RESEARCH

Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study

OPEN ACCESS

Emily Herrett *research fellow*[1], Anoop Dinesh Shah *clinical research fellow*[2], Rachael Boggon *research statistician*[3][4], Spiros Denaxas *senior research associate*[2], Liam Smeeth *professor of clinical epidemiology and general practitioner*[1], Tjeerd van Staa *professor of pharmacoepidemiology*[1][3][4], Adam Timmis *professor of clinical cardiology*[5], Harry Hemingway *professor of clinical epidemiology*[2]

[1]London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; [2]Department of Epidemiology and Public Health, Clinical Epidemiology Group, University College London, UK; [3]Clinical Practice Research Datalink Group, Medicines and Healthcare products Regulatory Agency, London, UK; [4]Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, Netherlands; [5]Barts and the London School of Medicine and Dentistry, London, UK

Abstract

Objective To determine the completeness and diagnostic validity of myocardial infarction recording across four national health record sources in primary care, hospital care, a disease registry, and mortality register.

Design Cohort study.

Participants 21 482 patients with acute myocardial infarction in England between January 2003 and March 2009, identified in four prospectively collected, linked electronic health record sources: Clinical Practice Research Datalink (primary care data), Hospital Episode Statistics (hospital admissions), the disease registry MINAP (Myocardial Ischaemia National Audit Project), and the Office for National Statistics mortality register (cause specific mortality data).

Setting One country (England) with one health system (the National Health Service).

Main outcome measures Recording of acute myocardial infarction, incidence, all cause mortality within one year of acute myocardial infarction, and diagnostic validity of acute myocardial infarction compared with electrocardiographic and troponin findings in the disease registry (gold standard).

Results Risk factors and non-cardiovascular coexisting conditions were similar across patients identified in primary care, hospital admission, and registry sources. Immediate all cause mortality was highest among patients with acute myocardial infarction recorded in primary care, which (unlike hospital admission and disease registry sources) included patients who did not reach hospital, but at one year mortality rates in cohorts from each source were similar. 5561 (31.0%) patients with non-fatal acute myocardial infarction were recorded in all three sources and 11 482 (63.9%) in at least two sources. The crude incidence of acute myocardial infarction was underestimated by 25-50% using one source compared with using all three sources. Compared with acute myocardial

Correspondence to: E Herrett emily.herrett@lshtm.ac.uk

infarction defined in the disease registry, the positive predictive value of acute myocardial infarction recorded in primary care was 92.2% (95% confidence interval 91.6% to 92.8%) and in hospital admissions was 91.5% (90.8% to 92.1%).

Conclusion Each data source missed a substantial proportion (25-50%) of myocardial infarction events. Failure to use linked electronic health records from primary care, hospital care, disease registry, and death certificates may lead to biased estimates of the incidence and outcome of myocardial infarction.

Trial registration NCT01569139 clinicaltrials.gov.

Introduction

Electronic health records inform patient decision making and policy and are increasingly used to define disease and the outcomes of care in observational cohorts of genetic and environmental factors[1-4] and randomised trials.[5-7] Recent initiatives to expand the use of health records for research have been announced in many countries,[8-11] and in the United Kingdom the National Health Service is now legally required to evaluate patient outcomes.[12] The UK government has also recently announced plans to drive improvement in cardiovascular disease care through use of information in linked health records.[13] In various settings across the world these initiatives are being met by the linkage of electronic health records from disparate sources. Underpinning these uses of electronic health records is the need for a better understanding of the quality of data within a single source as well as between multiple sources. Indeed, it is a concern that electronic records from one part of the health system, such as primary care, may not capture health events occurring in other parts of the health system, such as hospital care.

As part of the CArdiovascular disease research using LInked Bespoke studies and Electronic health Records (CALIBER) programme[14] we carried out new linkage between records from primary care,[15] hospitals, an acute coronary syndrome registry,[16] and death certificates. Although data from these types of source are increasingly available in different countries,[17 18] for acute myocardial infarction the overlap between these four electronic health record sources, patient risk factors, and subsequent mortality have not been compared. Previous cross referencing studies have typically compared one or two electronic sources, such as coded hospital discharge diagnoses and cause of death, with case note review,[19-21] questionnaires to general practitioners,[22] or active case finding in a prospective consented study[1-24] (see supplementary table 1). Linkages with the national, ongoing acute coronary syndrome registry allowed detailed diagnoses of myocardial infarction (with coded electrocardiographic findings and markers of myocardial necrosis, not available in other sources) to be compared with diagnoses in primary care and hospital admissions. Linkages with primary care allowed evaluation of risk factors in patients with a record of acute myocardial infarction in any source. Linkages with the death record allowed evaluation of cause specific mortality of myocardial infarction recorded in any source, including among cases not admitted to hospital.

We compared the incidence, recording, agreement of dates and codes, risk factors, and all cause mortality of acute myocardial infarction recorded in primary care, hospital care, the national acute coronary syndrome registry, and the national death registry.

Methods

We used a cohort study design, identifying patients with acute myocardial infarction in four prospectively collected, linked electronic health record sources in England (the CALIBER programme[14]). Briefly, the CALIBER linkage included anonymised primary care electronic patient records from the Clinical Practice Research Datalink[15] (www.cprd.com, formerly known as the General Practice Research Database), data on hospital admissions from Hospital Episode Statistics, the national registry of acute coronary syndromes (Myocardial Ischaemia National Audit Project, MINAP),[16] and the death registry, curated by the Office for National Statistics (see supplementary table 2).

Of the 630 primary care practices in Clinical Practice Research Datalink, 244 consented to data linkage with Hospital Episode Statistics, MINAP, and the Office for National Statistics. These practices contained 3.9% of the population of England in 2006. The linkage was carried out in October 2010 by a trusted third party, using a deterministic match between NHS number, date of birth, and sex. Overall, 96% of patients with a valid NHS number were successfully matched.

