ascertainment, and that both are required for robust estimation of acute serious events.

## Acknowledgements

We are grateful to the general practices and general practitioners who participate in MedicinesInsight and the patients whose de-identified data makes this work possible. We also acknowledge NPS MedicineWise staff, particularly Lisa Quick and Josephine Belcher who contributed to this study and CHeReL for conducting the data linkage necessary for this analysis.

## Statements on conflicts of interest

SA, AP and KC are employees of NPS MedicineWise, the custodian of MedicinesInsight data.

## Ethics statement

In December 2017, NPS MedicineWise was granted ethics approval for the standard operations and uses of the MedicinesInsight database by the Royal Australian College of General Practitioners (RACGP) National Research and Evaluation Ethics Committee (NREEC 17-017). Additional ethical approval for privacy preserving linkage of de-identified MedicinesInsight data with NSW hospital and death data held by CHeReL was obtained from the NSW Population and Health Service Research Ethics Committee (PHSREC) (Application number 2021/ETH00023). The project also received approval from the independent MedicinesInsight Data Governance Committee, on 12 July 2020 (Application number 2020–24).

## Declaration of funding

The study was funded by the Australian Government Department of Health and Aged Care. The funding body had no role in the design of the study, data collection, analysis or interpretation, nor in writing the manuscript.

## References

1. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. *Eur J Public Health*. 2016;26(1):60–4. <https://doi.org/10.1093/2016>
2. Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: an evaluation of information system impact and contingency factors. *International journal of medical informatics*. 2014;83(11):779–96. <https://doi.org/10.1016/2014>
3. Gentil ML, Cuggia M, Fiquet L, Hagenbourger C, Le Berre T, Banatre A, et al. Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. *BMC Med Inform Decis Mak*. 2017;17(1):139. <https://doi.org/10.1186/2017>
4. Youens D, Moorin R, Harrison A, Varhol R, Robinson S, Brooks C, et al. Using general practice clinical information system data for research: the case in Australia. *Int J Popul Data Sci*. 2020;5(1):1099. <https://doi.org/10.23889/2020>
5. Busingye D, Gianacas C, Pollack A, Chidwick K, Merrifield A, Norman S, et al. Data Resource Profile: MedicinesInsight, an Australian national primary health care database. *Int J Epidemiol*. 2019;48(6):1741–h. <https://doi.org/10.1093/2019>
6. Canaway R, Boyle DI, Manski-Nankervis JE, Bell J, Hocking JS, Clarke K, et al. Gathering data for decisions: best practice use of primary care electronic records for research. *Med J Aust*. 2019;210 (Suppl 6):S12–S6. <https://doi.org/10.5694/2019>



