Introduction

Information on clinical diagnoses and outcomes derived from electronic health records (EHRs) is of increasing relevance for both clinicians and researchers [1]. These records represent a rich source of clinical information, collected at minimal cost, in large numbers of people and with potential for linkage to other data sources [2]. Acute myocardial infarction (AMI) represents an important clinical outcome, and unlike many diseases, has internationally accepted and well-defined diagnostic criteria which are widely used in clinical practice [3]. EHRs have been used in different types of research: to assess quality and performance of healthcare providers for managing patients [4,5], to monitor national trends in mortality and morbidity, along with intra and inter-country comparisons of healthcare policy [6] and to generate outcome data for prospective studies and clinical trials [7–10]. There is a growing interest in EHR phenotypes to gain insights into the aetiology of AMI through “omic” association studies [11].

However there have been no systematic, contemporary evaluations of the diverse sources of EHR data relevant for AMI ascertainment, and of the validity of EHR data on AMI. In this context, EHRs encompass any electronic source of information relevant to the definition of AMI, including hospital EHRs containing clinical details such as markers of myocardial necrosis values, electrocardiogram (ECG) data, and administrative data on diagnoses used for billing purposes; registries (including disease and mortality registries); and primary care EHRs.

We sought to (1) evaluate the extent to which electronically stored information on markers of myocardial necrosis, ECG findings, symptoms and diagnoses has been used to ascertain AMI, (2) evaluate the accuracy of such EHR information, in different clinical settings, countries and for different phenotypes and (3) make recommendations where improvements are required. In order to do so, we carried out a systematic review of contemporary studies according to MOOSE [12] and PRISMA [13] guidelines.
2. Methods

2.1. Search strategies

We searched MEDLINE and EMBASE databases for studies reporting on EHR-derived AMI diagnosis published between 1 January 2000 and 31 July 2014. Keywords for EHRs, AMI, sensitivity, specificity, positive predictive value (PPV), markers of myocardial necrosis and ECG were searched using Medical Subject Headings (MeSH) terms and combined using Boolean operators as appropriate (Supplementary material online, Appendix A).

2.2. Inclusion criteria

Eligible studies (1) were published after the year 2000; (2) ascertained information relevant to an AMI diagnosis available in EHRs; (3) compared EHR data with manual chart review, or other relevant information; and, (4) provided a sensitivity (true positive fraction of all AMI diagnosed in EHRs). Where available, we report the sensitivity and specificity.

The medical classification systems used to identify AMI diagnosis were International Classification of Diseases revision 8 (ICD-8), ICD-9 (–CM, Clinical Modification) [14] or ICD-10, Diagnosis-related Group (DRG) [15], Current Procedural Terminology (CPT) [16], and in primary care Read Codes [17] and International Classification of Primary Care (ICPC) [18]. Studies using unstructured data (free text) were also included and studies published in a foreign language with an abstract in English were translated by a native speaker.

2.3. Study screening and data extraction

Two authors (BR and NKF) independently reviewed all abstracts for eligibility and obtained full text studies where inclusion criteria were met or there was uncertainty. Studies were excluded when both reviewers agreed the inclusion criteria were not met and conflicts were resolved by discussion with a third author (RSP) to reach consensus. Additional studies were identified by hand-searching reference lists. BR and NKF extracted quantitative and qualitative data from eligible studies. Multiple publications from one study dataset were deemed eligible where results were reported for two or more AMI phenotypes.

2.4. Quantitative and qualitative measurements

Accuracy of AMI diagnosis in an EHR source compared to a reference was assessed by PPV, which we defined arbitrarily as high if equal to or above 90% and moderate if between 70 and 89%. The Wilson method for binomial proportions [19] was used in STATA 13.1 to calculate 95% confidence intervals (CI) for studies that did not report it. For calculated values, decimal places were only reported when study sample sizes (n) were equal to or above 200. Eleven quality criteria adapted from a standardised checklist [13] were used to evaluate the quality of studies included in this review (Supplementary material online, Appendix B).