Study population: patients with acute myocardial infarction

We identified records of acute myocardial infarction with reference to previously described definitions for each source. In primary care, diagnoses are recorded using Read codes[25] and previous studies have published lists of Read codes used to identify acute myocardial infarction.[26 27] We identified myocardial infarction using the 62 Read codes listed in supplementary table 3. In Hospital Episode Statistics and the Office for National Statistics death registry, diagnoses are coded using the International Classification of Diseases, 10th revision[28]; in common with previous studies,[29] we defined acute myocardial infarction by ICD-10 codes I21 (acute myocardial infarction), I22 (subsequent myocardial infarction), or I23 (current complications following acute myocardial infarction). In Hospital Episode Statistics, to be included in our study myocardial infarction had to be recorded as the primary diagnosis in the first episode of an admission to hospital (where the first episode refers to the first period of care for an admitted patient overseen by a healthcare professional[30]). We performed a sensitivity analysis to assess the influence of inclusion of secondary diagnoses. In MINAP, ST elevation and non-ST elevation myocardial infarction were identified using hospital discharge diagnosis, markers of myocardial necrosis, and coded electrocardiographic findings, in accordance with the internationally agreed definition of myocardial infarction.[31] For MINAP and Hospital Episode Statistics, we took the hospital admission date to represent the date of acute myocardial infarction.

The study period was 1 January 2003 to 31 March 2009 (when all record sources were concurrent) and confined to patients who had been registered with their general practice for at least a year and the practice had been submitting data for at least one year that met Clinical Practice Research Datalink data quality standards for continuity and plausibility of data recording. In the main analysis we included only patients with at least one record of admission to hospital in Hospital Episode Statistics at any time (for any cause) as these patients were shown to be linkable, but we conducted a sensitivity analysis including all patients. We selected the first record of myocardial infarction during the patient's study period as the index event and considered myocardial infarction records in the other data sources as representing the same event if they were dated within 30 days of the index event.

Cardiovascular risk factors and non-cardiovascular coexisting conditions

For patients with acute myocardial infarction we identified risk factors recorded in primary care, including age, sex, social deprivation,[32] smoking, use of antihypertensives or lipid lowering drugs, diabetes mellitus, Charlson comorbidity index,[33] and primary care consultation rate before the event. We used mean measures of systolic blood pressure and total cholesterol and high density lipoprotein cholesterol levels before myocardial infarction along with age, sex, and smoking status (where these variables were present) to estimate the 10 year Framingham risk for acute myocardial infarction or coronary death.[34] We used these measures to compare the cohorts of myocardial infarction identified in each data source.

Follow-up for mortality

We followed all patients with a record of myocardial infarction in any source for one year for death as recorded in the Office for National Statistics death registry. We categorised patients as having fatal or non-fatal myocardial infarction by whether they died of any cause within seven days of the myocardial infarction. If a patient had a myocardial infarction record in Clinical Practice Research Datalink, Hospital Episode Statistics, or MINAP after their date of death, we considered that they died on the day of their myocardial infarction.

Agreement in recording

If the time difference between the earliest date of acute myocardial infarction in one source and the date in another source was no more than 30 days we considered that the records of acute myocardial infarction in the different sources agreed. A myocardial infarction recorded more than 30 days after the earliest date was considered a new event and was not included, ensuring that each patient appeared only once in the analysis. We chose 30 days to account for any delay in recording of myocardial infarction in primary care, assuming that any record within 30 days of a hospital admission was likely to represent the same event and anything after 30 days could feasibly be a subsequent myocardial infarction. We carried out a sensitivity analysis using a 90 day threshold.

Statistical analysis

Incidence

We estimated population based incidence rates of fatal and non-fatal acute myocardial infarction using the denominator of all adults in the CALIBER primary care population aged 18 and over (2.2 million), followed up for a mean 4.1 years between 2003 and 2009. We used each of the data sources separately and together to identify incident myocardial infarction, ending the follow-up period for a patient on the date of their first myocardial infarction during the study period, death, or deregistration from the general practice.

Cardiovascular risk factors and non-cardiovascular coexisting conditions

We compared patients with fatal and non-fatal acute myocardial infarction identified in the four data sources for risk factors and coexisting conditions recorded in primary care.

Death after acute myocardial infarction

We produced cumulative incidence curves for coronary and non-coronary mortality for patients recorded in each data source and compared mortality using a Cox proportional hazards model adjusted for age and sex.

Agreement in recording

We would expect patients who survived seven days after myocardial infarction to be recorded in the primary care, disease registry, and hospital admissions sources, and we assessed agreement between these three sources in a Venn diagram. For patients who died within seven days, we examined the proportion recorded by each source but did not compare agreement across all four sources as we would not expect the hospital discharge data and disease registry to record patients who died before reaching hospital.

In patients who did not have a record of acute myocardial infarction in one or more data sources, we looked for other codes that may have been used to describe the event. In the disease registry, we looked for unstable angina or admission diagnoses of any acute coronary syndrome. In primary care and hospital discharge data, we sought other acute coronary syndromes, coronary disease, chest pain, or other cardiac diagnoses (for example, atrial fibrillation, heart failure, cardiac arrest). In primary care data we also examined codes indicating contact with secondary care. Where none of these codes was recorded, we tabulated all recorded codes in the 30 days before and after the date of myocardial infarction to see if there were any relevant codes we had overlooked.

We performed a logistic regression analysis to establish whether age, sex, deprivation, rate of primary care consultation, year of myocardial infarction, or mortality at 30 days explained suboptimal recording of acute myocardial infarction in primary care, hospital discharge, or disease registry sources.

We calculated the positive predictive value of primary care or hospital discharge diagnoses of acute myocardial infarction among patients who also had a record in the acute coronary syndrome registry. Data were analysed using Stata 12 and R 2.14.1.[35]

Results

We identified 21 482 patients with fatal or non-fatal acute myocardial infarction recorded in any of the four data sources.

Incidence

Among the single source crude estimates for incidence of myocardial infarction, primary care data (Clinical Practice Research Datalink) gave the highest estimate, of 187 per 100 000 patient years (95% confidence interval 184 to 190), followed by hospital discharge data (Hospital Episode Statistics) with 154 per 100 000 patient years (152 to 157), acute coronary syndrome registry (MINAP) with 115 per 100 000 patient years (113 to 118), and death registry with 45 per 100 000 patient years (43 to 46). Combining these three sources yielded an estimate of 243 per 100 000 patient years (239 to 246, fig 1⇓). The crude incidence of acute myocardial infarction was 25% lower using only Clinical Practice Research Datalink and 50% lower using only MINAP compared with using all three sources. See supplementary figure 1 and table 4 for standardised incidence by age, sex, and region.

