BMJ Open

Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study

Rosa Gini,[1,2] Martijn J Schuemie,[3,4] Giampiero Mazzaglia,[5] Francesco Lapi,[5] Paolo Francesconi,[1] Alessandro Pasqua,[5] Elisa Bianchini,[5] Carmelo Montalbano,[6] Giuseppe Roberto,[1] Valentina Barletta,[1] Iacopo Cricelli,[6] Claudio Cricelli,[7] Giulia Dal Co,[8] Mariadonata Bellentani,[8] Miriam Sturkenboom,[2] Niek Klazinga[9]

For numbered affiliations see end of article.

Correspondence to
Dr Rosa Gini;
rosa.gini@ars.toscana.it

ABSTRACT

Objectives: The Italian project MATRICE aimed to assess how well cases of type 2 diabetes (T2DM), hypertension, ischaemic heart disease (IHD) and heart failure (HF) and their levels of severity can be automatically extracted from the Health Search/CSD Longitudinal Patient Database (HSD). From the medical records of the general practitioners (GP) who volunteered to participate, cases were extracted by algorithms based on diagnosis codes, keywords, drug prescriptions and results of diagnostic tests. A random sample of identified cases was validated by interviewing their GPs.

Setting: HSD is a database of primary care medical records. A panel of 12 GPs participated in this validation study.

Participants: 300 patients were sampled for each disease, except for HF, where 243 patients were assessed.

Outcome measures: The positive predictive value (PPV) was assessed for the presence/absence of each condition against the GP's response to the questionnaire, and Cohen's κ was calculated for agreement on the severity level.

Results: The PPV was 100% (99% to 100%) for T2DM and hypertension, 98% (96% to 100%) for IHD and 55% (49% to 61%) for HF. Cohen's kappa for agreement on the severity level was 0.70 for T2DM and 0.69 for hypertension and IHD.

Conclusions: This study shows that individuals with T2DM, hypertension or IHD can be validly identified in HSD by automated identification algorithms. Automatic queries for levels of severity of the same diseases compare well with the corresponding clinical definitions, but some misclassification occurs. For HF, further research is needed to refine the current algorithm.

Strengths and limitations of this study

- This is the first validation study specifically performed in the Italian General Practitioners' (GP) medical records database Health Search.
- The positive predictive value of automatic case-finding algorithms for four chronic diseases and their levels of severity was estimated, using manual assessment of the GPs themselves as a gold standard.
- A total of 12 GPs contributed to the study and 300 cases were assessed for type 2 diabetes, hypertension and ischaemic heart disease, 243 for heart failure.
- This study is part of a national Italian project funded by the Ministry of Health and aimed to validate information systems available in the country to monitor quality of healthcare for chronic diseases.
- Participation of the GPs was voluntary, and sensitivity could not be assessed.

INTRODUCTION

Italy is facing an increasing burden of chronic health conditions due to ageing of the population. To provide adequate and fair healthcare across regions, Italy was advised by the Organisation for Economic Co-operation and Development to develop a set of standards around the processes and outcomes of primary care and to develop a national quality governance model to support regions in delivering care of uniform quality across the country.[1] The Italian National Agency for Regional Healthcare Services started the MATRICE Project in 2011. MATRICE was

aimed at developing tools to compare quality of health-care across Italian regions of four chronic diseases: type 2 diabetes (T2DM), hypertension, ischaemic heart disease (IHD) and heart failure (HF). One of the objectives was to assess the validity of routine care data to monitor quality of healthcare supply.[2][3]

The Health Search IMS Health Longitudinal Patient Database (HSD) is a longitudinal primary care medical record database that was set up by members of the Italian College of General Practitioners (SIMG). More than 900 physicians, uniformly distributed across Italy, share their de-identified clinical records in the HSD. These data are extensively used for epidemiological and public health research.[4][5][6][7][8][9][10][11][12]

The HSD database is very similar to other Primary Care databases: for instance, in the UK, the Clinical Practice Research Datalink (CPRD, formerly GPRD),[13] The Health Information Network (THIN)[14] and QResearch;[15] in Canada, the Canadian Primary Care Sentinel Surveillance Network (CPCSSN);[16] in the Netherlands, the Integrated Primary Care Information database (IPCI);[17] in Spain, the database Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP).[18] A common feature of these countries is that the GP serves a well-defined population and is the gatekeeper to secondary care.