3. Results

The initial search strategy identified 2561 abstracts (Supplementary material online, Appendix C). After excluding duplicates, 1862 abstracts were reviewed for eligibility, with thirty three studies meeting the full inclusion criteria, three of which were published in a foreign language (Supplementary material online, Appendix D). A total of 128658 EHR-derived AMI diagnosis were identified and cross-referenced, of which 18164 potential cases were validated using manual chart review.

3.1. Clinical data features used for ascertainment and validation of AMI

Studies were grouped into three different groups according to the EHR source from which AMI diagnosis was derived, with twenty three studies cross-referencing AMI diagnosis from secondary care, four from mortality registries and three from primary care. Despite being collected and used for different purposes, studies using administrative billing databases (13/23) and hospital databases (10/23) to identify AMI diagnosis in secondary care EHRs were grouped together because AMI diagnosis was mostly derived using the same clinical data feature (ICD-coded diagnosis).

Only one study cross-referenced EHR-derived abnormal troponin levels with ICD-9 coded diagnosis [20]. None of the studies used electronically stored ECG data (digital wave form, computer interpretation or physician interpretation), while one used symptoms in the form of free text, combined with coded diagnosis, to search EHR sources for AMI cases (n = 213) [21]. Heriot et al. also used unstructured data to search for AMI diagnosis in electronic databases (n = 48) and compared these with post-mortem diagnosis obtained from autopsy reports [22]. Remaining studies relied on a mixture of structured data from coded admission and/or discharge diagnosis and death certificates (Fig. 1).

Chart review was the preferred reference against which EHR-derived AMI diagnosis was compared (24/33), despite five studies reporting concerns over the lack of completeness of medical charts. Other references included disease and mortality registries, questionnaires sent to general practitioners, autopsy findings and computerised algorithms based on information obtained from medical charts (Supplementary material online, Appendix E). Twenty nine studies obtained cross referencing information on markers of myocardial necrosis, twenty five on ECG findings and twenty one on clinical symptoms. ‘True’ AMI status was mostly based on the WHO MONICA [23] criteria (10/33), or the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC) Committee [24] and American Heart Association (AHA) Council on Epidemiology and Prevention [25] criteria (9/33). Other criteria are listed on the Supplementary material online, Appendix E.

3.2. Algorithms used to define EHR-derived AMI diagnosis

Studies used a range of coding algorithms to ascertain AMI diagnosis in EHRs. Eighteen studies confined the search to ICD-10 code I21 and/or ICD-9 code 410, four combined those with codes for subsequent acute myocardial infarction (ICD-10 code I22 or ICD-9 code 412), while five also used other forms of acute ischaemic heart disease (ICD-9 codes 411, 413, 414 or ICD-10 codes I20 and I24) in their search algorithm. Studies that ascertained STEMI and NSTEMI from hospital EHR used algorithms based on a combination of ICD codes to account for the lack of a specific ICD-9 and ICD-10 code for these events (Supplementary material online, Appendix E).

3.3. PPV of AMI diagnosis in EHR sources (Fig. 2)

Twenty three studies ascertained AMI diagnosis from a secondary EHR source against a non-electronic reference. Of those, twenty used chart review as reference and nineteen reported moderate to high PPVs (range 76–100%). Despite observing low PPV (20.7%), Gonski et al. [20] found high sensitivity (100%), specificity (78.4%) and negative predictive value (NPV) (100%) when comparing troponin levels in electronically stored troponin lists to ICD-coded AMI discharge diagnosis derived from chart review. The three studies that did not use chart review as reference observed the lowest PPVs (≤75%). Two used a computerised algorithm based on information extracted by chart review, of which one reported a PPV of 40% [26].

PPVs for the three studies that compared AMI diagnosis in primary care EHRs with a reference varied between 75.0% and 96.6%, while four studies using mortality registries found PPV between 67.1% and
74%. Six studies compared EHR-derived AMI diagnosis with another electronic source such as disease registry and reported high PPVs (range 82–97%). Only two studies, both from Denmark, ascertained AMI cases from disease registries and reported high PPVs (93.6% and 100%).

