Validation of diagnosis of acute myocardial infarction and stroke in electronic medical records: a primary care cross-sectional study in Madrid, Spain (the e-MADVEVA Study)

Carmen de Burgos-Lunar,1,2 Isabel del Cura-Gonzalez3,2,3 Juan Cárdenas-Valladolid,3,4 Paloma Gómez-Campelo,5 Juan C Abánades-Herranz,6 Ana Lopez-de-Andres,7 Mercedes Sotos-Prieto,8 Victor Iriarte-Campo,9 Carmen Y Fuentes-Rodriguez,1 Rafael Gómez-Coronado,1 Miguel A Salinero-Fort6,9

ABSTRACT

Objectives To validate the diagnoses of acute myocardial infarction (AMI) and stroke recorded in electronic medical records (EMR) and to estimate the population prevalence of both diseases in people aged ≥18 years.

Design Cross-sectional validation study.

Setting 45 primary care centres.

Participants Simple random sampling of diagnoses of AMI and stroke (International Classification of Primary Care-2 codes K75 and K90, respectively) registered by 55 physicians and random age-matched and sex-matched sampling of the records that included in primary care EMRs in Madrid (Spain).

Primary and secondary outcome measures Sensitivity, specificity, positive and negative predictive values and overall agreement were calculated using the kappa statistic. Applied gold standards were ECGs, brain imaging studies, hospital discharge reports, cardiology reports and neurology reports. In the case of AMI, the ESC/ACCF/AHA/WHF Expert Consensus Document was also used. Secondary outcomes were the estimated prevalence of both diseases considering the sensitivity and specificity obtained (true prevalence).

Results The sensitivity of a diagnosis of AMI was 98.11% (95% CI, 96.29 to 99.03), and the specificity was 97.42% (95% CI, 95.44 to 98.55). The sensitivity of a diagnosis of stroke was 97.56% (95% CI, 95.56 to 98.68), and the specificity was 94.51% (95% CI, 91.96 to 96.28). No differences in the results were found after stratification by age and sex (both diseases). The prevalence of AMI and stroke was 1.38% and 1.27%, respectively.

Conclusion The validation results show that diagnoses of AMI and stroke in primary care EMRs constitute a helpful tool in epidemiological studies. The prevalence of AMI and stroke was lower than 2% in the population aged over 18 years.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The major strength of the e-MADVEVA Study is the individual validation (manual validation) of electronic medical records (EMRs) by comparing each case and non-case, matched by age, with an accepted reference gold standard.

⇒ The validation method allows for calculating positive predictive value, negative predictive value, sensitivity and specificity, in contrast with other methods such as questionnaires for healthcare practitioners or patients and comparing rates in a comparable population.

⇒ In-hospital mortality due to acute myocardial infarction (AMI) and stroke may not have been recorded in primary care EMRs, with the result that the prevalence of patients who die in hospital due to a first AMI or stroke may be underestimated.

⇒ The International Classification of Primary Care (ICPC-2) codes studied for AMI (ICPC-2 K75) and stroke (ICPC-2 K90) do not allow differentiation between the different types of AMI and stroke.

INTRODUCTION

Electronic medical records (EMR) are digital versions of paper charts in primary care settings and hospitals. EMRs contain notes and information collected by and for clinicians and are used mostly by care providers for diagnosis and treatment. EMRs are more valuable than paper records because they enable providers to track data over time, identify patients for preventive care and screening, monitor patients and improve the quality of health care.1 In the last 20 years, EMRs have become an increasingly common source of information for research. They
allow for more efficient analyses and generalisation of findings and minimise selection and recall bias.

Despite these advantages, the validity of some diagnoses recorded in EMRs is uncertain, as the records were not created specifically for research purposes. Errors and inconsistencies in diagnoses can lead to misclassification bias, affecting the quality of research. The EMR must be able to distinguish between those who have had a disease (according to an accepted ‘gold standard’ reference diagnosis) and those who have not.

Healthcare in Spain is publicly funded and universal. Primary care is usually the gateway to the health system and the point to which most patients treated at other levels of care return. In the Community of Madrid, all primary care centres have had EMRs for more than 20 years, and hospital care reports can be accessed. These characteristics make EMRs a valuable source of epidemiological information.

Validation studies determine the degree of systematic measurement error and aid in the interpretation of results. Researches increase the reliability of their findings by quantifying the accuracy of data in EMRs. In some database validation studies, diagnoses are validated using a direct method (individual validation), which is more laborious than indirect methods that compare results (eg, prevalence) with national health surveys or other epidemiological studies. In addition, the direct or individual method checks for correct classification against a gold standard on a patient-by-patient basis and requires specific resources for each diagnosis, so that full validation of the database is sequential in time, as has been observed in other databases such as THIN UK. Previously, our group used the direct method to validate codes for arterial hypertension and diabetes mellitus in primary care EMRs and, more recently, those for atrial fibrillation. In both cases, we obtained high sensitivities, specificities, predictive values and diagnostic agreement.

Few studies have validated acute myocardial infarction (AMI) and stroke in primary care EMRs.

We chose the diagnoses AMI and stroke because they are major causes of morbidity and mortality and have internationally accepted diagnostic criteria.