In this type of medical records, every visit is recorded and all diagnoses, prescriptions and measurements are recorded as part of a general practitioner's daily practice. Moreover, information from specialist referrals is reported back to the general practitioner (GP) and stored in the same medical records. The medical records replace the paper records that once existed and may be considered a rather comprehensive list of health problems requiring care. Owing to its longitudinal, population-based nature, this type of databases serves well for many research purposes, such as estimate of burden of disease, pharmacoepidemiologic and health services research. Since information is collected primarily for the provision of care, the quality of coding diagnoses may not always be accurate and also this varies by type of disease.[19][20] Moreover, the same code may be used for different clinical definitions of a disease, when diagnostic standards are not uniform across healthcare communities or change over time. As a consequence, case-finding algorithms that retrieve participants from such data sources using diagnostic codes may unintentionally retrieve participants whose clinical condition does not correspond to the one intended for a study. On the other hand, GPs may record clinical conditions as a free text note, and a retrieval strategy using only coded diagnoses may miss some cases. As a result, algorithms using other sources, such as results from laboratory tests, have been developed to query this type of data sources.[21] Disease-specific validation studies of case-finding algorithms in GP medical record databases have been performed in CPRD,[22][23] CPCSSN,[20][24] IPCI and HSD itself.[25][26]

A panel of experts in the MATRICE Project established precise clinical definitions of T2DM, hypertension, IHD and HF, and identified levels of severity of the conditions that had to be distinguished for the purpose of monitoring healthcare quality. The aim of this study was to estimate the positive predictive value (PPV) of case-finding algorithms to detect such conditions and levels of severity from the GP medical records collected by HSD, against a gold standard based on manual comparison with the clinical definitions chosen by the MATRICE panel.

METHODS
Setting
Italy has a tax-based, universal coverage national health system. Every Italian resident is entitled to choose a GP, although parents might instead opt for a specialist paediatrician for their children, up to the age of 15. Therefore, each resident from the age of 16 onwards is specifically registered with a GP. GPs are the 'gate-keepers' of the system, meaning that patients can only access secondary care within the healthcare system on referral of their GP.[27][1] Secondary care is accessed either free of charge or on a small copayment.

During their daily practice, GPs record all clinical findings, diagnoses and prescriptions in their electronic medical records. GPs participating in HSD all use the same software, which requires that each prescription is associated with a specific disease code. A disease code may be labelled as 'suspect' when further clinical ascertainment is needed. Results from laboratory tests may be recorded as well. Moreover, free-text fields are available in the software to collect clinical notes on diagnoses, signs, symptoms and referral letters from specialists or from hospitals. Every 6 months, GPs send their data to a central repository, after anonymisation. The central repository performs quality controls, like estimation of prevalence of common diseases, and selects GPs whose data prove to be accurate.[4] Currently, data of 700 of 900 GPs, uniformly distributed across Italy, are considered accurate according to data quality checking.[12]

Clinical definition of the diseases and of their levels of severity
A panel of cardiologists, diabetologists, epidemiologists and experts in organisation of primary care services participating in the MATRICE Project first established clinical definitions of the four diseases and of their levels of severity. The levels of severity were selected according to whether national and international clinical guidelines contained specific indications for treatment or diagnostic follow-up in the patients with that condition. For instance, a patient with IHD after an episode of acute myocardial infarction (AMI) has an indication for treatment with β-blockers;[28] hence, history of AMI is a relevant level of severity for IHD.

The detailed definitions of the diseases and of the levels of severity are depicted in table 1. For T2DM, the