3.4. Phenotypes of EHR-derived AMI diagnosis (Fig. 3)

We found three studies that used EHR-derived STEMI events (PPV range 71–100%), of which two ascertained both STEMI and NSTEMI finding slightly higher PPVs for NSTEMI events (91% and 92.1% as opposed to 79.2% and 88.1% for STEMI) [27,28]. The third study cross-referenced STEMI diagnosis in three different countries, reporting lower PPV for the Italian primary care EHRs (PPV 71.4%) compared to both the Dutch primary care (PPV 92.6%) and Danish secondary care EHRs (PPV 100%) [21].

Three studies validated EHR-derived non-fatal AMI diagnosis and reported high PPVs (range 82–92.2%), while diagnosis of fatal AMI was generally found to be less accurate (PPV range 64–91.7%). A possible explanation for this finding is that fatal AMI included unheralded AMI and events occurring outside hospital settings, where potentially less information was available to validate the AMI diagnosis and, therefore, these cases could not be confirmed when applying the criteria adopted by the study. Additionally, sudden unexplained deaths might have been inaccurately classified as fatal AMI, particularly when occurring outside the hospital. In order to explore these possibilities, we further stratified fatal AMI into in-hospital (death after admission to the hospital) and out-of-hospital (OOH) AMI deaths. OOH deaths were defined as those that occurred outside hospital care such in emergency care (before admission to hospital), in residential care homes and dead-on-arrival cases. We found that in-hospital AMI deaths were more accurately diagnosed in EHRs (PPV range 74–91.7%) than OOH AMI deaths (PPV range 64–88.7%). However, additional studies are needed to investigate this hypothesis as we found very few studies validating these phenotypes.

Incident AMI diagnosis was validated by seven studies (PPV range 72.7–97%), though case definitions adopted by the studies varied from no previous AMI ever to no record of AMI in the previous 8 years.

3.5. International comparisons

We found thirteen studies ascertaining EHR-derived AMI diagnosis from countries across Europe, ten from the United States (USA), three from Canada, two from Australia, two from Brazil, and two from South Korea. Accuracy of AMI diagnosis was similar between the majority of studies published in the USA (PPV range 73.0–96.7%) and Europe (PPV range 72.7–100.0%). When comparing EHRs in three different countries, Coloma et al. found higher PPV for AMI diagnosis in the Italian primary care database (PPV 96.6%) and the Danish secondary care registry (100%) than in the Dutch primary care database (PPV 75.0%), despite significant differences in coding algorithms [21].

4. Discussion

Among thirty three studies ascertaining over 120000 potential AMI events we found that diagnoses (mostly from administrative EHRs) had moderate to high PPVs when cross referenced against chart review or another EHR source. These results, however, are limited by significant heterogeneity and lack of standardisation between studies, both in the reference adopted and in the clinical data features used to define AMI in the EHR source and in the reference. EHR curated troponin values,
ECG findings and symptom descriptions – the core clinical attributes in the internationally agreed and implemented definition [3] – were rarely used to identify AMI patients electronically. The use of administrative billing codes only is not optimal as they do not represent the true patient state but a reflection of it through the data recording process of the healthcare workflow. If readily available, detailed phenotypic biomarkers can be used to re-construct AMI diagnosis according to different definitions, decreasing heterogeneity and improving data sharing between studies. These findings emphasize the need to improve the electronic capture of high quality data for use in clinical care and reuse in research. Professional societies such as the ESC and AHA should make clinical practice recommendations for the extraction and harmonisation of hospital data to enhance the accuracy and depth of AMI phenotyping as a pre-requisite to high quality care (Box 1).