This MADrid electronic medical records for Vascular Events Validation (e-MADVEVA Study) aims to validate the diagnoses of AMI and stroke recorded in EMR and to estimate the population prevalence of both diseases in people aged ≥18 years in the Community of Madrid (Spain).

**METHODS**

**Design**

Cross-sectional validation study of diagnoses of AMI and stroke recorded in primary care EMRs.

**Setting**

The e-MADVEVA Study was carried out in 45 primary care centres in the Community of Madrid. These centres provide care to a population of 1080000 people. Fifty-five volunteer general practitioners (GPs) took part in the study.

**Sources of information**

The e-MADVEVA Study was based on individualised patient data obtained from patients’ primary care EMRs. The EMR, which is managed by the AP-Madrid computer application, is structured around a list of episodes consisting of a code and a description or label. The code corresponds to the second edition of the International Classification of Primary Care (ICPC-2) and can have several different descriptions. An episode of care is defined as a health problem or disease from its first presentation to a healthcare provider to the completion of the last encounter for that same health problem or disease. After several visits by the patient for the same health problem, it is sometimes necessary to change the diagnosis. Therefore, the diagnostic code should be replaced by the definitive one. Unfortunately, the AP-Madrid application allows the ICPC-2 label to be modified without changing the code, thus generating a diagnostic classification error.

**Study population**

The e-MADVEVA Study population comprised persons aged ≥18 years with active primary care EMRs and at least one entry in the EMR before 1 January 2015.

Patients were not included if they were not regular users, if they were temporarily displaced from their usual residence or if they were transients who required occasional urgent medical care at the primary care centres.

**Sampling**

The sampling procedure has been described elsewhere. Briefly, four samples were obtained, two to validate the diagnosis of AMI and two to validate the diagnosis of stroke. Samples 1 and 3 were obtained from clinical records with codes ICPC-2 K-75 (myocardial infarction) and K-90 (stroke), respectively. Samples 2 and 4 were control samples paired by age and sex, without codes K-75 and K-90, respectively.

Given the absence of reference data on the expected proportion of misclassifications (FNs or FPs), maximum possible indeterminacy (p=q=0.5) was assumed.

With this assumption, and for a confidence level of 95% and a precision of 5%, 384 patients needed to be assessed for each variable. This was increased to 423 in anticipation of a 10% loss from sampling to validation of the diagnoses (EMR of patients who had become inactive owing to a change of residence, death, or other reasons).

Four patient samples were obtained as follows:

To validate AMI episodes:

- Sample 1: 423 patients with an AMI code (ICPC-2 K75).
- Sample 2: 423 patients without AMI codes.

To validate stroke episodes:
Sample 3: 423 patients with a stroke code (ICPC-2 K90).
Sample 4: 423 patients without stroke codes.

Given that the probability of presenting these episodes increases with age and that they are not equally prevalent in both sexes, it was decided that samples 2 and 4 (patients without episodes) should be similar to the respective samples 1 and 3 (patients with episodes) in the variables year of birth and sex. This strategy avoids overestimating specificity if the sample comprised younger patients or was characterised by the predominance of one sex over the other.

Samples 1 and 3 were obtained by simple random sampling from the patient list of participating GPs. Samples 2 and 4 were obtained by individual age and sex matching techniques with their corresponding samples 1 and 3.

Our group collaborates with 153 GPs. Of these, 55 volunteers were selected by random sampling, and another four of the 153 GP’s participated in the validation process.

A diagnostic test is validated by comparing its positive and negative results with those obtained by the best available instrument for measuring the phenomenon under study (gold standard). In this study, the diagnoses of AMI and stroke in primary care EMRs could be considered equivalent to diagnostic tests. However, our main purpose is not to see how EMR can diagnose but how it can correctly classify the sick as sick (sensitivity) and the healthy as healthy (specificity). We also look at the accuracy of EMR compared with the gold standard.

Patients in samples 1 and 2 were considered to have AMI if they met any of the following criteria:
- Clinical record of hospital discharge report or cardiology outpatient report with a diagnosis of AMI.
- Meeting the diagnostic criteria set out in the third universal definition of AMI established in the ESC/ACCF/AHA/WHF Expert Consensus Document.

Patients in samples 3 and 4 were considered to have had a stroke if they met any of the following criteria:
- Hospital or neurological outpatient discharge report with a diagnosis of stroke.
- Sudden onset of a focal neurological deficit, with clinical or imaging evidence of infarction lasting 24 hours or more and not attributable to a non-ischaemic cause (ie, not associated with brain infection, trauma, tumour, seizure, metabolic disease or degenerative neurological disease).
- Acute extravasation of blood into the brain parenchyma or subarachnoid space associated with neurological symptoms.

The evaluators validated the diagnoses by accessing the primary care EMR and, based on the information collected there, verifying that the criteria were met.

The flow chart of EMR and patients’ selection is shown in figure 1. The evaluators were eight GPs with experience in the management of the AP-Madrid application. The assessment was peer-reviewed, and discrepancies were resolved by consensus.

Statistical analysis
First, a descriptive analysis of the study populations and samples was performed. The quantitative variables are expressed as the mean and SD, and age is expressed as the median and IQR. Qualitative variables were summarised with their relative frequency.