Table 1 Clinical definition of diseases and levels of severity

Clinical definition	Levels of severity
Type 2 diabetes mellitus Syndrome diagnosed on the basis of the following criteria outlined in a first test and confirmed with a second test in an adult, non-pregnant patient, without typical symptoms of the disease: Fasting plasma glucose ≥126 mg/dL (no caloric intake for at least 8 hours), or 2-hour plasma glucose ≥200 mg/dL during an oral glucose tolerance test (OGTT) after a load of 75 g glucose, or glycated haemoglobin ≥6.5%. Or, in a patient with classic symptoms of hyperglycaemia or hyperglycaemic crisis, a random plasma glucose of plasma glucose ≥200 mg/dL (regardless of food intake).	Level 1: Clinical definition of the disease, no indication for insulin therapy and no complications listed in level 3 Level 2: Clinical definition of the disease, indication for insulin therapy and absence of complications listed in level 3. Level 3: Clinical definition of the disease, no indication for insulin therapy and one of the following: (1) arterial stenosis (coronary, carotid, peripheral arteries of lower extremities), angina pectoris, MI, TIA, ischaemic stroke of atherosclerotic origin, intermittent claudication, diabetic foot ulcer, lower limb amputation, (2) retinopathy, (3) incipient diabetic nephropathy (microalbuminuria) or overt (albuminuria or GFR abnormal)/dialysis Level 4: As in level 3, except that insulin is indicated
Hypertension Syndrome characterised by arterial systolic blood pressure above 140 mm Hg and/or diastolic blood pressure above 90 mm Hg in at least two measurements (patient at rest) confirmed by Holter blood pressure measurements or by home blood pressure monitoring (two measurements in the morning and two in the evening for 7 days) and then calculating the average of all measurements after discarding those of the first day (as recommended by the ESH guidelines)	Level 1: Clinical definition of the disease, absence of organ damage and of diabetes Level 2: Clinical definition of the disease, no HF in level at least C of the ACC/AHA classification and at least one of the following conditions: type 2 diabetes; hypertrophy (ECG or Echo), dilation or left ventricular asynergy (Echo); hypertensive retinopathy; GFR abnormal; microalbuminuria or proteinuria; atherosclerotic plaques in carotid arteries; atherosclerotic peripheral arterial occlusive disease; angina pectoris; coronary revascularisation; AMI; TIA or ischaemic stroke due to atherosclerosis; hypertensive encephalopathy; abdominal aortic aneurysm; aortic dissection; cerebral haemorrhage Level 3: Clinical definition of the disease and HF in stage C or D of the ACC/AHA classification.
Ischaemic heart disease Clinical syndrome characterised by typical angina chest pain, and/or transient myocardial ischaemia verified by stress ECG or imaging, and/or significant coronary arteries occlusion verified with angiography, or history of previous AMI.	Level 1: Clinical definition of the disease, no evidence of previous AMI nor PTCA, no evidence of HF in stage C or D of the ACC/AHA classification. Level 2: Evidence of previous PTCA, no evidence of previous AMI, no evidence of HF in stage C or D of the ACC/AHA classification. Level 3: Evidence of previous AMI, no evidence of previous PTCA, no evidence of HF in stage C or D of the ACC/AHA classification. Level 4: Evidence of previous PTCA and AMI, no evidence of HF in stage C or D of the ACC/AHA classification. Level 5: Clinical definition of the disease and evidence of HF in stage C or D of the ACC/AHA classification.
Heart failure Stage C or D of the ACC/AHA classification: syndrome characterised by the presence of symptoms and signs, current or prior dyspnoea and/or fatigue and/or fluid retention (peripheral oedema and/or pulmonary stasis), and the presence of structural heart disease (left ventricular systolic dysfunction with EF <50% and/or left ventricular diastolic dysfunction and/or right ventricular dysfunction) detected by echocardiography	None

ACC/AHA: American Cardiology Association and American Heart Association; AMI, acute myocardial infarction; EF, ejection fraction; HF, heart failure; GFR, glomerular filtration rate; PTCA, percutaneous transluminal coronary angioplasty.

clinical definition was at least two abnormal measurements among fasting plasma glucose, or 2-hour plasma glucose after a load of glucose, or glycated haemoglobin; or just one abnormal measurement of plasma glucose if symptoms of hyperglycaemia were observed. Four levels of severity of the disease were identified, according to the presence/absence of indication for insulin and the presence/absence of complications or organ damage. For hypertension, diagnostic criteria were two abnormal measurements for systolic and diastolic blood pressure confirmed either by a Holter blood pressure measurement or by home blood pressure monitoring. Three levels of severity were identified: no organ damage or diabetes or stroke; organ damage or diabetes or stroke without HF; hypertension with HF. For IHD, the clinical definition referred to symptoms (angina pain) or to a history of AMI or to a bioimaging observation of coronary ischaemia. Five levels of severity of IHD were identified: the most severe was HF, and among those free from HF the presence/absence of history of AMI and the presence/absence of percutaneous transluminal coronary angioplasty (PTCA) classified the four levels. HF was also identified as a condition on its own, and the clinical definition was stage C or D of the classification by the American College of Cardiology and of the American Heart Association.[29]

Case identification in the primary care medical records

A panel comprising epidemiologists from HSD and GPs belonging to SIMG, with expertise in the clinical areas of interest, developed ad hoc algorithms to identify from the GP medical records cases matching the clinical definitions of the MATRICE Project. In each algorithm, the inclusion and exclusion criteria of the clinical definition were mapped to a list of ICD9CM codes or, when deemed necessary, to free text and to conditions on diagnostic tests.