Despite investment in hospital EHRs such as the HITECH Act [29], rich phenotypic and clinical data held in hospital information systems still remain largely inaccessible for research, primarily due to use of proprietary platforms with restricted capabilities and/or access (Box 1). Troponin quantification has revolutionised diagnosis of AMI [30], especially NSTEMI, and lends itself to electronic and structured storage within any EHR system, yet we found only one study [20] that used this information to identify AMI cases. Furthermore, application of standards for reporting and storage of data [31] can rapidly facilitate sharing of such data for research but are not widely used. In contrast, use of ECG findings relates to a fundamental problem of data acquisition and storage, with the majority of healthcare institutions still working with paper copies rather than electronic, and unsurprisingly we found no studies identifying AMI based on electronically stored ECG data in combination with other clinical data features. National or regional initiatives in conjunction with industry are needed to migrate to an electronic system, as exemplified by healthcare organisations such as Kaiser Permanente introducing electronic ECG capture across sites in the USA [32]. Access to ECG data in EHRs may also allow examination of other relevant phenotypes [33,34], as exemplified by the CHARGE consortium [35], as well as the discovery of new ‘deeper’ electrocardiographic phenotypes (Box 1).

A major problem in evaluating studies of EHR-derived AMI diagnoses is the implementation of a ‘gold standard’ for cross referencing purposes. We identified few studies that could faithfully report against a ‘gold standard’ i.e. against sources with complete information available to classify cases according to external, internationally agreed criteria. The majority of the studies used manual chart review despite it being well recognised that paper charts may not record troponin levels, ECG or symptoms and four of the included studies did not report the use of any of these features to confirm AMI diagnosis. Furthermore, none of the studies utilised procedure codes for primary PCI or thrombolysis as surrogate definitions of AMI to either ascertain or validate an AMI diagnosis, despite being highly specific treatments for STEMI.
Importantly, agreement within and between studies depends on which external criteria for AMI are used. For example, Barchielli et al. [37] applied two sets of criteria to the same cohort and found that PPV varied considerably (PPV 94.6% using AHA Council of Epidemiology and Prevention [25] criteria; PPV 65.3% using WHO MONICA [23] criteria). These findings further emphasize the importance of raw data on markers of myocardial necrosis, ECG and symptoms in order to reconstruct AMI according to different definitions, facilitating comparison of study results and data sharing. Additionally, such information is essential when combining multiple sources that used different classification criteria and to reclassify cases after revisions of the current definition of AMI.

We found few studies investigating the validity of diagnosis among different AMI phenotypes, with most focusing on composite AMI despite emerging data suggesting important differences in survival and outcomes among those with different AMI phenotypes [38]. Association studies using “omic” technology will benefit from greater phenotyping of AMI in order to derive mechanistic insights [11,39,40]. There are questions regarding the accuracy of diagnoses derived solely from death certificates [41], with a recent systematic review reporting low accuracy for AMI diagnosis derived from vital statistics databases [42]. While we found better agreement in studies published after 2000 using mortality registries, PPV was still considerably lower than in those using other sources of EHRs. This conclusion, however, is limited by the high level of heterogeneity between studies included in both groups and within each group.

Comparisons across different EHR sources and countries were limited by the heterogeneity in the algorithms used to define EHR-derived AMI diagnosis. These findings mirror those of a previous systematic review [43] and highlight the need for internationally agreed phenotyping EHR algorithms, spanning clinical coding systems, to be developed and used for defining AMI cases in EHRs [44,45].

AMI diagnoses derived from secondary care, used in the majority of the studies we report on, are collected primarily for administrative purposes, rather than as a clinical tool. As such they may not reflect the true patient state but rather the healthcare setting or specific coding processes they occur in and may be influenced by financial or other pressures [46]. These processes are inherently subject to variability, poorly documented and usually not entirely understood by researchers. Furthermore the quality or accuracy of coding itself is also variable. In some healthcare systems, coding is performed by clinicians while in others non-clinical personnel follow a predefined protocol. Thus, to improve quality of diagnosis coding as well as comparability of studies, coding processes should be integrated into clinical care according to guidelines issued by relevant professional bodies. Contextual information essential to understand the clinical data, such as laboratory reference values, units of measurements and assay methods, also should be consistently documented and made available (Box 1). Similarly, additional information that might alter interpretation of the data, such as renal function, evidence of infection, co-morbidities and medications, should be systematically recorded and made accessible.