Second, we used a 2×2 table between the gold standard (yes/no AMI and stroke) and the EMRs (yes/no AMI and stroke) to calculate true positive (TP) and true negative (TN) and false positive (FP) and false negative (FN). TP is the number of EMRs classified as true when they were true according to the gold standard. FN is the number of EMRs classified as false when they were true according to the gold standard. FP is the number of EMRs classified as true when they were false according to the gold standard. TN is the number of EMRs classified as false when they were false according to the gold standard.

The metrics obtained from the 2×2 table are as follows:
- Accuracy: (TP+TN)/(TP+FP+FN+TN). Generally, the metric accuracy measures the ratio of correct classifications over total number of cases evaluated.
- Sensitivity: TP/(TP+FN). It is the proportion of cases with AMI or stroke codes in the EMR among all cases where the diagnostic criteria could be verified.
- Specificity: TN/(TN+FP). It is the proportion of cases with no AMI or stroke code among those who did not meet the diagnostic criteria.
- Positive predictive value (PPV) or precision (P): TP/(TP+FP). Measures the positive patterns that are correctly classified from the total predicted patterns in a positive class (AMI and stroke). The PPV can be calculated for any prevalence as follows:
  \[
  \text{PPV} = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1-\text{specificity}) \times (1-\text{prevalence})}
  \]

Negative predictive value (NPV): TN/(TN+FN): measures the TN patterns that are correctly classified from the total predictive patterns in a negative class (healthy people).

The NPV can be calculated for any prevalence as follows:
  \[
  \text{NPV} = \frac{\text{specificity} \times (1-\text{prevalence})}{(1-\text{sensitivity}) \times \text{prevalence} + \text{specificity} \times (1-\text{prevalence})}
  \]

If the prevalence of the disease is very high, the PPV will be high if both the sensitivity and specificity are fixed. As our data have a prevalence of 50%, we have adjusted the predictive values to the prevalence in EMRs using the MedCalc programme.

Recall (R): TP/(TP+FN). The recall is used to measure the fraction of positive patterns that are correctly classified.

F1-Score: (2*precision*recall)/(precision+recall). This metric represents the harmonic mean between recall and precision values (PPV).

Likelihood ratio (LR): It is defined as the ratio of the expected test results in subjects with a particular condition/disease to those without the condition. As such, the LR directly relates a given patient’s pretest and post-test
probability of having the disease. In simple terms, LR tells us how much more likely a particular positive test result or positive diagnostic classification is in subjects with the disease than in those without the disease. If both probabilities are equal, then that test or diagnostic classification has no value and its LR=1. The likelihood ratio for positive test or positive diagnostic classification (LR+) tells us how likely the positive test or positive diagnostic classification result is to occur in subjects with the disease compared with those without the disease. LR+ can be calculated using the following formula:

\[ LR^+ = \frac{sensitivity}{1 - specificity} = \frac{TP}{FP} \]

Good diagnostic tests or diagnostic classifications have LR+>10, and their positive result significantly contributes to the diagnosis. LR− can be calculated using the following formula:

\[ LR^- = \frac{1 - sensitivity}{specificity} = \frac{FN}{TN} \]

LR− is a good indicator for ruling out the diagnosis. Good diagnostic tests or diagnostics classification have LR−<0.1.

Area under the ROC curve (AUC): \( \frac{Sp–np (nn+1)/2}{np nn} \). The AUC value reflects the overall ranking performance of a classifier, where Sp is the sum of all positive examples ranked. At the same time, np and nn denote the number of positive or negative examples, respectively.

Accuracy, sensitivity and specificity, PPV and NPV, positive likelihood ratio (LR+) and negative likelihood ratio (LR−) and AUC with their 95% CI were calculated overall and stratified by sex and age groups. We tested whether the sensitivity and specificity differed according to the different categories of the variables using the \( \chi^2 \) test for homogeneity of the contrast statistic. When the conditions for its application were not met (any expected frequency less than 5), a two-sided Fisher’s exact test was used.

The proportion of individuals with a disease code in the EMR (apparent prevalence) should not be used as an estimate of the prevalence of a disease in that population, given that the sensitivity and specificity of these diagnoses are usually less than 100%. Thus, the proportion of individuals with a positive result includes FPs and excludes FNs. Consequently, estimating the true prevalence of a disease requires an adjustment for misclassification.
resulting from the sensitivity and specificity. In this study, the formula proposed by Rogan and Gladen\(^\text{16}\) was used for this adjustment, as follows: true prevalence=\((\text{apparent prevalence} \times \text{specificity}−1)/(\text{sensitivity} + \text{specificity}−1)\).

The degree of overall agreement between the recorded diagnosis and the reference standard, as well as the interobserver agreement, was determined using the kappa index and its CIs. According to this value, agreement is considered poor (≤0.20), low (0.21–0.40), moderate (0.41–0.60), good (0.61–0.80) or very good (≥0.81).\(^\text{17}\)

The statistical analysis was performed using MedCalc, statistical software V. 20.211 (Ostend, Belgium), SPSS V.19.0 (IBM Corp) and the CIs of the kappa index and the predictive values were calculated with the macros for SPSS of the Laboratory of Applied Statistics of the Autonomous University of Barcelona (Spain), !KAPPA and !DT, respectively.\(^\text{18,19}\)

### Confidentiality of the data
To guarantee the confidentiality of the data in the validation process, we acted under the provisions of the Spanish ‘Organic Law on Personal Data Protection and guarantee of digital rights’ and the provisions of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on Data Protection.