7. Dawda P, True A, Dickinson H, Janamian T, Johnson T. Value-based primary care in Australia: how far have we travelled? *Med J Aust.* 2022;216 Suppl 10(S10):S24–S7. <https://doi.org/10.5694/2022>
8. Australian Institute of Health and Welfare. Developing a National Primary Health Care Data Asset: consultation report. Canberra: AIHW; 2019. Available from: [https://www.aihw.gov.au/getmedia/023846dd-b30e-4149-a442-5dc0694aab26/aihw\\_phc\\_1.pdf.aspx](https://www.aihw.gov.au/getmedia/023846dd-b30e-4149-a442-5dc0694aab26/aihw_phc_1.pdf.aspx)
9. Canaway R, Boyle D, Manski-Nankervis J, Gray K. Primary Care Data and Linkage: Australian dataset mapping and capacity building. Melbourne Academic Centre for Health for the Australian Health Research Alliance; 2020. Available from: [https://machaustralia.org/wp-content/uploads/2021/05/Primary-care-data-and-linkage\\_Dec-2020.pdf](https://machaustralia.org/wp-content/uploads/2021/05/Primary-care-data-and-linkage_Dec-2020.pdf)
10. Havard A, Manski-Nankervis JA, Thistlethwaite J, Daniels B, Myton R, Tu K, et al. Validity of algorithms for identifying five chronic conditions in MedicinesInsight, an Australian national general practice database. *BMC Health Serv Res.* 2021;21(1):551. <https://doi.org/10.1186/2021>
11. Kitsos A, Peterson GM, Jose MD, Khanam MA, Castelino RL, Radford JC. Variation in Documenting Diagnosable Chronic Kidney Disease in General Medical Practice: Implications for Quality Improvement and Research. *Journal of primary care & community health.* 2019;10:2150132719833298. <https://doi.org/10.1177/2019>
12. Chidwick K, Kiss D, Gray R, Yoo J, Aufgang M, Zekry A. Insights into the management of chronic hepatitis C in primary care using MedicinesInsight. *Aust J Gen Pract.* 2018;47(9):639–45. <https://doi.org/10.31128/2018>
13. Gallagher AM, Dedman D, Padmanabhan S, Leufkens HGM, de Vries F. The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf.* 2019;28(5):563–9. <https://doi.org/10.1002/2019>
14. Vinogradova Y, Coupland C, Hippisley-Cox J. Use of combined oral contraceptives and risk of venous thromboembolism: nested case-control studies using the QResearch and CPRD databases. *BMJ.* 2015;350:h2135. <https://doi.org/10.1136/2015>
15. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015;44(3):827–36. <https://doi.org/10.1093/2015>
16. Okoli GN, Myles P, Murray-Thomas T, Shepherd H, Wong ICK, Edwards D. Use of Primary Care Data in Research and Pharmacovigilance: Eight Scenarios Where Prescription Data are Absent. *Drug Saf.* 2021;44(10):1033–40. <https://doi.org/10.1007/2021>
17. The University of Melbourne. GRHANITE Health Informatics Unit 2022 [11 April 2020]. Available from: <http://www.grhanite.com/>
18. Precedence Health Care. cdmNET 2020 [cited 2022 1/9/2022]. Available from: <http://precedencehealthcare.com/cdmnet/>
19. World Health Organisation. ATC: Structure and principles 2021 [cited 2022 1/9/2022]. Available from: [https://www.whocc.no/atc/structure\\_and\\_principles/](https://www.whocc.no/atc/structure_and_principles/)
20. Myton R PA, Havard A, Belcher J, Annear K, Chidwick K. MedicinesInsight report: Validation of the MedicinesInsight database: the accuracy of death recording in the MedicinesInsight general practice data compared with the National Death Index in Australia. Sydney: NPS MedicineWise; 2021. Available from: <https://www.nps.org.au/assets/NPS/pdf/MedicinesInsight-NDI-death-validation-study.pdf>
21. Centre for Health Record Linkage (CheReL). Data dictionaries: CheReL; 2022 [cited 2022 18/02/2022]. Data dictionaries for datasets in CheReL master linkage map]. Available from: <https://www.cherel.org.au/data-dictionaries#section5>
22. Irvine K, Hall R, Taylor L. A profile of the Centre for Health Record Linkage. *Int J Popul Data Sci.* 2019;4(2):1142. <https://doi.org/10.23889/2018>
23. Irvine K, Smith M, De Vos R, Brown A, Ferrante A, Boyd J, et al. Real world performance of privacy preserving record linkage. *International Journal of Population Data Science.* 2018;3(4). <https://doi.org/10.23889/2018>
24. Sax Institute. Introduction to SURE: Sax Institute Sydney, Australia; 2017 [cited 2022 1/9/2022]. Available from: <https://www.saxinstitute.org.au/wp-content/uploads/Intro-Guide-to-SURE-V3.0-Mar17.pdf>
25. Royal Australian College of General Practitioners (RACGP). RACGP Standards for General Practices. 2014
26. Health Communication Network Limited. Practice Incentives Program (PIP) eHealth Incentive: Requirement 3—Data Records and Clinical Coding. 2014
27. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pitfalls, and Pitfalls in Research and Practice. *Frontiers in Public Health.* 2017;5. Available from: <https://www.frontiersin.org/articles/10.3389/fpubh.2017.00307>
28. NPS MedicineWise. General Practice Insights Report July 2019–June 2020 including analyses related to the impact of COVID-19. Sydney: NPS MedicineWise; 2021. Available from: <https://www.nps.org.au/assets/NPS/pdf/GPIR-Report-2019-20.pdf>
29. McBrien KA, Souri S, Symonds NE, Rouhi A, Lethebe BC, Williamson TS, et al. Identification of validated case definitions for medical conditions used

in primary care electronic medical record databases: a systematic review. *Journal of the American Medical Informatics Association : JAMIA*. 2018;25(11):1567–78. <https://doi.org/10.1093/2018>

30. Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic diseases and injuries in Canada*. 2013;33(3). Available from: <https://pubmed.ncbi.nlm.nih.gov/23735455/>
31. Persson R, Sponholtz T, Vasilakis-Scaramozza C, Hagberg KW, Williams T, Kotecha D, et al. Quality and Completeness of Myocardial Infarction Recording in Clinical Practice Research Datalink Aurum. *Clinical epidemiology*. 2021;13:745–53. <https://doi.org/10.2147/2021>
32. Arana A, Margulis AV, Varas-Lorenzo C, Bui CL, Gilson A, McQuay LJ, et al. Validation of cardiovascular outcomes and risk factors in the Clinical Practice Research Datalink in the United Kingdom. *Pharmacoepidemiol Drug Saf*. 2021;30(2):237–47. <https://doi.org/10.1002/2021>
33. Morgan A, Sinnott SJ, Smeeth L, Minassian C, Quint J. Concordance in the recording of stroke across UK primary and secondary care datasets: a population-based cohort study. *BJGP Open*. 2021;5(2):BJGPO.2020.0117. <https://doi.org/10.3399/2020>
34. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *Journal of clinical epidemiology* 2011;64(8):821–9. <https://doi.org/10.1016/2010>
35. Sutherland K, Chessman J, Zhao J, Sara G, Shetty A, Smith S, et al. Impact of COVID-19 on healthcare activity in NSW, Australia. *Public Health Res Pract*. 2020;30(4). <https://doi.org/10.17061/2020>
36. Australian Institute of Health and Welfare. Australian Burden of Disease Study: Methods and supplementary material 2018. Canberra: AIHW; 2021. Available from: <https://www.aihw.gov.au/reports/burden-of-disease/abds-methods-supplementary-material-2018>

## Abbreviations

- 95% CI: 95% confidence interval – a range of values that are likely to encompass the true value
- ABS: Australian Bureau of Statistics
- APDC: Admitted Patient Data Collection
- ASGS-RA: Australian Statistical Geography Standard Remoteness Areas
- BP: Best Practice- a general practice clinical information system

- CHeReL: Centre for Health Record Linkage (NS)
- CIS: Clinical information system
- COVID-19: Coronavirus SARS-CoV2
- CODURF: Cause of Death Unit Record File. Also referred to as ACR CODURF (Australian Coordinating Registry Cause of Death Unit Record File)
- CPRD: Clinical Practice Research Datalink
- DUSC: Drug Utilisation Subcommittee (Commonwealth Department of Health and Aged Care)
- EDDC: Emergency Department Data Collection
- EHR: Electronic Health Record-the patient’s health record within the general practice clinical information system
- GP: General practice
- GRHANITE: GeneRic Health Network Information Technology for the Enterprise- a data extraction tool used with GENERAL PRACTICE clinical information systems
- ICD-10 & ICD-10 AM: International Classification of Diseases version 10 (and the Australian Modification)
- INCA: Integrated Care- a data extraction tool used with GENERAL PRACTICE clinical information systems
- IRSAD: ABS Index of Relative Socioeconomic Advantage and Disadvantage
- MBS: Medicare Benefits Schedule
- MD: Medical Director- a general practice clinical information system
- NHMRC: National Health and Medical Research Council
- NREEC: National Research and Evaluation Ethics Committee
- NPV: Negative predictive value
- NSW: New South Wales
- p-value: A test of statistical significance
- PHSREC: NSW Population and Health Services Research Ethics Committee
- PBAC: Pharmaceutical Benefits Advisory Committee
- PBS: Pharmaceutical Benefits Scheme
- PoA: Percentage of agreement
- PPN: Project specific Person Number
- PPRL: Privacy preserving record linkage – a method using a hashed key generated at source from first name, last name, date of birth and address (ie a Bloom filter) to link records across datasets without needing to transfer or expose identifying information to researchers
- PPV: Positive predictive value
- RACGP: Royal Australian College of General Practitioners
- RID: Reference Index Date
- RBDM: Register of Births Deaths and Marriages
- SAS: A statistical software package
- SNOMED-CT: Systematized Nomenclature of Medicine Clinical Terms
- SURE: Secured Unified Research Environment
- SEIFA: ABS Socio-Economic Indexes for Areas
- TGA: Therapeutic Goods Administration
- VTE: Venous thromboembolism