Case validation

An invitation to participate in the validation study was circulated by SIMG to a sample of GPs, and participation was voluntary.

A data collection plugin for the medical record software was developed by HSD and installed in the computers of the GPs who accepted to participate in the study.

For each disease, the plugin applied the algorithm to the whole list of active patients on the date of data collection and selected a random sample of 25 patients from the resulting list of cases. The plugin then showed the names of the patients in the sample to the GP for assessment. On the same screen, the clinical definitions of the disease and of its levels of severity were presented. The GP was aware that the cases were selected for a specific condition among T2DM, hypertension, IHD and HF, but was blinded to the level of severity. The GP had the choice of indicating that the patient was not affected by the disease (false positive: FP) or that the patient was affected but the level of severity could not be assessed on the basis of the information available to the GP (not staged: NS), or to assign a level. In the

process, the GP was free to access the patient's medical record. Figure 1 shows a screenshot of the data collection plugin (patients listed are not real).

When the GP had completed the manual assessment of the four samples of patients, the plugin applied the algorithms for the level of severity, linked the new columns to the data set resulting from the questionnaire, anonymised the final data set and transmitted it to HSD for statistical analysis.

Data collection was performed in July 2013.

Statistical analysis

The formula to compute the PPV, for presence/absence of the condition and for each of the levels of severity, was

$$PPV = true\ positives/(automatically\ assigned)$$

and 95% CIs were estimated. For the three diseases where levels of severity had been validated (T2DM, hypertension and IHD), Cohen's κ was computed to the categorical distribution of levels of severity. Cohen's kappa discounts from observed concordance (the percentage of participants who are classified in the same level by the algorithms and the GP), the expected concordance (the percentage of participants who would be classified the same if assignment had been performed randomly) by means of the following formula

$$K = (observed\ concordance - expected\ concordance)/ \\ (1 - expected\ concordance)$$

Cohen's κ provides an overall measure of agreement about levels of severity. Analysis was performed on the whole sample of patients, irrespectively on the GP. Analysis was performed using Stata V.12.

RESULTS
Algorithms for primary care medical records

The algorithms detecting the diseases and levels of severity from the primary care medical records are listed in table 2. Each algorithm consists of a sequence of rules, each acting as an inclusion criterion, a refinement criterion (linked to the inclusion criterion with the logical connector AND) or a refinement criterion (linked to the inclusion criterion with the logical connectors AND NOT). Each rule is itself composed by subqueries (represented in the table by keywords in round parentheses), and subjects matching at least one of the subqueries are included in the rule, that is, the subqueries are linked to each other with the logical connector OR. No specific temporal sequence between subqueries is requested. Every subquery selects records matching a specific list of codes, free text keywords and/or diagnostic test levels, which are listed in Supplementary material 1. All the subqueries are applied to the whole set of longitudinal observations of the patients up to the index date, except when specified otherwise.

Figure 1 Screenshot of the data collection plugin. The GP participating in the study was asked to fill in the four tables, one per disease. In the table, 'S.P.' meant 'without the disease' and 'N.S.' meant that the patient had the disease but the GP was not able to assess the level of severity. The clinical definition of the condition and of each level of severity could be browsed in the upper part of the screen, while, whenever the pointer stopped on a cell in the table, a tooltip suggested the label of the corresponding level. The patients listed in the figure are not real. GP, general practitioner.

In all subqueries, records labelled with 'suspect' were excluded, except in the case of the subquery detecting patients with AMI, used to detect levels of severity 3 and 4 in patients with IHD.

Validation

A panel of 12 GPs participated in the validation study. A total of 300 patients were identified and validated for each disease, except for HF, where due to low prevalence of the condition only 243 patients were included.

The PPV of the algorithms were 100% (CI 99% to 100%) for T2DM and hypertension and 98% (CI 96% to 100%) for IHD. For HF, PPV was 55% (CI 49% to 61%).