### RESULTS
The main demographic characteristics of the population of patients over 18 years of age seen in primary care centres of the Community of Madrid with episodes of AMI (ICPC-2 K75) and stroke (ICPC-2 K90) in their EMR, as well as those of the samples selected, are described in table 1.

Of the 5 040 988 patients aged≥18 years with an active medical history and regular users of primary care centres, 47.36% were male, with a mean age of 49.82 (SD 17.74).

An AMI code was identified in the EMR in 1.34% of the population; of these, 76.8% were male, with a mean age of 70.45 (SD 12.84). On the other hand, patients with an AMI code in sample 1 had a mean age of 72.70 (SD 12.3) years, and 71.63% were male.

In the same population, a diagnosis of stroke was found in the EMR in 1.22%; of these, 52.81% were male, with

### Table 1 Main demographic characteristics of the patients aged ≥18 years attended in PHC centres and of samples selected

<table>
<thead>
<tr>
<th></th>
<th>N (%)</th>
<th>Age (years) median (IQR)</th>
<th>Age≥70 years (%)</th>
<th>Female sex (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with diagnostic code for AMI (K-75)</td>
<td>67516 (1.34)</td>
<td>71.04 (61.31–80.54)</td>
<td>52.78</td>
<td>23.20</td>
</tr>
<tr>
<td>Sample of patients with diagnostic code for AMI (sample 1)</td>
<td>423 (0.63)</td>
<td>73.07 (64.02–83.12)</td>
<td>59.34</td>
<td>28.37</td>
</tr>
<tr>
<td>Correct diagnosis (TP)</td>
<td>412 (97.40)</td>
<td>72.81 (63.93–82.99)</td>
<td>58.74</td>
<td>27.43</td>
</tr>
<tr>
<td>Incorrect diagnosis (FP)</td>
<td>11 (2.60)</td>
<td>85.25 (75.85–90.14)</td>
<td>81.82</td>
<td>63.64</td>
</tr>
<tr>
<td>Patients without diagnostic code for AMI (K-75)</td>
<td>4973472 (98.66)</td>
<td>47.40 (36.60–61.48)</td>
<td>14.81</td>
<td>53.04</td>
</tr>
<tr>
<td>Sample of patients without diagnostic code for AMI (sample 2)</td>
<td>423 (0.01)</td>
<td>73.28 (63.88–82.80)</td>
<td>59.10</td>
<td>28.37</td>
</tr>
<tr>
<td>Correct diagnosis (TN)</td>
<td>415 (98.11)</td>
<td>73.08 (63.76–82.64)</td>
<td>58.80</td>
<td>28.67</td>
</tr>
<tr>
<td>Incorrect diagnosis (FN)</td>
<td>8 (1.89)</td>
<td>84.70 (69.89–86.13)</td>
<td>75.00</td>
<td>12.50</td>
</tr>
<tr>
<td>Patients with diagnostic code for stroke (K-90)</td>
<td>61718 (1.22)</td>
<td>74.99 (63.99–83.81)</td>
<td>62.74</td>
<td>47.19</td>
</tr>
<tr>
<td>Sample of patients with diagnostic code for stroke (sample 3)</td>
<td>423 (0.69)</td>
<td>75.24 (64.00–83.83)</td>
<td>64.07</td>
<td>46.57</td>
</tr>
<tr>
<td>Correct diagnosis (TP)</td>
<td>399 (94.33)</td>
<td>75.66 (64.00–83.83)</td>
<td>64.66</td>
<td>45.86</td>
</tr>
<tr>
<td>Incorrect diagnosis (FP)</td>
<td>24 (5.67)</td>
<td>72.68 (62.12–83.99)</td>
<td>54.17</td>
<td>58.33</td>
</tr>
<tr>
<td>Patients without diagnostic code for stroke (K-90)</td>
<td>4979270 (98.78)</td>
<td>47.44 (36.62–61.48)</td>
<td>14.73</td>
<td>52.71</td>
</tr>
<tr>
<td>Sample of patients with diagnostic code for stroke (sample 4)</td>
<td>423 (0.01)</td>
<td>75.30 (63.88–84.07)</td>
<td>63.83</td>
<td>46.57</td>
</tr>
<tr>
<td>Correct diagnosis (TN)</td>
<td>413 (97.64)</td>
<td>75.27 (63.76–84.02)</td>
<td>63.20</td>
<td>46.73</td>
</tr>
<tr>
<td>Incorrect diagnosis (FN)</td>
<td>10 (2.36)</td>
<td>78.84 (71.98–85.16)</td>
<td>90.00</td>
<td>40.00</td>
</tr>
</tbody>
</table>

AMI, acute myocardial infarction; FN, false negative; FP, false positive; PHC, primary healthcare; TN, true negative; TP, true positive.
The overall agreement between the diagnosis recorded in the EMR and the reference standard, measured as the kappa concordance index, was very good for diagnosing AMI ($\kappa=0.955$) and stroke ($\kappa=0.920$).

The sensitivity of a diagnosis of stroke was 97.56%, and according to the reference standard (true prevalence), both overall and stratified by age group and sex.

The misclassification cases (FPs and FNs) were few, increasing specificity from 97.42% to 99.96% for AMI and from 94.215% to 99.93% for stroke.

The different strata of the variables sex and age over 69 years. In all cases, $\kappa$ indices above 0.880 were achieved.