Importantly, data from a patient with AMI may be recorded in a growing number of different EHR sources (primary care, secondary care and registries) and this represents an opportunity to increase the ascertainment and accuracy of AMI. Reassuringly, we found that AMI diagnosis extracted from coded primary care diagnosis is as reliable as secondary care, supporting the combined use of such data. Thus, the
Box 1

Actions to improve the use of EHR data for health outcome assessment, using AMI as an example.

<table>
<thead>
<tr>
<th>Stage in data cycle</th>
<th>Specific for AMI (n studies)</th>
<th>Barriers</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical acquisition of data features used in diagnosis</td>
<td>Laboratory blood values</td>
<td>Troponin values (1); Not specified laboratory values (1)</td>
<td>Data fragment and stored in different clinical information systems</td>
</tr>
<tr>
<td></td>
<td>Diagnostic tests</td>
<td>Electrocardiogram (0)</td>
<td>Proprietary, different machines; different access Not always stored in electronic format</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Chest pain (1)</td>
<td>Text contains personal identifiers</td>
<td>Standardised and shareable</td>
</tr>
</tbody>
</table>

| Combination of clinical data features to ascertain patients with diagnosis | Algorithms | None | Mostly limited to use of diagnosis codes from primary or secondary care Different coding systems and versions used | Standardised and shareable |
| Contextual information | None | Largely not known, not systematically recorded and often unclear | Centralised metadata portals and mechanisms to record and disseminate contextual information Greater clinician input into coding and documentation of process |
| EHR sources | | Multiple EHR sources not systematically linked | Combine multiple EHR sources for better accuracy |

The ‘barriers’ column depicts some of the obstacles currently hampering the availability of data in easily accessible and usable electronic format; the ‘actions’ column offers actions to overcome the obstacles described in the previous column. *LOINC: Logical Observation Identifiers Names and Codes, internationally standardised database for identifying medical laboratory observations

The combination of multiple EHR sources may enhance the accuracy of AMI ascertainment, despite some studies identifying low rates of concordance [27,47].

We found that disease prevalence was underreported (3/33) and most studies did not specify if cases were spontaneous AMI (type 1), MI secondary to an ischaemic imbalance (type 2), cardiac death due to MI (type 3) or procedure-related AMI (types 4 and 5) [3]. In addition, only a third of the studies reported sensitivity and five calculated specificity. These findings indicate that there is a clear need to standardise conducting and reporting of validation studies.

This systematic review has limitations. Grey literature was not searched and it is possible that relevant literature was missed. Studies were so heterogeneous that formal meta-analysis was not carried out, and therefore we could not estimate the extent of publication bias. Additionally, studies might have used some of the clinical data features to ascertain or validate AMI despite these not being reported so greater efforts are needed to standardise the reporting of validation studies.

5. Conclusions

Clinical coding of EHR-derived AMI diagnosis in primary care and secondary care has moderate to high accuracy in different clinical settings and for different phenotypes. However, markers of myocardial necrosis values, ECG findings and symptom descriptions, the cornerstones of a clinical AMI diagnosis, are underutilised and remain a challenge to access and retrieve from EHRs. Efforts are needed to tackle the barriers that currently impede optimised, and more accurate, use of EHRs for AMI ascertainment.

**Contributors**

BR, NKF, MD, HH participated in the original conception and design of the study. BR, NKF acquired the data. BR, MD, NY interpreted the data, performed statistical analysis and created the figures. BR, NKF, SD, RSP drafted the manuscript. NKF, SD, MD, NY, RSP, HH, UK Biobank Follow-up and Outcomes Working Group revised the manuscript critically for important intellectual content. All authors have approved the submitted version.

**Conflict of interests**

None.

**Acknowledgments**

The authors would like to acknowledge the following members of the UK Biobank Follow-up and Outcomes Working Group for their valuable input: John Danesh (chair), Naomi Allen, Mark Atkinson, Ekaterini Blaveri, Rachael Brannan, Carol Brayne, Sinead Brophy, Nish Chaturvedi, Rory Collins, Simon delusignan, Spiros Denaxas, Parul Desai, Sophie Eastwood, John Gallacher, Harry Hemingway, Matthew Hotopf, Martin Landray, Ronan Lyons, Mark McGilchrist, Henrik Moller, Terence O’Neil, Mike Pringle, Tim Sprosen, David Strachan, Cathie Sudlow, Frank Sulli-van, Rebecca Woodfield, Qiuli Zhang and Robin Flaig.