Cardiovascular risk factors and comorbidity

Overall, the cohorts identified from the primary care, hospital, and disease registry sources had a similar prevalence of cardiovascular risk factors and comorbidities. However, compared with those recorded in the acute coronary syndrome

registry or hospital discharge data only, patients with fatal or non-fatal myocardial infarction recorded only in primary care were on average two years younger and more likely to be current smokers and in the most deprived fifth (P<0.001 for these comparisons, also see supplementary table 5). Patients recorded by the death registry were older than patients recorded in the other sources and had a higher burden of risk factors reflecting their age. However, other demographic characteristics and cardiovascular risk factors were broadly similar across patients recorded in primary care and hospital care sources (table 1⇓, also see supplementary table 5).

Death after acute myocardial infarction

Patients with myocardial infarction identified in the disease registry had lower crude 30 day mortality (10.8%, 95% confidence interval 10.2% to 11.4%) than those identified in hospital care (13.9%, 13.3% to 14.4%) or in primary care (14.9%, 14.4% to 15.5%, fig 2⇓). At one year, however, mortality was similar in all three groups, at around 20%.

In the linked data, patients with acute myocardial infarction recorded in only one source had higher mortality than those recorded in more than one source (age and sex adjusted hazard ratio 2.29, 95% confidence interval 2.17 to 2.42; P<0.001). Among patients with myocardial infarction recorded in only one source (Hospital Episode Statistics, Clinical Practice Research Datalink, or MINAP), those recorded only in primary care had the highest mortality on the first day but the lowest mortality thereafter (see supplementary figures 2 and 3). Among patients with myocardial infarctions recorded in one of Hospital Episode Statistics or MINAP but not both, those in MINAP had lower coronary mortality in the first month (age and sex adjusted hazard ratio 0.33, 0.28 to 0.39, P<0.001) but similar mortality for non-coronary events (1.12, 0.90 to 1.40, P=0.3). After the first month, patients with myocardial infarctions recorded only in primary care had about half the hazard of mortality of patients with myocardial infarctions recorded in one of MINAP or Hospital Episode Statistics (hazard ratio adjusted for age and sex for coronary causes 0.49, 95% confidence interval 0.40 to 0.60, P<0.001 and for other causes 0.57, 0.49 to 0.67, P<0.001). Of the 3518 patients with myocardial infarction recorded in any of the four sources who died of any cause within seven days, 54.4% (n=1914) had a myocardial infarction code recorded in primary care within 30 days. The underlying cause of death was acute myocardial infarction in 2924 patients (83.0%); a further 164 patients (4.7%) had ischaemic heart disease recorded as the underlying cause of death (ICD-10 code I20, I24, or I25), 60 (1.7%) had cerebrovascular disease (I60-I69), and 85 (2.4%) had respiratory disease (J00-J99). However, 3375 of these 3518 patients (95.9%) had a coronary diagnosis (I20-I25) as either the underlying cause or a secondary cause of death.

Fatal myocardial infarctions identified by death registry data (underlying cause of death ICD-10 I21, I22, or I23, n=2919) were unlikely to be recorded in hospital sources; 36.7% (n=1072) were recorded in Hospital Episode Statistics and just 17.1% (n=498) in the MINAP disease registry within 30 days, but 55.9% (n=1631) were recorded in primary care (see supplementary table 6).

Non-fatal acute myocardial infarction: agreement between record sources

Among the 17 964 patients with at least one record of non-fatal acute myocardial infarction, 13 380 (74.5%) were recorded by Clinical Practice Research Datalink, 12 189 (67.9%) by Hospital Episodes Statistics, and 9438 (52.5%) by MINAP. Overall, 5561

(31.0%) of patients had the event recorded in all three sources and 11 482 (63.9%) in at least two sources (fig 3⇓). When we extended the recording window from 30 days to 90 days, the proportion recorded in all three sources increased only slightly, to 32.0% (n=5747). When we included patients who had never had a record of a hospital admission in Hospital Episode Statistics, the proportion of non-fatal myocardial infarctions recorded in all three sources decreased slightly to 30.0% (5561/18 536) and the proportion recorded only in primary care increased from 17.7% (3188/17 964) to 20.3% (3760/18 536). A sensitivity analysis in which the Hospital Episode Statistics case definition included secondary diagnoses of myocardial infarction (where myocardial infarction was not the reason for admission) produced only a slight increase in the proportion recorded in all three sources (5812/18 283, 32.0%), and identified 306 additional myocardial infarctions that were not in any other source.

The exact date of admission agreed in over 80% of 6851 patients with acute myocardial infarction recorded in both hospital care and disease registry sources (see supplementary figure 4), but the date recorded in primary care was the same as the disease registry admission date or hospital admission date for only 50% of patients (n=15 753). There was a smaller peak in primary care recording between five and seven days after admission. When the time window was extended to 90 days, there was little change in these proportions.

Among patients with non-fatal myocardial infarction, 88.0% (8304/9438) of those recorded in MINAP and 89.1% (10 859/12 189) recorded in Hospital Episode Statistics had a Read code for any cardiac diagnosis or chest pain within 30 days in primary care, and in over 70% the Read code stated myocardial infarction (see supplementary table 6). Only 25.1% (3364/13 380) of the non-fatal myocardial infarctions recorded in primary care stated the type—that is, ST elevation or non-ST elevation—compared with 100% for the disease registry. If a non-fatal myocardial infarction was recorded in primary care, hospital discharge data recorded a cardiac diagnosis within 30 days in 84.9% (11 355/13 380) of patients, with a primary diagnosis of myocardial infarction in 72.6% (9720/13 380). However, this proportion varied depending on the Read term used to identify myocardial infarction in primary care; for terms that state the anatomical location (for example, acute anterolateral infarction) it was around 80% but was lower for less precise terms. For example, of the 74 patients with the Read term "heart attack," only 32 (43%) had a primary hospital diagnosis of myocardial infarction. Supplementary tables 7-9 describe the agreement between sources according to the way in which acute myocardial infarction was recorded in each source.