As shown in table 2, the diagnosis of AMI was confirmed in 98.11% of cases (sensitivity), and for no stroke and stroke codes (0.6854%). This increased progressively with age, reaching 5.5% in those over 80 years (3.23% in women and 9.77% in men).

The true prevalence of stroke was 1.27% (1.13% in women and 1.42% in men); in the group aged over 80 years, the true prevalence of stroke was 7.35% (3.23% in women and 9.77% in men).

Table 3 shows the differences in the prevalence of AMI and stroke, as recorded in the EMR (apparent prevalence) and according to the reference standard (true prevalence), both overall and stratified by age group and sex.

Another approach would have been to achieve the same sample fraction for non-AMI and AMI codes in the EMR (0.6265%) and for no stroke and stroke codes (0.6854%). To do this, we would have had to include 31,158 individuals without AMI codes and 34,130 individuals without stroke codes. Given our limited resources, this size would have made the study unfeasible. However, we have included a simulation example (see online supplemental tables 4 and 5) holding TP, FP and sensitivity (96.29% to 99.03) constant. The results suggest that we would have slightly increased specificity from 97.42% to 99.96% for AMI and from 94.215% to 99.93% for stroke.
Table 3  Apparent prevalence (diagnosis in the EMR) and true prevalence (fulfilment of the diagnostic criteria) of AMI and stroke in persons aged≥18 years in primary healthcare centres of the Community of Madrid

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Apparent prevalence (95% CI)</td>
<td>True prevalence (95% CI)</td>
<td>Apparent prevalence (95% CI)</td>
</tr>
<tr>
<td>AMI</td>
<td>Total</td>
<td>1.34 (1.33 to 1.35)</td>
<td>1.38 (1.37 to 1.39)</td>
</tr>
<tr>
<td></td>
<td>18–50 y</td>
<td>0.15 (0.14 to 0.15)</td>
<td>0.14 (0.13 to 0.14)</td>
</tr>
<tr>
<td></td>
<td>51–60 y</td>
<td>1.24 (1.22 to 1.27)</td>
<td>1.28 (1.25 to 1.3)</td>
</tr>
<tr>
<td></td>
<td>61–70 y</td>
<td>2.7 (2.66 to 27.37)</td>
<td>2.78 (2.74 to 2.83)</td>
</tr>
<tr>
<td></td>
<td>71–80 y</td>
<td>4.07 (4.02 to 4.13)</td>
<td>4.21 (4.15 to 4.27)</td>
</tr>
<tr>
<td></td>
<td>&gt;80 y</td>
<td>5.31 (5.24 to 5.39)</td>
<td>5.5 (5.42 to 5.58)</td>
</tr>
<tr>
<td>Stroke</td>
<td>Total</td>
<td>1.22 (1.21 to 1.23)</td>
<td>1.27 (1.26 to 1.28)</td>
</tr>
<tr>
<td></td>
<td>18–50 y</td>
<td>0.18 (0.17 to 0.18)</td>
<td>0.13 (0.13 to 0.14)</td>
</tr>
<tr>
<td></td>
<td>51–60 y</td>
<td>0.77 (0.75 to 0.79)</td>
<td>0.77 (0.76 to 0.79)</td>
</tr>
<tr>
<td></td>
<td>61–70 y</td>
<td>1.81 (1.78 to 1.84)</td>
<td>1.9 (1.87 to 1.94)</td>
</tr>
<tr>
<td></td>
<td>71–80 y</td>
<td>3.62 (3.56 to 3.68)</td>
<td>3.87 (3.82 to 3.93)</td>
</tr>
<tr>
<td></td>
<td>&gt;80 y</td>
<td>6.82 (6.74 to 6.91)</td>
<td>7.35 (7.26 to 7.44)</td>
</tr>
</tbody>
</table>

True prevalence: (apparent prevalence+specificity−1)/(sensitivity+specificity−1). Apparent prevalence: proportion of individuals with a disease code in the EMR.

AMl, acute myocardial infarction; EMR, electronic medical record; y, years.

DISCUSSION

The results of the e-MADVEVA Study showed very good agreement with the reference standard and high sensitivity and specificity overall and in each sex category and age group.

These results are consistent with those of other published studies, although most validate the coding of these diagnoses in hospital registries or specific registries using ICD-9 or ICD-10 codes.

Thus, for the diagnosis of AMI, the systematic review by McCormick et al of 30 studies published between 1984 and 2010 found that sensitivity was, in most cases, higher than 86%, specificity higher than 89%, PPV higher than 93% and NPV higher than 75%. In their review of 31 studies published between 2000 and 2014, Rubbo et al found PPVs greater than 70%.

In the case of stroke, our results are superior to those found in other studies, such as that of Baldereschi et al in Italy (sensitivity 70.54%, PPV 97.3%), Hall et al in Canada (sensitivity 82.2%, PPV 68.8%), Johnsen et al in Denmark (PPV 87.6%) and Porter et al in Canada (sensitivity 97.3%). In a systematic review of 77 studies published between 1976 and 2015 by McCormick et al, the sensitivity of ICD-9 and ICD-10 codes for any cerebrovascular disease was over 82% in most studies, and the PPV was over 81%.