**References**

[4] Quality of Care and Outcomes Research in CVD and Stroke Working Groups, Mea-
suring and improving quality of care: a report from the american heart associa-
tion/american college of cardiology first scientific forum on assessment of
healthcare quality in cardiovascular disease and stroke, Circulation 101 (12)

mortality – how to count does matter for patients hospitalized for acute myocardial

ison of short-term survival in national outcome registries in Sweden and the UK,

follow-up of the west of scotland coronary prevention study, N. Engl. J. Med. 357

[8] O. Fröhbert, B. Lagerqvist, G.K. Olvecrona, et al., Thrombus aspiration during ST-
1587–1597.

cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific

[10] The West of Scotland Coronary Prevention Study Group, Computerised record
linkage: compared with traditional patient follow-up methods in clinical trials
and illustrated in a prospective epidemiological study, J. Clin. Epidemiol. 48 (12)

I. Hofmans-Okkes, H. Lamberts, The International Classi-
fication of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease
Prevention and Control and Prevention; and the National Heart, Lung, and Blood Institute, Circula-

[12] V.L. Roger, J. Killian, M. Henkel, et al., Coronary disease surveillance in Olmsted

[13] E. Herrett, A.D. Shah, R. Boggon, et al., Completeness and diagnostic validity of re-
corded acute myocardial infarction events in primary care, hospital care, disease
registry, and national mortality records: cohort study, BMJ 346 (2013).

incidence and outcomes of acute myocardial infarction, N. Engl. J. Med. 362 (23)


[16] J.K. French, H.D. White, Clinical implications of the new definition of myocardial

400–403.

and the ECG: evidence of an adverse QT effect on corrected QT interval, Ann. Nonin-


e206–e211.

in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-
Analyses of Genome-Wide Association Studies From 5 Cohorts, Circ. Cardiovasc.

[22] Task force on the management of ST-segment elevation acute myocardial infarc-
tion: effects on treatment and mortality, and implications for National Service
Delivery/american college of cardiology/american heart association forum on assessment of
cardiac conduction identi-


diagnosis-related groups, Med. Care 18 (2) (1980) i-53.


[27] I. Hofmans-Ookse, H. Lamberts, The International Classification of Primary Care (KPC):
new applications in research and computer-based patient records in family

[28] R.G. Newcombe, Two-sided confidence intervals for the single proportion: compar-

and the ECG: evidence of an adverse QT effect on corrected QT interval, Ann. Nonin-

a phenotype-wide scan to discover gene–disease associations, Bioinformatics 26

749–756.

[32] N. McCormick, D. Lacalle, V. Bhole, J.A. Avina-Zubieta, Validity of myocardial infarc-
tion diagnoses in administrative databases: a systematic review, PLoS ONE 9 (3)
(2014) e92286.

[33] A. Metcalfe, A. Neudam, S. Forde, et al., Case definitions for acute myocardial infarc-
tion in administrative databases and their impact on in-hospital mortality rates,

[34] J.S. Alpert, K. Thygesen, E. Antman, J.P. Bassand, Myocardial infarction redefi-
ned—a consensus document of The Joint European Society of Cardiology/American College
of Cardiology Committee for the Redefinition of Myocardial Infarction, J. Am. Coll.
Cardiol. 37 (3) (2001) 973.

heart disease in epidemiology and clinical research studies: a statement from the
AHA Council on Epidemiology and Prevention; AHA Statistics Committee: World
Heart Federation Council on Epidemiology and Prevention; the European Society of
Cardiology Working Group on Epidemiology and Prevention; Centers for Disease
Control and Prevention; and the National Heart, Lung, and Blood Institute, Circula-

[36] L. Gronski, W. Martinson, K.U. Heitmann, LOINC® codes for hospital infor-
400–403.