For T2DM, the second and fourth levels of severity had very high PPV (88% and 93%, respectively). Around 20% of patients without indication for insulin (first and third levels of severity) had their presence of complications misclassified. Both possibilities took place: patients with complications were identified as being free from them, and vice versa. Overall, Cohen's kappa was 0.70, indicating a good level of agreement (table 3).

Among hypertensive patients, every level of severity was misclassified in <20% of patients, and in the case of the middle level (organ damage and/or diabetes, no HF) patients were in fact almost all less severe with respect to the level they were automatically assigned to. Cohen's κ was 0.69, showing good agreement (table 3).

In the case of IHD, the first and fifth levels of severity had excellent PPV, while AMI was incorrectly identified by the algorithm in 22% of cases. The two levels of severity characterised by the presence of PTCA were never manually indicated by the GPs, and the automatic algorithm was in almost perfect agreement in both cases. Overall, Cohen's κ was 0.69, showing good agreement (table 3).

DISCUSSION

This study shows that almost all of the automatically detected cases of T2DM, hypertension and IHD, but only the 55% of cases of HF were true cases as assessed by the GP, on the basis of their own records and personal knowledge. Automatic classification of levels of severity of T2DM, hypertension and IHD was acceptable although less accurate. In the case of IHD, mild cases were misclassified as severe, while in T2DM and

Table 2 Algorithms to detect diseases and levels of severity

Algorithm for the disease	Algorithms for levels of severity
Type 2 diabetes mellitus [(DM) OR (GFR) OR (DM TESTS)] **AND NOT** (DM1)	Algorithm for the disease **AND** (DM UNCOMPLICATED) **AND NOT** [(INSULIN) OR (DM2 CHRONIC COMPLICATIONS) OR (DM2 ASYMPTOMATIC COMPLICATIONS) OR (CVD)] Algorithm for the disease **AND** (DM UNCOMPLICATED) **AND** (INSULIN) **AND NOT** [(DM2 CHRONIC COMPLICATIONS) OR (DM2 ASYMPTOMATIC COMPLICATIONS) OR (CVD)] Algorithm for the disease **AND** [(DM CHRONIC COMPLICATIONS) OR (DM ASYMPTOMATIC COMPLICATIONS) OR (CVD)] **AND NOT** (INSULIN) Algorithm for the disease **AND** [(DM CHRONIC COMPLICATIONS) OR (DM ASYMPTOMATIC COMPLICATIONS) OR (CVD)] **AND** (INSULIN)
Hypertension (HYPERTENSION) OR (BP)	Algorithm for the disease **AND NOT** [LEVEL 2 OR LEVEL 3] Algorithm for the disease **AND** [(DM2) OR (HYPERTENSION COMPLICATIONS) OR (HYPERTENSIVE RETINOPATHY) OR (CHRONIC KIDNEY DISEASE) OR (ATHEROSCLEROTIC ARTERIOPATHY) OR (CAROTID ATHEROSCLEROSIS) OR (HYPERTENSIVE ENCEPHALOPATHY) OR (AORTIC ANEURYSM)] **AND NOT** [(HF) OR (LVEF)] Algorithm for the disease **AND** [(HF) OR (LVEF) OR (VENTRICULAR DYSFUNCTION)]
Ischaemic heart disease (CHD)	Algorithm for the disease **AND** (CHD NO AMI) **AND NOT** [(AMI) OR (PAST AMI) OR (CORONARY REVASCULARIZATION ICD9CM) OR (CORONARY REVASCULARIZATION FREE TEXT) OR (HF) OR (LVEF) OR (VENTRICULAR DYSFUNCTION)] Algorithm for the disease **AND** [(CORONARY REVASCULARIZATION ICD9CM) OR (CORONARY REVASCULARIZATION FREE TEXT)] **AND NOT** [(AMI) OR (PAST AMI) OR (HF) OR (LVEF) OR (VENTRICULAR DYSFUNCTION)] Algorithm for the disease **AND** [(AMI) OR (PAST AMI)] **AND NOT** [(CORONARY REVASCULARIZATION ICD9CM) OR (CORONARY REVASCULARIZATION FREE TEXT) OR (HF) OR (LVEF) OR (VENTRICULAR DYSFUNCTION)] Algorithm for the disease **AND** [(AMI) OR (PAST AMI) OR (CORONARY REVASCULARIZATION ICD9CM) OR (CORONARY REVASCULARIZATION FREE TEXT)] **AND NOT** [(HF) OR (LVEF) OR (VENTRICULAR DYSFUNCTION)] Algorithm for the disease **AND** [(HF) OR (LVEF)]
Heart failure (HF) OR (LVEF) OR (VENTRICULAR DYSFUNCTION)	None