## Appendix 1: Description of the MedicinesInsight algorithms for acute serious events

### Acute Pancreatitis

This flag indicates those records where Acute Pancreatitis (or a relevant synonym) is reported in MedicinesInsight either as a coded condition, using a drop down list in the clinical information system (CIS)), or as a non-coded condition (free text) in one or more of the 'Diagnosis', 'Reason for visit' or 'Reason for prescription' fields.

The following coded terms (from Docle or Pyefinch) have been used to identify records for inclusion:

- ACUTE PANCREATITIS
- PANCREATITIS, ACUTE
- PANCREATITIS - ACUTE

The following free text strings have been used to search for terms which may indicate records for inclusion:

- INFLAMMATION
- PANCREAS
- PANCREATITIS
- RECURRENT PANCREATITIS

Records identified by a free text string alone are not automatically flagged but are individually reviewed by a clinical coder to determine whether the text string actually refers to the condition indicated or is present in another context (eg, a search for 'cancer' may identify 'partner died from cancer'). Each record is flagged accordingly.

This condition flag was created in October 2020.

### Myocardial Infarction (MI)

This flag indicates those records where Myocardial infarction (or a relevant synonym) is reported in MedicinesInsight either as a coded condition (using a drop down list in the CIS) or as a non-coded condition (free text) in one or more of the 'Diagnosis', 'Reason for visit' or 'Reason for prescription' fields.

The following coded terms (from Docle or Pyefinch) have been used to identify records for inclusion:

- MYOCARDIAL INFARCTION - ANTEROLATERAL
- INFERIOR MYOCARDIAL INFARCTION
- MYOCARDIAL INFARCTION - INFERIOR
- MYOCARDIAL INFARCTION - POSTERIOR
- POSTERIOR MYOCARDIAL INFARCT
- MYOCARDIAL INFARCT - SILENT
- SILENT MYOCARDIAL INFARCTION
- MYOCARDIAL INFARCTION - SUBENDOCARDIAL
- SUBENDOCARDIAL MYOCARDIAL INFARCT
- MYOCARDIAL INFARCTION - SUPERIOR
- SUPERIOR MYOCARDIAL INFARCT
- ACUTE MYOCARDIAL INFARCTION
- AMI
- HEART ATTACK
- MI
- MYOCARDIAL INFARCTION
- ANTERIOR MYOCARDIAL INFARCT
- MYOCARDIAL INFARCTION, ANTERIOR
- ANTEROLATERAL MYOCARDIAL INFARCT
- MYOCARDIAL INFARCTION, ANTEROLATERAL
- INFERIOR MYOCARDIAL INFARCTION
- MYOCARDIAL INFARCTION, INFERIOR
- MYOCARDIAL INFARCTION, POSTERIOR
- POSTERIOR MYOCARDIAL INFARCTION
- MYOCARDIAL INFARCTION, SUBENDOCARDIAL