The specificity found in our study was 94.51%, and the NPV was 97.64%, similar to those found by Baldereschi et al (sensitivity 70.54%, PPV 97.3%). In a systematic review of 20 of 30 studies published between 1984 and 2014, Johnsen et al found PPVs greater than 81%.

Porter et al published studies, although most validate the coding of these diagnoses in hospital registries or specific registries using ICD-9 or ICD-10 codes.

Thus, for the diagnosis of AMI, the systematic review by McCormick et al of 30 studies published between 1984 and 2010 found that sensitivity was, in most cases, higher than 86%, specificity higher than 89%, PPV higher than 93% and NPV higher than 75%. In their review of 31 studies published between 2000 and 2014, Rubbo et al found PPVs greater than 70%.

In the case of stroke, our results are superior to those found in other studies, such as that of Baldereschi et al in Italy (sensitivity 70.54%, PPV 97.3%), Hall et al in Canada (sensitivity 82.2%, PPV 68.8%), Johnsen et al in Denmark (PPV 87.6%) and Porter et al in Canada (sensitivity 97.3%). In a systematic review of 77 studies published between 1976 and 2015 by McCormick et al, the sensitivity of ICD-9 and ICD-10 codes for any cerebrovascular disease was over 82% in most studies, and the PPV was over 81%.

The specificity found in our study was 94.51%, and the NPV was 97.64%, similar to those found by Baldereschi et al (95.56% and 85.64%, respectively) and by McCormick et al (both above 95%).

Most epidemiological studies on AMI and stroke in Spain estimate incidence, hospital admissions and mortality. The few prevalence studies found show great variability with respect to terminology, definition and methodology, as well as in the age groups assessed.
The adult questionnaire of the 2017 National Health Survey of the Spanish National Institute of Statistics includes the question ‘Has a doctor ever diagnosed you with an acute myocardial infarction?’ and ‘Has a doctor ever diagnosed you with stroke (embolism, cerebral infarction, cerebral haemorrhage)?’. Figure 2 compares the prevalence of AMI and stroke found in this study with the rates of positive responses to these questions by sex and age group. As can be seen, the results are very similar.

Our study found a few FPs and FNs. FPs are usually due to a presumed diagnosis that has not subsequently been confirmed later, forgetting to replace the initial diagnosis with the definitive one. For example, at first attention, it may have been registered as AMI and later not subsequently modified to unstable angina.

FNs may occur due to limitations of the registration system when using the ICPC-2. Both symptoms and established diseases are used as diagnostic codes in this classification. Therefore, at first attention, it may have been registered as AMI and later not subsequently modified to unstable angina.

The misclassification could be favoured by the characteristics of saturation of primary care clinical practice in Spain, where the patients visit the GP, on average, 7.25 times in 2017 (https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20191219-1?ininheritRedirect=true&redirect=%2FSpanish%2Fen%2Fhome). Each GP has assigned approximately 1500 patients; subsequently, the number of visits per year is 11,250 (43 visits per day), being too high to ensure an excellent quality record.

In addition, changes made by clinics in episode labelling have been detected primarily in older diagnoses, suggesting that GPs tend to make fewer coding errors over time.

On the other hand, our results can be applied in some real-life situations. For example, GPs often have to rule out AMI in patients with typical chest pain and a low to intermediate pretest probability of disease. In primary care, the proportion of all consults for acute chest pain is 3%,46 and the pretest probability of ischaemic heart disease is 8%.

The use of an ECG allows us to increase the pretest probability to a post-test probability high enough to make a clinical decision. However, the ECG is not always available (outside normal working hours, being used on another patient at the same time). Therefore, increasing the probability of AMI without a diagnostic test is helpful in this situation. In this hypothetic case, using our results, which show a likelihood ratio of 73.28 for identifying male patients with AMI, and if the patient had a history of previous AMI in the EMR, the pretest probability would increase to 85% using Fagan nomogram,38 which would allow a decision to be administer antithrombotic drugs39 and then transport the patient to hospital by ambulance, as soon as possible.

Since specificity and sensitivity area used to calculate the likelihood ratio, neither LR+ nor LR− depend on the disease prevalence. Consequently, one study’s likelihood ratios apply to other clinical settings.
The major strength of the e-MADVEVA Study is the individual validation (manual validation) of EMR by comparing each case and non-case, matched by age, with an accepted reference gold standard. The validation method allows for calculating PPV, NPP, Sensitivity and Specificity, in contrast with other methods such as questionnaires for healthcare practitioners or patients and comparing rates in a comparable population.40

Our study is limited by the fact that the study design makes it impossible for researchers to remain blind to the medical records. Blinding could have led to more remarkable agreement than would have been the case if they had been blind to this information. In addition, in-hospital mortality due to AMI and stroke may not have been recorded in primary care EMRs, with the result that the prevalence of patients who die in hospital due to a first AMI or stroke may be underestimated.

On the other hand, our study is influenced by using the same sample size for cases and non-cases of both diseases, so the 2×2 table shows a prevalence of 50% for each disease. Using Bayes’ theorem, the prevalence affects the calculation of the predictive values. Therefore, we had to recalculate the predictive values for the prevalence found. However, the predictive values are not the primary indicator of our study, and the sensitivity and specificity are unaffected. An alternative would have been to use the same sample fraction for cases and non-cases recorded in the EMR, which would have required 31,000 non-case records. As mentioned above, this option was unfeasible for us. Furthermore, to the best of our knowledge, no validations require more than 30,000 records, except for algorithmic artificial intelligence methods.