The algorithms are described by means of subqueries, represented by keywords in upper case between parentheses, whose details are in online supplementary table 1. In particular, measurements are as follows: glomerular filtration rate (GFR) <mL/min/1.73 m^2; systolic blood pressure (SBP) >140 mm Hg at least twice ever or diastolic blood pressure (DBP) >90 mm Hg at least twice ever; Left ventricular ejection fraction (LVEF) <50%; DM TESTS: plasma glucose >200 mg/dL after an oral load of 75 g glucose or fasting plasma glucose ≥126 mg/dL or glycated haemoglobin ≥6.5%. In all subqueries, records labelled with 'suspect' were excluded, except in the case of the subquery (acute myocardial infarction, AMI) used to detect levels of severity 3 and 4 of ischaemic heart disease (IHD).

Table 3 Results of the validation

Level	Description	Automatically assigned	True positives	True level of false positives	PPV (95% CI)	Cohen's K
Type 2 diabetes mellitus (N=300, true positives overall: 300)						
1	Diabetes with no evidence of organ damage nor complications and no indication for insulin therapy	129	101	Level 3: 28	0.78 (0.70 to 0.85)	0.70
2	Diabetes with indication for insulin but no evidence of organ damage nor complications	17	15	Level 4: 2	0.88 (0.64 to 0.99)	
3	Diabetes with evidence of organ damage or complications and no indication for insulin	127	104	Level 1: 23	0.82 (0.74 to 0.88)	
4	Diabetes with evidence of organ damage or complications and with indication for insulin	27	25	Level 3: 2	0.93 (0.76 to 0.99)	
Hypertension (N=300, true positives overall: 300)						
1	Hypertension with no organ damage nor diabetes	138	119	Level 2: 18 Level 3: 1	0.86 (0.79 to 0.91)	0.69
2	Hypertension with organ damage and/or stroke and/or diabetes, no HF	155	127	Level 1: 28	0.82 (0.75 to 0.88)	
3	Hypertension and HF	7	6	Level 2: 1	0.86 (0.42 to 0.99)	
Ischaemic heart disease (N=300, true positives overall: 298)						
1	IHD, no AMI, no HF, no PTCA	56	54	No IHD: 2	0.96 (0.88 to 0.99)	0.71
2	IHD with PTCA, no AMI, no HF	1	0	Level 5: 1	0	
3	IHD with AMI, no PTCA, no HF	195	152	Level 1: 43	0.78 (0.71 to 0.84)	
4	IHD with PTCA and AMI, no HF	0	0	–	–	
5	IHD with HF	48	45	Level 3: 3	0.94 (0.83 to 0.99)	
Heart failure (N=243, true positives: 134)						

AMI, acute myocardial infarction; IHD, ischaemic heart disease; HF, heart failure; PPV, positive predictive value; PTCA, percutaneous transluminal coronary angioplasty.

hypertension both possibilities took place: severe patients were automatically classified as mild and vice versa. Our results provide guidance on the interpretation of results of studies using those algorithms to define variables in medical records or HSD, for instance to monitor quality of healthcare.

The excellent PPV of some algorithms is not unexpected: algorithms to detect diabetes (irrespective of type) and hypertension, as well as other chronic conditions, had similarly high validity in the CPCSSN.[20] [24] The low PPV that we observed in the case of HF is unsurprising as well. Indeed, HF is a syndrome, and several different clinical definitions of HF have been used among clinicians in the recent past, such as the Framingham and European Society of Cardiology criteria. It has been shown that the changing definition has a noticeable impact on the epidemiology of the condition.[30] Our definition comprises two of the four stages of the classification of the American College of Cardiology and of the American Heart Association,

which has been itself revised repeatedly in recent years.[29] It is most likely that the GPs in the sample themselves adopt a different definition to diagnose HF with respect to the one proposed by the MATRICE panel. Further research could investigate whether a clinical definition modelled on the Italian guidelines may be more easily identified in primary care medical records. Finally, poor performance of ICD9CM codes in identifying a specific clinical definition of HF has been consistently reported in the literature.[31]