Positive predictive value

For primary care or hospital discharge patients with an associated record in the disease registry (MINAP), the positive predictive value of the acute myocardial infarction diagnosis (the probability that the diagnosis recorded in the disease registry was myocardial infarction rather than unstable angina or a non-cardiac diagnosis) was 92.2% (6660/7224, 95% confidence interval 91.6% to 92.8%) in primary care and 91.5% (6851/7489, 90.8% to 92.1%) in hospital care (table 2⇓). Eighty five per cent of patients recorded in primary care and hospital discharge (7386/8707) had a record of raised cardiac markers and half (3766/8707) had a record of ST segment elevation on electrocardiography.

Non-fatal acute myocardial infarction: reasons for disagreement

Compared with patients who had a record of acute myocardial infarction in only one source, those with records in multiple sources had a lower rate of primary care consultation before the event, were younger, were more likely to be male, and more likely to have experienced acute myocardial infarction in one of the later years of data collection. Among patients with myocardial infarction recorded in primary care, an additional record in Hospital Episode Statistics or MINAP was strongly associated with increased mortality at 30 days (see supplementary table 10).

Discussion

We compared electronic health records on one major disease event—acute myocardial infarction—across four English, ongoing sources of health record data: primary care (Clinical Practice Research Datalink), hospital admissions (Hospital Episode Statistics), a quality improvement disease registry (Myocardial Ischaemia National Audit Project, MINAP), and the death registry (Office for National Statistics). In over 20 000 patients each data source missed a substantial proportion of myocardial infarction events. We also found evidence for the validity of myocardial infarction recording across all sources, in terms of risk factor profiles and mortality at one year. Taken together, these findings support the wider use of linkage of multiple record sources by clinicians, policy makers, and researchers.

Fatal myocardial infarction

Both primary care and death registry data can be used to capture fatal myocardial infarction occurring out of hospital among people without a record of myocardial infarction in the Hospital Episode Statistics or disease registry (MINAP). The death registry is a useful source of fatal acute myocardial infarction for research, as most (83.0%) patients who were identified as having acute myocardial infarction in any of the data sources and died within seven days had myocardial infarction recorded as their underlying cause of death. These figures agree with results from the Oxford Record Linkage Study, where among 5686 patients admitted to hospital with myocardial infarction 85.2% who died within 30 days had myocardial infarction recorded as the underlying cause of death.[36]

Non-fatal myocardial infarction

Primary care captures most cases, but all sources miss non-fatal myocardial infarction

We found that each record source misses cases. Only one third of non-fatal myocardial infarctions were recorded in all three data sources (primary care, hospital admissions, and disease registry) and two thirds were recorded in at least two sources. Clinical Practice Research Datalink was the single most complete source of non-fatal myocardial infarction records (one quarter of all non-fatal myocardial infarction events not recorded), Hospital Episode Statistics missed one third, and MINAP missed nearly half (fig 3). This agrees with the results of other studies; a two source study of myocardial infarction in Scotland (see supplementary table 1) compared the incidence based on primary care records with that based on hospital data and showed that in combination they provided the highest estimates of incidence.[37] Further two source comparisons in Australia,[38] Denmark,[39] and the Netherlands[24] (see supplementary table 1) have shown that hospital records alone underestimate

the true incidence of myocardial infarction. Despite the low sensitivity of these data sources, in our study the positive predictive value of myocardial infarction records in primary care and hospital admission sources were over 90% compared with the disease registry gold standard based on the international definition of myocardial infarction (table 2).

However, some of the myocardial infarctions recorded only in primary care are likely to be historical diagnoses because the associated mortality rate in the first month is much lower than those also recorded in hospital sources (see supplementary figure 2). Our results using cross referencing of electronic health records in 20 000 patients are consistent with previous manual approaches to validation in primary care, which cross reference a few hundred patients with disease diagnoses recorded using Read codes against anonymised free text, death certificates, paper medical records, or hospital discharge summaries, or questionnaires to general practitioners.[40-48] Our much larger sample size, however, allowed us to evaluate individual Read terms that are used to record myocardial infarction. This type of validation has not been done previously for myocardial infarction and may be relevant to other common conditions that can be recorded using a variety of codes, such as stroke.[49]

Hospital admission data

To our knowledge, no studies have examined the positive predictive value of ICD-10 coded myocardial infarction diagnosis in hospital admission data against an ongoing disease registry. We found that 62.8% of non-fatal myocardial infarctions recorded in primary care and 72.6% recorded in the disease registry were recorded by hospital admissions data. This is consistent with a single electronic health record source (Hospital Episode Statistics ICD-10 I21 and I22) capturing 53% of myocardial infarctions in an investigator led cohort with active follow-up.[50]

Disease registry and maximising true positives

The strengths of the disease registry MINAP lie in the fact that its diagnostic records (troponin values, electrocardiographic findings, and cardiologist diagnosis of ST elevation and non-ST elevation myocardial infarction) are not available in other sources, which offer validated endpoints electronically from all hospitals in England and Wales. An acute myocardial infarction recorded in MINAP is thus an electronic health record gold standard, as a myocardial infarction recorded by a registry is likely to fulfil international diagnostic criteria.[16] The registry may be important for detecting endpoints in cohort studies and trials, where false positives can dilute any observed effect and reduce the power of a study. Furthermore, in such studies it has been shown that avoiding false positives is more important than avoiding false negatives.[51] Validation of myocardial infarctions recorded by primary care and hospital admissions against those recorded by the disease registry showed a positive predictive value of over 90%, making them suitable for detecting endpoints in cohort studies and trials, where poor endpoint resolution can dilute any observed effect and reduce the power of a study.[51] The positive predictive value was not 100% because some myocardial infarction records in primary care may actually have been related to unstable angina or chest pain of an unknown cause.

Limitations of this study

Our data were from a sample of 244 English general practices contributing to the Clinical Practice Research Datalink. However, the primary care patients included in this CALIBER

study are representative of the UK population. Furthermore, patients in practices that participated in the linkage were representative of the Clinical Practice Research Datalink as a whole in terms of age, social deprivation, body mass index, and prescription of key drugs.[52] Hospital admissions with linked primary care data were also representative of all admissions to hospital in England in terms of the distribution of age, sex, and diagnostic group.[53] The UK life science strategy aims to increase the proportion of the UK population with primary care data available for research linked electronic health record through the Clinical Practice Research Datalink.[15] A second limitation concerns the generalisability of our findings for the quality of primary care data. Practices contributing data to the Clinical Practice Research Datalink are advised of recording guidelines and their data are accepted only when they meet standards of data completeness,[15] so they are likely to record disease events better than general practices that do not contribute. Our estimates of agreement from this study may therefore be higher than for practices that do not contribute to the Clinical Practice Research Datalink. Thirdly, our validation of Hospital Episode Statistics and Clinical Practice Research Datalink myocardial infarctions against MINAP was (inherently) limited to the subset of patients with a MINAP record, and caution must be exercised in extending these conclusions to patients with myocardial infarctions without a MINAP record.