The gold standard of our study was the fulfillment of diagnostic criteria that can be verified with information collected in the primary care EMR. For this reason, we cannot assume that FPs are diagnostic errors, only that they could not be verified, thus potentially leading us to underestimate prevalence.

We selected episodes using codes, regardless of the labels. This entails a potential selection bias in cases where the professional recording the episode modifies the label. However, only 0.26% and 0.74% of the AMI and stroke codes, respectively, had a label that did not correspond to the episode. Conversely, 0.07% and 0.11% of the labels of patients without cardiovascular episode codes were modified and corresponded to AMI and stroke, respectively.

Finally, the ICPC-2 codes studied for AMI (ICPC-2 K75) and stroke (ICPC-2 K90) do not allow differentiation between the different types of AMI and stroke and, therefore, cannot be used in studies requiring discrimination between types.

CONCLUSIONS

The validation results show that diagnoses of AMI and stroke recorded in primary care EMRs are a helpful tool for epidemiological studies.

The prevalence of AMI and stroke was lower than 2% in the population aged over 18 years. However, the prevalence of both diseases increases progressively with age.

Author affiliations
1Departamento Medicina Preventiva, Hospital Clínico San Carlos, Madrid, Comunidad de Madrid, Spain
2Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPP), Instituto de Salud Carlos III, Madrid, Comunidad de Madrid, Spain
3Research Support Unit, Primary Health Care Management, Madrid, Madrid, Spain
4Facultad de Ciencias de la Salud, Universidad Alfonso X el Sabio Facultad de Ciencias de la Salud, Villanueva de la Canada, Madrid, Spain
5Institute for Health Research (IdiPAZ), Hospital La Paz, Madrid, Madrid, Spain
6Monovar Health Centre, Servicio Madrileño de Salud, Madrid, Madrid, Spain
7Department of Public Health and Maternal and Child, Complutense University of Madrid Faculty of Medicine, Madrid, Comunidad de Madrid, Spain
8Department of Preventive Medicine, Autónoma University, Madrid, Spain
9MADiabetes Research Group, Foundation for Biomedical Research and Innovation in Primary Care of the Community of Madrid, Madrid, Spain

Acknowledgements We thank Thomas O’Boyle for writing assistance.

Contributors CBL conceived the study, and MASF, ICG, MSP and CBL developed its design. CBL, MASF, JAH and JCV contributed to data acquisition. JCV, VIC and PGC coordinated the research group. CBL, RGC, MSP, VIC and ALA performed the statistical analysis. All authors contributed to the interpretation of results. CBL, CYFR and RGC worked with MASF to develop the first draft of the manuscript. All authors contributed to revisions of the manuscript and the final content. All authors have approved the final manuscript, take responsibility for parts of the content and have agreed to be accountable for all aspects of the work. CBL and MASF were responsible for the concept of validation procedures. MASF is the guarantor of the manuscript.

Funding This work was supported by the FIS (Fondo de Investigaciones Sanitarias—Health Research Fund, Instituto de Salud Carlos III) grant no. PI13/00632, and cofunded by the European Union through the Fondo Europeo de Desarrollo Regional (FEDER), ‘A way of shaping Europe’. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval This study involves human participants and the e-MADVEVA Study protocol was approved by the Ethics Committee of the Ramón y Cajal University Hospital, Madrid (code 345) and the Central Research Commission of the Madrid Primary Care Management (code 02/2015). The Ethics Committee did not require informed consent because the research was performed with secondary data. In addition, all evaluators signed a confidentiality clause.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The data that support the findings of this study are available on request from the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.
REFERENCES


Supplementary Table 1. Distribution of misclassification of AMI in the electronic medical records (false positive - FP- and false negative - FN- diagnoses) according to the gold standard

<table>
<thead>
<tr>
<th>AMI</th>
<th>GOLD STANDARD</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic Medical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>412</td>
<td>11 (FP)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>8 (FN)</td>
<td>415</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>420</td>
<td>426</td>
</tr>
</tbody>
</table>

Accuracy: (412 + 415) / (412 + 415 + 11 + 8) = 827 / 846 = 0.9775 (97.75%)
Misclassification: 11 FP + 8 FN = 19

False positives of AMI: 11 cases of 423 (0.026)
- 7 unstable angina (2-3-vessel disease)
- 1 atypical chest pain
- 3 unknowns

False negatives of AMI: 8 cases of 423 (0.019)
- 5 considered as angina were AMI
- 3 unknowns

Supplementary Table 2. Distribution of misclassification of Stroke in the electronic medical records (false positive - FP- and false negative - FN- diagnoses) according to the gold standard

<table>
<thead>
<tr>
<th>STROKE</th>
<th>GOLD STANDARD</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic Medical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>399</td>
<td>24 (FP)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>10 (FN)</td>
<td>413</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>409</td>
<td>437</td>
</tr>
</tbody>
</table>

Accuracy: (399 + 413) / (399 + 413 + 24 + 10) = 812 / 846 = 0.9598 (95.98%)
Misclassification: 24 FP + 10 FN = 34

False positives of Stroke: 24 cases of 423 (0.056)
- 14 transitory ischemic attacks (TIA)
- 5 epileptic attacks
- 1 meningioma
- 1 hypertensive crisis
- 3 unknowns

False negatives of Stroke: 10 cases of 423 (0.023)
- 7 transitory ischemic attacks (TIA) with evolution to stroke
- 1 stroke verified by emergency report
- 2 unknowns
Supplementary Table 3: Diagnostic performance metrics for acute myocardial infarction and stroke, stratified by sex and age group.