- SUBENDOCARDIAL INFARCT
- SUBENDOCARDIAL MYOCARDIAL INFARCT
- MYOCARDIAL INFARCTION, SUPERIOR
- SUPERIOR MYOCARDIAL INFARCTION
- DRESSLER'S SYNDROME
- POSTMYOCARDIAL INFARCTION SYNDROME
- POSTPERICARDIOTOMY SYNDROME
- MYOCARDIAL INFARCTION, NON STEMI
- NON ST ELEVATION MYOCARDIAL INFARCTION
- NSTEMI
- ST ELEVATION MYOCARDIAL INFARCTION
- MYOCARDIAL INFARCTION, STEMI
- ST ELEVATION MYOCARDIAL INFARCTION
- STEMI

The following free text strings have been used to search for terms which may indicate records for inclusion:

- DRESSLER
- HEART ATTACK
- INFARCT
- M.I
- M.I.
- MI
- MYOCARDIAL DAMAGE
- POSTPERICARDIOTOMY SYNDROME
- STEMI

Records identified by a free text string alone are not automatically flagged but are individually reviewed by a clinical coder to determine whether the text string actually refers to the condition indicated or is present in another context (eg, a search for 'cancer' may identify 'partner died from cancer'). Each record is flagged accordingly.

This condition flag was updated in October 2020.

### Stroke

This flag indicates those records where Stroke, including transient ischaemic attacks, (or a relevant synonym) is reported in MedicinesInsight either as a coded condition (selected from a list in the CIS) or as a free text entry in one or more of the 'Diagnosis', 'Reason for visit' or 'Reason for prescription' fields. The following coded terms (from Docle or Pyefinch) have been used to identify records for inclusion:

- ARTERIAL EMBOLISM - MINOR
- CEREBRAL HAEMORRHAGE
- CEREBRAL INFARCTION
- CEREBRAL TIA
- CEREBRAL TRANSIENT ISCHAEMIA
- CEREBRAL TRANSIENT ISCHAEMIC ATTACKS
- CEREBROVASCULAR ACCIDENT
- CVA
- CVA (CEREBROVASCULAR ACCIDENT)
- HAEMORRHAGE - INTRACEREBRAL
- HAEMORRHAGE, INTRACEREBRAL
- HAEMORRHAGIC CVA
- HAEMORRHAGIC STROKE
- INTRACEREBRAL BLEED
- INTRACEREBRAL HAEMORRHAGE
- INTRACRANIAL HAEMORRHAGE
- ISCHAEMIC STROKE
- LACUNAR INFARCT
- LACUNAR STROKE
- MIGRAINOUS STROKE
- MIGRANOUS STROKE



- STROKE
- STROKE - HAEMORRHAGIC
- STROKE - ISCHAEMIC
- STROKE - LACUNAR
- STROKE - MIGRANOUS
- STROKE - THROMBOTIC
- STROKE, HAEMORRHAGIC
- STROKE, ISCHAEMIC
- STROKE, LACUNAR
- STROKE, MIGRAINOUS
- STROKE, THROMBOTIC
- SYNCOPE, TIA
- THROMBOTIC - STROKE
- THROMBOTIC STROKE
- TIA
- TIA (TRANSIENT ISCHAEMIC ATTACK)
- TRANSIENT ISCHAEMIC ATTACK
- VISUAL CORTEX STROKE

The following free text strings have been used to search for terms which may indicate records for inclusion:

- ARTERIAL EMBOLISM - MINOR
- C.V.A
- CEREBRAL HAEMORRHAGE
- CEREBRAL TRANSIENT ISCHAEMIA
- CEREBROVASCULAR ACCIDENT
- CVA
- HAEMORRHAGE – INTRACEREBRAL
- HAEMORRHAGE, INTRACEREBRAL
- INTRACEREBRAL BLEED
- INTRACRANIAL HAEMORRHAGE
- LACUNAR INFARCT
- STROKE
- T.I.A
- TIA
- TRANSIENT ISCHAEMIC ATTACK

Records identified by a free text string alone are not automatically flagged but are individually reviewed by a clinical coder to determine whether the text string actually refers to the condition indicated or is present in another context (eg, a search for 'cancer' may identify 'partner died from cancer'). Each record is flagged accordingly. This condition flag was last updated in January 2019.