The results we observed for levels of severity are probably due to reasons which are more specific to HSD. In the case of T2DM, some misclassification occurred between level 1 and level 3, that is, patients with or without complications but without indication for insulin use. Adding rules that explore free text comments to the diagnostic codes may improve those algorithms. In the algorithm developed for this study, patients with IHD labelled with 'suspect' AMI were included in the 'AMI' level of severity, because in a previous study specific

algorithms to identify AMI had been observed to have low sensitivity in HSD.[25] As a result, our sensitive algorithm did indeed capture all the cases of AMI in patients with IHD; however, 22% of patients had not had an AMI, so this strategy needs to be reconsidered for future research. Remarkably, GPs were not aware of a single case of PTCA in their patients, as confirmed by automatic querying of their records. The absence of PTCA could be due to the fact that this procedure is only performed in hospital. In Italy, hospitals do not provide GPs with discharge letters; rather patients themselves must describe to the GP what happened during an inpatient care episode, and the PTCA procedure may be communicated inadequately to the GP.

Strategies to improve communication between hospital and primary care should be implemented in Italy for the purpose of improving quality of primary care medical records, as well as to improve healthcare for patients with severe cardiovascular conditions.

Limitations

The sample of GPs was self-selected and may have been composed of those with more accurate data recording attitudes. In particular, two GPs in the validation sample also participated in the panel that created the algorithms. The PPV of the algorithms may be lower in the general group of GPs contributing to HSD.

Providing a direct estimate of sensitivity of the case-finding algorithms was not an aim of this study, because it would have been unfeasible for GPs to assess a large enough sample of their patients. Indeed, the number of patients needed to estimate the sensitivity of a test is bigger with less prevalent conditions: for instance, to estimate sensitivity of a case-finding algorithm for T2DM with a marginal error of 5%, even assuming 10% prevalence (an overestimation, according to common estimates in Italy[32]) and 95% sensitivity would have required an additional sample of more than 700 participants; to obtain a valid estimate of sensitivity for HF, which has much lower prevalence and a lower expected sensitivity, GPs would have needed to assess thousands of participants.[33]

CONCLUSIONS

This study shows that participants with T2DM, hypertension or IHD can be validly identified in HSD by automated identification algorithms. Automatic queries for levels of severity of the same diseases compare well with the corresponding clinical definitions, but some misclassification occurs. For HF, further research is needed to refine the current algorithm.

Author affiliations
[1]Agenzia regionale di sanità della Toscana, Osservatorio di epidemiologia, Florence, Italy
[2]Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands
[3]Department of Epidemiology Janssen Research & Development, Titusville, New Jersey, USA
[4]Observational Health Data Sciences and Informatics (OHDSI), New York, New York, USA
[5]Health Search, Italian College of General Practitioners and Primary Care, Florence, Italy
[6]Genomedics, Florence, Italy
[7]Italian College of General Practitioners and Primary Care, Florence, Italy
[8]Agenzia Nazionale per il Servizi Sanitari Regionali, Rome, Italy
[9]Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

REFERENCES

1. OECD. *OECD reviews of health care quality. Italy 2014: raising standards*. OECD Publishing, 2015.
2. Agenzia nazionale per i Servizi Sanitari Regionali. Programma Mattoni del SSN—Progetto MATRICE. http://www.agenas.it/images/agenas/In%20primo%20piano/Matrice/Progetto_MATRICE_Scheda_informativa.pdf (accessed Aug 2016) [Italian].
3. Agenzia regionale di sanità della Toscana. Data integration for chronic diseases management in outpatient settings (MATRICE Project). https://www.ars.toscana.it/en/project/chronic-diseases/2460-matrice-project.html (accessed Aug 2016).
4. Cricelli C, Mazzaglia G, Samani F, *et al*. Prevalence estimates for chronic diseases in Italy: exploring the differences between self-report and primary care databases. *J Public Health Med* 2003;25:254–7.
5. Savica R, Beghi E, Mazzaglia G, *et al*. Prescribing patterns of antiepileptic drugs in Italy: a nationwide population-based study in the years 2000–2005. *Eur J Neurol* 2007;14:1317–21.
6. Sacchetti E, Trifirò G, Caputi A, *et al*. Risk of stroke with typical and atypical anti-psychotics: a retrospective cohort study including unexposed subjects. *J Psychopharmacol (Oxford)* 2008;22:39–46.
7. Mazzaglia G, Ambrosioni E, Alacqua M, *et al*. Adherence to antihypertensive medications and cardiovascular morbidity among newly diagnosed hypertensive patients. *Circulation* 2009;120:1598–605.
8. Ravera M, Noberasco G, Re M, *et al*. Chronic kidney disease and cardiovascular risk in hypertensive type 2 diabetics: a primary care perspective. *Nephrol Dial Transplant* 2009;24:1528–33.
9. Cazzola M, Bettoncelli G, Sessa E, *et al*. Prevalence of comorbidities in patients with chronic obstructive pulmonary disease. *Respiration* 2010;80:112–9.
10. Cazzola M, Calzetta L, Lauro D, *et al*. Asthma and COPD in an Italian adult population: role of BMI considering the smoking habit. *Respir Med* 2013;107:1417–22.