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Accuracy % (95% CI)</th>
<th>AUC (95% CI)</th>
<th>F1-SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AMI</strong></td>
<td>412</td>
<td>11</td>
<td>8</td>
<td>415</td>
<td>97.75 (96.12-98.39)</td>
<td>0.98 (0.96-0.99)</td>
<td>0.978</td>
</tr>
<tr>
<td>- Female</td>
<td>113</td>
<td>7</td>
<td>1</td>
<td>119</td>
<td>96.67 (93.54-98.55)</td>
<td>0.97 (0.94-0.99)</td>
<td>0.984</td>
</tr>
<tr>
<td>- Male</td>
<td>299</td>
<td>4</td>
<td>7</td>
<td>296</td>
<td>98.19 (96.78-99.09)</td>
<td>0.98 (0.97-0.99)</td>
<td>0.982</td>
</tr>
<tr>
<td>- Age &lt;70</td>
<td>170</td>
<td>2</td>
<td>2</td>
<td>171</td>
<td>98.84 (97.06-99.68)</td>
<td>0.99 (0.97-1.00)</td>
<td>0.988</td>
</tr>
<tr>
<td>- Age ≥70</td>
<td>242</td>
<td>9</td>
<td>6</td>
<td>244</td>
<td>97.00 (95.10-98.30)</td>
<td>0.97 (0.95-0.98)</td>
<td>0.966</td>
</tr>
<tr>
<td><strong>Stroke</strong></td>
<td>399</td>
<td>24</td>
<td>10</td>
<td>413</td>
<td>95.98 (94.43-97.20)</td>
<td>0.96 (0.95-0.97)</td>
<td>0.959</td>
</tr>
<tr>
<td>- Female</td>
<td>183</td>
<td>14</td>
<td>4</td>
<td>193</td>
<td>95.43 (92.88-97.27)</td>
<td>0.96 (0.93-0.97)</td>
<td>0.953</td>
</tr>
<tr>
<td>- Male</td>
<td>216</td>
<td>10</td>
<td>6</td>
<td>220</td>
<td>96.46 (94.32-97.96)</td>
<td>0.97 (0.94-0.98)</td>
<td>0.964</td>
</tr>
<tr>
<td>- Age &lt;70</td>
<td>141</td>
<td>11</td>
<td>1</td>
<td>152</td>
<td>96.07 (93.23-97.95)</td>
<td>0.96 (0.94-0.98)</td>
<td>0.959</td>
</tr>
<tr>
<td>- Age ≥70</td>
<td>258</td>
<td>13</td>
<td>9</td>
<td>261</td>
<td>95.93 (93.91-97.43)</td>
<td>0.96 (0.94-0.97)</td>
<td>0.959</td>
</tr>
</tbody>
</table>

AMI: acute myocardial infarction; 95% CI: 95% confidence interval; FN: false negative; FP: false positive; TN: true negative; AUC: area under the ROC curve

Supplementary Table 4: Simulation of Acute Myocardial Infarction validation results with identical sample fractions, holding true positives, false positives, and sensitivity constant.

<table>
<thead>
<tr>
<th>AMI present</th>
<th>AMI absent</th>
<th>n</th>
<th>N</th>
<th>Sampling fraction (n/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI code</td>
<td>412</td>
<td>11</td>
<td>423</td>
<td>67,517</td>
</tr>
<tr>
<td>Non-AMI code</td>
<td>8</td>
<td>31,150</td>
<td>31,158</td>
<td>4,973,472</td>
</tr>
<tr>
<td>420</td>
<td>31,161</td>
<td>31,581</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI 95%

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>98.10</td>
</tr>
<tr>
<td>Specificity</td>
<td>99.96</td>
</tr>
<tr>
<td>Prevalence (%)</td>
<td>1.33</td>
</tr>
</tbody>
</table>

95% CI: 95% confidence interval.
Supplementary Table 5. Simulation of stroke validation results with identical sample fractions, holding true positives, false positives, and sensitivity constant.

<table>
<thead>
<tr>
<th>Stroke code</th>
<th>Stroke present</th>
<th>Stroke absent</th>
<th>n</th>
<th>N</th>
<th>Sampling fraction (n/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>399</td>
<td>24</td>
<td>423</td>
<td>423</td>
<td>61,718</td>
</tr>
<tr>
<td>Non-Stroke</td>
<td>34,120</td>
<td>34,130</td>
<td>34,130</td>
<td>4,979,270</td>
<td>0,6854</td>
</tr>
<tr>
<td>code</td>
<td>409</td>
<td>34,144</td>
<td>34,553</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI 95%

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>97.56</td>
<td>95.56-98.82</td>
</tr>
<tr>
<td>Specificity</td>
<td>99.93</td>
<td>99.89-99.96</td>
</tr>
<tr>
<td>Prevalence (%)</td>
<td>1.18</td>
<td>1.07-1.30</td>
</tr>
</tbody>
</table>