#### Suicide

This flag indicates those records where Suicide related conditions (or a relevant synonym) is reported in MedicineInsight either as a coded condition (using a drop down list in the CIS) or as a non-coded condition (free text) in one or more of the 'Diagnosis', 'Reason for visit' or 'Reason for prescription' fields. The following coded terms (from Docle or Pyefinch) have been used to identify records for inclusion:

- SUICIDAL IDEATION
- SUICIDAL THOUGHTS
- THINKING OF SUICIDE
- SUICIDE
- ATTEMPTED SUICIDE
- SUICIDE ATTEMPT
- SELF-HARM
- SUICIDE IDEAS
- SUICIDAL IDEATION
- SUICIDAL TENDENCIES
- SELF-MUTILATION

The following free text strings have been used to search for terms which may indicate records for inclusion:

- DEAD
- DEATH
- DECEASED
- HANGING
- LIFE EXTINCT
- POISON
- R.I.P
- RIP
- SELF HARM
- SELF INFLICT
- SELF INJURY
- SELF-MUTILATION
- SUIC

Records identified by a free text string alone are not automatically flagged but are individually reviewed by a clinical coder to determine whether the text string actually refers to the condition indicated or is present in another context (eg, a search for 'cancer' may identify 'partner died from cancer'). Each record is flagged accordingly. This condition flag was created in October 2020.

#### Venous Thromboembolism (VTE)

This flag indicates those records where VTE, including Deep Vein Thrombosis and Pulmonary Embolism (or a relevant synonym) is reported in MedicineInsight either as a coded condition (selected from a list in the CIS) or as a free text entry in one or more of the 'Diagnosis', 'Reason for visit' or 'Reason for prescription' fields.

The following coded terms (from Docle or Pyefinch) have been used to identify records for inclusion:

- DEEP VENOUS THROMBOSIS
- DVT
- EMBOLISM – PULMONARY
- EMBOLISM, PULMONARY
- PULMONARY EMBOLISM
- PULMONARY EMBOLISM – SADDLE TYPE
- SADDLE PULMONARY EMBOLISM
- THROMBOSIS – DEEP VEIN

The following free text strings have been used to search for terms which may indicate records for inclusion:

- D.V.T
- DEEP VENOUS THROMBOSIS
- DVT
- EMBOLISM
- EMBOLUS
- PE
- PULM EMB
- PULMONARY EMB
- THROMBOSIS – DEEP VEIN
- VTE

Note: The abbreviation 'PE' can refer to many terms such as 'Physical Examination', 'Pre-eclampsia', 'Premature ejaculation' etc. 'PE' was only assumed to refer to 'Pulmonary embolism' when reported with a related condition or other qualifier, or in conjunction with an anticoagulant eg

Flagged as 'Pulmonary embolism' – WARFARIN FOR PE', 'CLEXANE FOR PE', 'DVT AND PE', 'SADDLE PE', 'LEFT LUNG PE', 'UNPROVOKED PE', 'MALIGNANT PE'



Not flagged as 'Pulmonary embolism' – 'PE', 'POST OP PE', 'PE UNREMARKABLE', 'HYERTENSION POSSIBLE PE', 'PE ED LUTS'

Records identified by a free text string alone are not automatically flagged but are individually reviewed by a clinical

coder to determine whether the text string actually refers to the condition indicated or is present in another context (eg, a search for 'cancer' may identify 'partner died from cancer'). Each record is flagged accordingly. This condition flag was last updated in October 2020.