Clinical and policy implications

With the current emphasis on measuring clinical outcomes in health systems and recent plans to use linked data to drive improvements in the care of patients with cardiovascular disease,[4 13] our study has important implications for practice and policy. Firstly, we propose much wider use of linked record sources in commissioning and in research to estimate disease occurrence and outcome, because of the biases inherent in using only one source of records. Changing the estimates for incidence of myocardial infarction could potentially alter the modelled effect of population based healthcare interventions. Our findings underscore the importance of international initiatives to accelerate availability of linked data in America,[11-55] in Europe,[56] and elsewhere. Secondly, a national strategy for biomedical informatics is required to tackle manifest system failings: a single health event should, ideally, have a single record that is propagated in multiple record systems. Efforts to reform the process of death certification is already underway,[57] but this needs to be broadened to include other health records. Thirdly, and more specifically, primary care records could be improved if the admission date rather than discharge date was used to record the myocardial infarction (as reflected by the current "tail" of myocardial infarction records recorded up to 20 days after admission; see supplementary figure 4) and if acute myocardial infarction was recorded only for its occurrence rather than repeated entries for consultations related to a history of myocardial infarction. Fourthly, disease registries, such as MINAP, could be improved if embedded in real time clinical care of all patients with myocardial infarction, rather than the current situation in which hospitals employ audit staff to retrospectively enter records on patients in coronary care units. This needs to be dealt with to obtain a more complete understanding of the quality of care provided to patients with acute coronary syndromes.

Future research

Several lines of research are warranted by our findings. Firstly, research is required to understand how electronic health record data are coded—historically under-resourced and lacking audit

against quality standards—and how this can be improved. Secondly, more extensive cross referencing is required against additional sources of information on myocardial infarction. These include self reported myocardial infarction (which may be less dependent on specific setting in the health system), manual review of all the available local case records (paper and electronic), and investigation of electronic free text recorded by general practitioners (for example, diagnoses that are not recorded using a Read code). Such efforts are underway in the UK Biobank cohort (n=500 000).[3] There is a need for investigator led cohorts and trials to link with the primary care record.[58] Although cancer registries do not record gold standard diagnostic criteria or cancer stage, it will be important to understand how linkages with primary care, admission to hospital, and mortality data compare.[59] This is essential for large studies where manual review of case records is not feasible. Evaluating the quality of the data available in these linked data sources is therefore a priority.

Conclusion

Failure to use linked electronic health records from primary care, hospital care, disease registry, and death certificates may lead to biased estimates of the incidence and outcome of myocardial infarction.

What is already known on this topic

Electronic health records are increasingly used to measure outcomes of healthcare and health policy, and for research in observational cohorts and randomised trials

Records from one part of the health system, such as primary care, may not capture health events occurring in other parts of the health system, such as hospital care

No studies have addressed the completeness and validity of recording of myocardial infarction across four national health record sources: primary care, hospital care, disease registry, and death records

What this study adds

About one third of patients had a record of non-fatal acute myocardial infarction in all three of primary care, hospital care, and disease registry and two thirds in two sources

Risk factor profiles and one year all cause mortality were comparable across myocardial infarction records from different sources

Crude incidence of acute myocardial infarction was underestimated by 25-50% using one source compared with using all three sources