11. Sultana J, Italiano D, Spina E, *et al*. Changes in the prescribing pattern of antidepressant drugs in elderly patients: an Italian, nationwide, population-based study. *Eur J Clin Pharmacol* 2014;70:469–78.

12. Health Search Web page. https://www.healthsearch.it/?lang=en (accessed Aug 2016).

13. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997;350:1097–9.

14. Lewis JD, Schinnar R, Bilker WB, *et al*. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2007;16:393–401.

15. Hippisley-Cox J, Stables D, Pringle M. QRESEARCH: a new general practice database for research. *Inform Prim Care* 2004;12:49–50.

16. Birtwhistle RV. Canadian Primary Care Sentinel Surveillance Network: a developing resource for family medicine and public health. *Can Fam Physician* 2011;57:1219–20.

17. Vlug AE, van der Lei J, Mosseveld BM, *et al*. Postmarketing surveillance based on electronic patient records: the IPCI project. *Methods Inf Med* 1999;38:339–44.

18. Salvador Rosa A, Moreno Pérez JC, Sonego D, *et al*. El Proyecto BIFAP: Base de datos para la Investigación Farmacoepidemiológica en. *Aten Primaria* 2002;30:655–61.

19. Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. *Fam Pract* 2004;21:396–412.

20. Kadhim-Saleh A, Green M, Williamson T, *et al*. Validation of the diagnostic algorithms for 5 chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a Kingston Practice-based Research Network (PBRN) report. *J Am Board Fam Med* 2013;26:159–67.

21. Tu K, Manuel D, Lam K, *et al*. Diabetics can be identified in an electronic medical record using laboratory tests and prescriptions. *J Clin Epidemiol* 2011;64:431–5.

22. Herrett E, Thomas SL, Schoonen WM, *et al*. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010;69:4–14.

23. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract* 2010;60:e128–36.

24. Williamson T, Green ME, Birtwhistle R, *et al*. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann Fam Med* 2014;12:367–72.

25. Coloma PM, Valkhoff VE, Mazzaglia G, *et al*. Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. *BMJ Open* 2013;3;pii: e002862.

26. Valkhoff VE, Coloma PM, Masclee GMC, *et al*. Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. *J Clin Epidemiol* 2014;67:921–31.

27. Lo Scalzo A, Donatini A, Orzella L, *et al*. Italy: Health system review. *Health Syst Transit* 2009;11:1–216.

28. Fihn SD, Gardin JM, Abrams J, *et al*. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the Diagnosis and Management of Patients With Stable Ischemic Heart Disease: Executive Summary. A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *Circulation* 2012;126:3097–137.

29. Hunt SA, Abraham WT, Chin MH, *et al*. ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure): developed in collaboration with the American College of Chest Physicians and the International Society for Heart and Lung Transplantation: Endorsed by the Heart Rhythm Society. *Circulation* 2005;112:e154–235.

30. Di Bari M, Pozzi C, Cavallini MC, *et al*. The diagnosis of heart failure in the community. Comparative validation of four sets of criteria in unselected older adults: the ICARe Dicomano Study. *J Am Coll Cardiol* 2004;44:1601–8.

31. Quach S, Blais C, Quan H. Administrative data have high variation in validity for recording heart failure. *Can J Cardiol* 2010;26:e306–12.

32. Gini R, Francesconi P, Mazzaglia G, *et al*. Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey. *BMC Public Health* 2013;13:15.

33. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193–204.