Appendix 2: Description of the codes used to identify acute serious events in the linked datasets

Condition	ICD-10-AM / ICD-10	SNOMED-CT
Acute pancreatitis	K85. Patients with acute pancreatitis are incident cases of short duration. This diagnosis code includes acute episodes within a diagnosis of chronic pancreatitis. [36]	4399003, 7881005, 197456007, 96081000119101, 235945004, 197458008, 197460005, 235942001, 235944000, 235949005, 235941008, 235943006, 277537008, 235948002, 235950005, 235946003
Myocardial Infarction	I20-I23	703164000, 703165004, 52035003, 62695002, 703213009, 703253007, 57054005, 17531000119105, 896689003, 23311000119105, 703212004, 54329005, 703252002, 70211005, 836294006, 59063002, 194809007, 282006, 64627002, 73795002, 703251009, 65547006, 836295007, 76593002, 58612006, 70998009, 15990001, 836293000, 79009004, 307140009, 233828006, 233826005, 233830008, 233832000, 233834004, 233837006, 401314000, 233838001, 233827001, 233825009, 233829003, 233831007, 233833005, 233836002, 304914007, 401303003, 840316004, 15713081000119108, 840312002, 840609007, 896691006, 846668006, 868217004, 846683001, 868224003, 868220007, 868225002, 868226001, 868214006, 840680009, 840309000, 15713121000119105, 285981000119103, 15712881000119105, 15712961000119108, 896696001, 12238111000119106, 12238151000119107, 15712921000119103, 15713041000119103, 15713201000119105, 15712841000119100, 896697005, 15713161000119100, 70422006, 233835003, 738011000168103, 394710008, 738061000168100, 42531007, 428196007, 22298006, 726499301000119105, 16837681000119104, 418044006, 879955009, 314207007, 129574000, 311796008, 311792005, 311793000, 233843008, 738051000168102, 738071000168106, 194802003
Stroke (including TIA)	I60–I68, G45.8,G45.9,G46 [36]	230714009, 413758000, 16371781000119100, 99451000119105, 788881005, 788880006, 724426006, 788882003, 788883008, 724425005, 724993002, 724424009, 724994008, 788884002, 915141931000119109, 107557061000119108, 152148641000119104, 58173271000119101, 939885431000119109, 168747591000119109, 732330391000119107, 251770561000119107, 859422751000119101, 517253051000119105, 849488701000119104, 384430101000119103, 806161651000119106, 239965291000119107, 720261501000119107, 346674811000119104, 188174841000119103, 86553761000119103, 898941951000119108, 182960891000119101,

Continued.

Appendix 2: Continued

		117776611000119102, 655081461000119101, 849579281000119106, 16661971000119104, 957319791000119104, 851365731000119106, 759950981000119101, 496369931000119104, 38595071000119104, 511452481000119102, 595899961000119100, 652287331000119104, 881694631000119107, 769023031000119104, 25133001, 16891111000119104, 371041009, 422504002, 140911000119109, 140921000119102, 725132001, 195216008, 738221000168102, 1142056004, 371121002, 111297002, 9901000119100, 373606000, 116288000, 722929005, 384993003, 230715005, 78569004, 57981008, 195217004, 723082006, 230690007, 724429004, 329641000119104, 230691006, 16026991000119107, 16644681000119102, 16024111000119109, 16024151000119105, 329461000119102, 16000511000119103, 329371000119101, 16023991000119104, 16000351000119109, 329431000119105, 292691000119103, 16024271000119107, 329451000119104, 16000431000119109, 329361000119107, 16024031000119100, 16000391000119104, 329421000119107, 16023911000119108, 16026951000119102, 292681000119101, 16661931000119102, 292671000119104, 292661000119105, 330791000119108, 16002111000119106, 329651000119102, 16002031000119102, 716051003, 275434003, 230713003, 371040005
Suicide (related conditions)	X60-X84, Y87.0 [36]	274228002, 418420002, 48981002, 44301001, 287189003, 287194003, 287188006, 287192004, 287193009, 287190007, 287195002, 287191006, 78070009, 891003, 23546003, 401231009, 401229000, 267073005, 425104003, 304594002, 6471006, 247650009, 711561004, 269725004, 8521000175109, 36153001, 440144004, 55554002, 53846008
Venous Thromboembolism (VTE) including deep vein thrombosis and pulmonary embolism	I26, I82.8 & I82.9	128053003, 429098002, 309735004, 15964701000119109, 233936003, 706870000, 707414004, 59282003, 723859005, 713078005, 1001000119102, 74315008, 233935004, 328511000119109, 233937007