1   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* 1995;48:1441-52.
2   Ford I, Murray H, Packard CJ, Shepherd J, Macfarlane PW, Cobbe SM. Long-term follow-up of the west of Scotland coronary prevention study. *N Engl J Med* 2007;357:1477-86.
3   Collins R. What makes UK Biobank special? *Lancet* 2012;379:1173-4.
4   Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
5   Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. *Health Technol Assess* 2000;4:1-55.
6   Nieminen P, Kotaniemi L, Hakama M, Tarkkanen J, Martikainen J, Toivonen T, et al. A randomised public-health trial on automation-assisted screening for cervical cancer in Finland: performance with 470,000 invitations. *Int J Cancer* 2005;115:307-11.
7   Gulliford MC, van Staa T, McDermott L, Dregan A, McCann G, Ashworth M, et al. Cluster randomised trial in the general practice research database: 1. Electronic decision support to reduce antibiotic prescribing in primary care (eCRT study). *Trials* 2011;12:115.
8   Office for Life Sciences. Strategy for UK Life Sciences, 2011. Department for Business, Innovation and Skills, 2013. www.bis.gov.uk/ols.
9   Steinbrook, R. Health Care and the American Recovery and Reinvestment Act. *New Engl J Med* 2009;360:1057-60.
10 Almqvist C, Adami HO, Franks PW, Groop L, Ingelsson E, Kere J, et al. LifeGene—a large prospective population-based study of global relevance. *Eur J Epidemiol* 2011;26:67-77.
11 Health Information Technology for Economic and Clinical Health Act Title XIII—Health Information Technology; in: American Recovery and Reinvestment Act of 2009. 2009. Feb 17, Public Law No 111-5, 123 Stat. 115.
12 A mandate from the Government to the NHS Commissioning Board: April 2013 to March 2015. Department of Health, 2012. 2013. http://mandate.dh.gov.uk/.
13 Cardiovascular Disease Outcomes Strategy, Department of Health, March 2013. www.wp.dh.gov.uk/publications/files/2013/03/9387-2900853-CVD-Outcomes_web1.pdf.
14 Denaxas S, George J, Herrett E, Shah A, Kalra D, Hingorani AD, et al. Data resource profile: Cardiovascular Disease Research using Linked Bespoke Studies and Electronic Records (CALIBER). *Int J Epidemiol* 2012;41:1625-38.
15 Clinical Practice Research Datalink, 2012. 2013. www.cprd.com.
16 Herrett E, Smeeth L, Walker L, Weston C. The Myocardial Ischaemia National Audit Project (MINAP). *Heart* 2010;96:1264-7. doi:10.1136/hrt.2009.192328S.
17 Jernberg T, Attebring MF, Hambraeus K, Ivert T, James S, Jeppsson A, et al. The Swedish web-system for enhancement and development of evidence-based care in heart disease evaluated according to recommended therapies (SWEDEHEART). *Heart* 2010;96:1617-21. doi:10.1136/hrt.2010.198804.
18 Institute for Clinical Evaluative Sciences (ICES), Canada. 2013. www.ices.on.ca/.
19 Rapola JM, Virtamo J, Korhonen P, Haapakoski J, Hartman AM, Edwards BK, et al. Validity of diagnoses of major coronary events in national registers of hospital diagnoses and deaths in Finland. *Eur J Epidemiol* 1997;13:133-8.
20 Hammar N, Alfredsson L, Rosén M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. *Int J Epidemiol* 2001;30(Suppl 1):S30-4.
21 Barchielli A, Balzi D, Naldoni P, Roberts AT, Profili F, Dima F, et al. Hospital discharge data for assessing myocardial infarction events and trends, and effects of diagnosis validation according to MONICA and AHA criteria. *J Epidemiol Community Health* 2012;66:462-7.
22 Wright FL, Green J, Canoy D, Cairns BJ, Balkwill A, Beral V. Million Women Study Collaborators. Vascular disease in women: comparison of diagnoses in hospital episode statistics and general practice records in England. *BMC Med Res Methodol* 2012;12:161.
23 Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Räihä P, Kärjä-Koskenkari P, et al. The validity of the Finnish Hospital Discharge Register and Causes of Death Register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil* 2005;12:132-7.
24 Merry AH, Boer JM, Schouten LJ, Feskens EJ, Verschuren WM, Gorgels AP, et al. Validity of coronary heart diseases and heart failure based on hospital discharge and mortality data in the Netherlands using the cardiovascular registry Maastricht cohort study. *Eur J Epidemiol* 2009;24:237-47.
25 Read Codes, NHS Connecting for Health, 2012. 2013. www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes.
26 Bhattarai N, Charlton J, Rudisill C, Gulliford MC. Coding, recording and incidence of different forms of coronary heart disease in primary care. *PLoS One* 2012;7:e29776. doi: 10.1371/journal.pone.0029776.
27 Delaney JAC, Daskalopoulou SS, Brophy JM, Steele RJ, Opatrny L, Suissa S. Lifestyle variables and the risk of myocardial infarction in the General Practice Research Database. *BMC Cardiovasc Disord* 2007;7:38. doi:10.1186/1471-2261-7-38.
28 World Health Organization. International statistical classification of diseases and health related problems: 10th revision, Volume 2. WHO, 2010.
29 Sekhri N, Feder GS, Junghans C, Hemingway H, Timmis AD. How effective are rapid access chest pain clinics? Prognosis of incident angina and non-cardiac chest pain in 8762 consecutive patients. *Heart* 2007;93:458-63.
30 Department of Health. Hospital Episode Statistics. 2013. www.hesonline.nhs.uk/.
31 Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, et al. Third universal definition of myocardial infarction. *Eur Heart J* 2012;33:2551-67.
32 McLennan D, Barnes H, Noble M, Davies J, Garratt E, Dibben C. The English Indices of Deprivation 2010. Department for Communities and Local Government, March 2011. 2013. www.communities.gov.uk/publications/corporate/statistics/indices2010technicalreport.
33 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373-83.
34 Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) (Adult Treatment Panel III). *JAMA* 2001;285:2486-97.
35 R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, 2012. ISBN 3-900051-07-0. www.R-project.org/.
36 Goldacre MJ, Roberts SE, Griffith M. Place, time and certified cause of death in people who die after hospital admission for myocardial infarction or stroke. *Eur J Public Health* 2004;14:338-42.
37 Payne RA, Abel GA, Simpson CR. A retrospective cohort study assessing patient characteristics and the incidence of cardiovascular disease using linked routine primary and secondary care data. *BMJ Open* 2012;2:e000723.
38 Boyle CA, Dobson AJ. The accuracy of hospital records and death certificates for acute myocardial infarction. *Aust NZ J Med* 1995;25:316-23.
39 Madsen M, Davidsen M, Rasmussen S, Abildstrom SZ, Osler M. The validity of the diagnosis of acute myocardial infarction in routine statistics: a comparison of mortality and hospital discharge data with the Danish MONICA registry. *J Clin Epidemiol* 2003;56:124-30.
40 Herrett EL, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010:69;4-14. doi:10.1111/j.1365-2125.2009.03537.x.
41 Garcia Rodriguez LA, Varas-Lorenzo C, Maguire A, Gonzalez-Perez A. Nonsteroidal antiinflammatory drugs and the risk of myocardial infarction in the general population. *Circulation* 2004;109:3000-6.
42 Van Staa T, Abenheim L. The quality of information recorded on a UK database of primary care records: a study of hospitalisations due to hypoglycemia and other conditions. *Pharmacoepidemiol Drug Saf* 1994;3:15-21.
43 Meier CR, Jick SS, Derby LE, Vasilakis C, Jick H. Acute respiratory-tract infections and risk of first-time acute myocardial infarction. *Lancet* 1998;351:1467-71.
44 Jick H, Jick S, Myers MW, Vasilakis C. Risk of acute myocardial infarction and low-dose combined oral contraceptives. *Lancet* 1996;347:627-8.
45 Hammad TA, McAdams MA, Feight A, Iyasu S, Dal Pan GJ. Determining the predictive value of Read/OXMIS codes to identify incident acute myocardial infarction in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2008;17:1197-201.
46 Hall GC, Brown MM, Mo J, MacRae KD. Triptans in migraine: the risks of stroke, cardiovascular disease, and death in practice. *Neurology* 2004;62:563-8.
47 Johansson S, Wallander MA, Ruigomez A, Garcia Rodriguez LA. Is there any association between myocardial infarction, gastro-oesophageal reflux disease and acid-suppressing drugs? *Aliment Pharmacol Ther* 2003;18:973-8.
48 Hammad TA, Graham DJ, Staffa JA, Kornegay CJ, Dal Pan GJ. Onset of acute myocardial infarction after use of non-steroidal anti-inflammatory drugs. *Pharmacoepidemiol Drug Saf* 2008;17:315-21.
49 Gulliford MC, Charlton J, Ashworth M, Rudd AG, Toschke AM; eCRT Research Team. Selection of medical diagnostic codes for analysis of electronic patient records. Application

to stroke in a primary care database. *PLoS One* 2009;4:e7168. doi:10.1371/journal.pone. 0007168.

50 Silver LE, Heneghan C, Mehta Z, Banning AP, Rothwell PM. Substantial underestimation of incidence of acute myocardial infarction by hospital discharge diagnostic coding data: a prospective population-based study. *Heart* 2009;95:A5.

51 UK Biobank Protocol, UK Biobank, 2007. 2013. www.ukbiobank.ac.uk/resources/.

52 Gallagher AM, Puri S, Van Staa T. Linkage of the General Practice Research Database (GPRD) with other data sources. *Pharmacoepidemiol Drug Saf* 2011;20:S230-1.

53 Herrett E. Myocardial infarction as the first manifestation of atherosclerotic disease. PhD dissertation. London School of Hygiene and Tropical Medicine, 2103.

54 Million Veteran Program. 2013. www.research.va.gov/mvp.

55 Kaiser Permanente Research Program on Genes, Environment, and Health. 2013. www. dor.kaiser.org/external/DORExternal/rpgeh.

56 Johannesdottir SA, Horváth-Puhó E, Ehrenstein V, Schmidt M, Pedersen L, Sørensen HT. Existing data sources for clinical epidemiology: the Danish National Database of Reimbursed Prescriptions. *Clin Epidemiol* 2012;4:303-13. doi:10.2147/CLEP.S37587.

57 Death Certification Reforms in England and Wales: update for coroners. Department of Health, September 2012. 2013. www.dh.gov.uk/health/2012/09/update-for-coroners/.

58 Van Staa TP, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, et al. Randomised Evaluations of Accepted Choices in Treatment (REACT) trials: large-scale pragmatic trials within databases of routinely collected electronic healthcare records. *BMJ* 2012;344:e55.

59 Boggon R, Van Staa T, Chapman M, Gallagher AM, Hammad TA, Richards AM. Cancer recording and mortality in the general practice research database and linked cancer registries. *Pharmacoepidemiol Drug Saf* 2013;22:168-75. doi:10.1002/pds.3374.

Tables

Table 1| Recording of risk factors in primary care before myocardial infarction recorded in primary care, hospital admission, disease registry, and death registry sources from 1 January 2003 to 31 March 2009

Characteristics	Primary care: CPRD	Hospital admissions: HES	Disease registry: MINAP	Cause specific mortality: ONS
No of patients	15 819	13 831	10 351	4017
Median (interquartile range) age (years)	73 (61-81)	73 (61-82)	72 (61-81)	81 (73-87)
Women	5810 (36.7)	5072 (36.7)	3649 (35.3)	1752 (43.6)
Most deprived fifth*	3211 (20.3)	2641 (19.1)	1997 (19.3)	849 (21.1)
Smoking:				
Current	4147 (26.2)	3608 (26.1)	2729 (26.4)	638 (15.9)
Former	9414 (59.5)	8176 (59.1)	6194 (59.8)	2622 (65.3)
None	1933 (12.2)	1745 (12.6)	1341 (13.0)	521 (13.0)
Missing	325 (2.1)	302 (2.2)	87 (0.8)	236 (5.9)
Mean (SD) systolic blood pressure (mm Hg)†	145 (15.4)	145 (15.6)	145 (15.2)	146 (16.1)
Missing	385 (2.4)	351 (2.5)	198 (1.9)	64 (1.6)
Use of blood pressure lowering drugs	9149 (57.8)	7907 (57.2)	5950 (57.5)	2919 (72.7)
Mean (SD) total serum cholesterol (mmol/L)†	5.4 (0.9)	5.4 (0.9)	5.4 (0.9)	5.2 (0.9)
Missing	4646 (29.3)	4291 (31.0)	2927 (28.3)	1101 (27.4)
Mean (SD) HDL cholesterol (mmol/L)†	1.3 (0.3)	1.3 (0.3)	1.3 (0.3)	1.4 (0.3)
Missing	6985 (44.2)	6214 (44.9)	4443 (42.9)	1703 (42.4)
Use of lipid lowering drugs	5632 (35.6)	4686 (33.9)	3669 (35.4)	1757 (43.7)
Framingham hard coronary disease risk score‡				
<10%	1273 (8.0)	1019 (7.4)	851 (8.2)	186 (4.6)
10-20%	4718 (29.8)	4121 (29.8)	3181 (30.7)	1248 (31.1)
>20%	2799 (17.7)	2439 (17.6)	1841 (17.8)	872 (21.7)
Missing	7029 (44.4)	6252 (45.2)	4478 (43.3)	1711 (42.6)
Diabetes	2885 (18.2)	2467 (17.8)	1858 (17.9)	927 (23.1)
Mean (SD) Charlson index	2.5 (1.7)	2.4 (1.6)	2.4 (1.6)	3.2 (1.9)
Median (interquartile range) primary care consultation rate per year	3.7 (1.6-7.9)	3.5 (1.5-7.7)	3.6 (1.6-7.8)	5.0 (2.3-9.8)

CPRD=Clinical Practice Research Datalink; HES=Hospital Episode Statistics; MINAP=Myocardial Ischaemia National Audit Project; ONS=Office for National Statistics; HDL=high density lipoprotein.

The total number of patients was 21 482. Patients might be represented in more than one column if their myocardial infarction was recorded in more than one source.

*Assessed by index of multiple deprivation.

†Mean of measurements before date of myocardial infarction.

‡Based on patients with complete data for blood pressure and cholesterol levels.

Table 2| Information recorded in disease registry (MINAP) within 30 days for non-fatal myocardial infarction recorded in primary care (Clinical Practice Research Datalink, CPRD) or hospital admissions (Hospital Episode Statistics, HES). Values are numbers (percentages) unless stated otherwise

| Information in MINAP	Source of myocardial infarction record		
	CPRD	HES	CPRD and HES
No of patients	7224	7489	6006
Electrocardiographic findings:			
ST elevation	3373 (46.7)	3455 (46.1)	3062 (51.0)
Other abnormality	2337 (32.4)	2485 (33.2)	1816 (30.2)
Normal	389 (5.4)	429 (5.7)	306 (5.1)
Not recorded	1125 (15.6)	1120 (15)	822 (13.7)
Cardiac markers:			
Raised	6149 (85.1)	6358 (84.9)	5121 (85.3)
Normal	368 (5.1)	411 (5.5)	299 (5.0)
Missing	707 (9.8)	720 (9.6)	586 (9.8)
Peak troponin:			
Level recorded	6109 (84.6)	6357 (84.9)	5029 (83.7)
Median (interquartile range) ng/ml	2.03 (0.47-10.0)	2.04 (0.45-10.2)	2.34 (0.53-11.6)
CALIBER diagnosis*:			
ST elevation myocardial infarction	3386 (46.9)	3441 (46.0)	3064 (51.0)
Non-ST elevation myocardial infarction	3274 (45.3)	3410 (45.5)	2497 (41.6)
Unstable angina	384 (5.3)	425 (5.7)	312 (5.2)
Other	180 (2.5)	213 (2.8)	133 (2.2)
Positive predictive value (95% CI) for myocardial infarction†	92.2 (91.6 to 92.8)	91.5 (90.8 to 92.1)	92.6 (91.9 to 93.3)

MINAP=Myocardial Ischaemia National Audit Project.

*MINAP contains details of admissions with suspected acute coronary syndromes. The CALIBER algorithm assigns a diagnosis based on troponin, ECG findings, and discharge diagnosis recorded in MINAP.

†Positive predictive value of HES or CPRD myocardial infarction is calculated considering a MINAP diagnosis of myocardial infarction as gold standard.

Figures
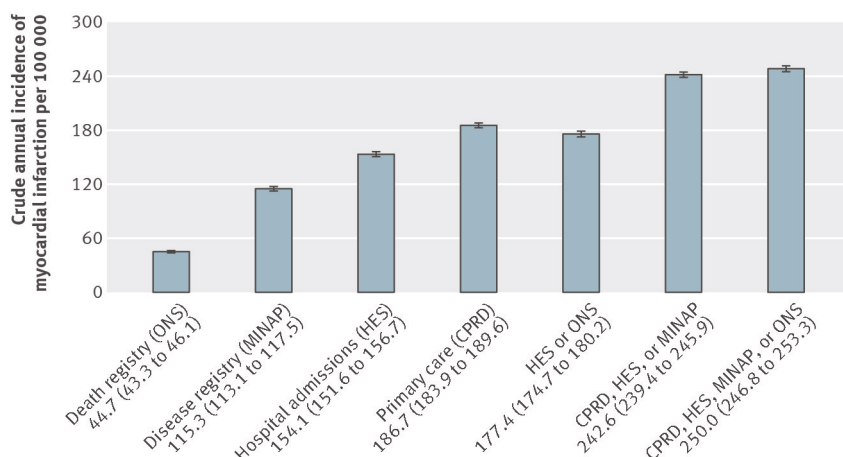
Fig 1 Crude incidence of acute fatal and non-fatal myocardial infarction estimated using different combinations of data from primary care (Clinical Practice Research Datalink), hospital admissions (Hospital Episode Statistics), disease registry (MINAP, Myocardial Ischaemia National Audit Project), and death registry (Office for National Statistics)
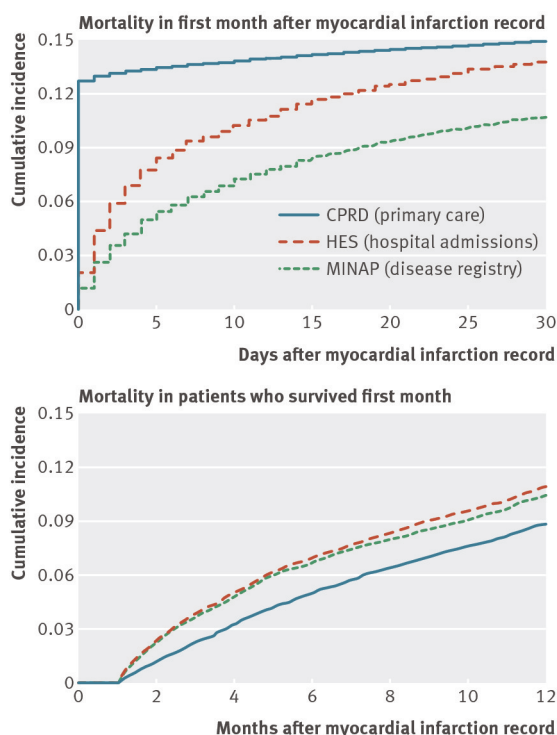
Fig 2 Kaplan Meier curves showing all cause mortality, stratified by record source in 20 819 patients: Clinical Practice Research Datalink (n=15 819), Hospital Episode Statistics (n=13 831), Myocardial Ischaemia National Audit Project (MINAP) (n=10 351). Myocardial infarctions recorded by the Office for National Statistics are not shown as they are by definition fatal on the date of myocardial infarction
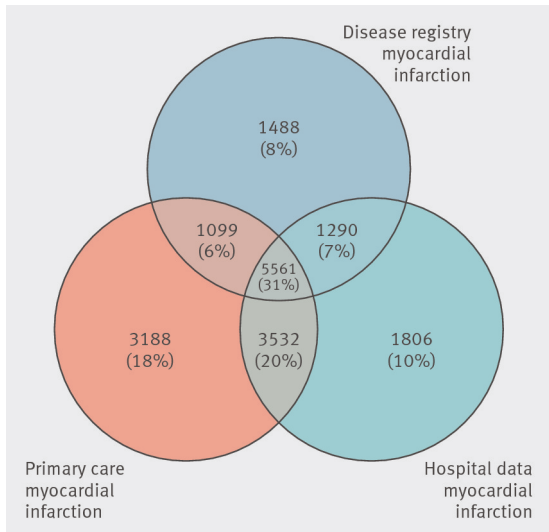
Fig 3 Number and percentage of records recorded in primary care (Clinical Practice Research Datalink), hospital care (Hospital Episode Statistics), and disease registry (Myocardial Ischaemia National Audit Project) for non-fatal myocardial infarction across the three sources (n=17 964 patients)