



## ORIGINAL ARTICLE

# Is diabetes mellitus correctly registered and classified in primary care? A population-based study in Catalonia, Spain



Manel Mata-Cases<sup>a,b,c</sup>, Dídac Mauricio<sup>a,b,d</sup>, Jordi Real<sup>a,e</sup>, Bonaventura Bolívar<sup>f,g</sup>, Josep Franch-Nadal<sup>a,b,h,\*</sup>

<sup>a</sup> DAP-Cat group, Unitat de Suport a la Recerca Barcelona Ciutat, Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain

<sup>b</sup> Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III (ISCIII), Spain

<sup>c</sup> Primary Health Care Center La Mina, Gerència d'Àmbit d'Atenció Primària Barcelona Ciutat, Institut Català de la Salut, Sant Adrià de Besòs, Spain

<sup>d</sup> Department of Endocrinology & Nutrition, Health Sciences Research Institute & Hospital Universitari Germans Trias i Pujol, Badalona, Spain

<sup>e</sup> Universitat Internacional de Catalunya, Epidemiologia i Salut Pública, Sant Cugat, Spain

<sup>f</sup> Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain

<sup>g</sup> Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain

<sup>h</sup> Primary Health Care Center Raval Sud, Gerència d'Àmbit d'Atenció Primària Barcelona Ciutat, Institut Català de la Salut, Barcelona, Spain

Received 4 June 2016; accepted 19 July 2016

Available online 6 September 2016

### KEYWORDS

Diabetes;  
Primary care;  
Classification;  
Coding

### Abstract

**Objective:** To assess the prevalence of miscoding, misclassification, misdiagnosis and under-registration of diabetes mellitus (DM) in primary health care in Catalonia (Spain), and to explore use of automated algorithms to identify them.

**Methods:** In this cross-sectional, retrospective study using an anonymized electronic general practice database, data were collected from patients or users with a diabetes-related code or from patients with no DM or prediabetes code but treated with antidiabetic drugs (unregistered DM). Decision algorithms were designed to classify the true diagnosis of type 1 DM (T1DM), type 2 DM (T2DM), and undetermined DM (UDM), and to classify unregistered DM patients treated with antidiabetic drugs.

\* Corresponding author.

E-mail address: [josep.franch@gmail.com](mailto:josep.franch@gmail.com) (J. Franch-Nadal).

**Results:** Data were collected from a total of 376,278 subjects with a DM ICD-10 code, and from 8707 patients with no DM or prediabetes code but treated with antidiabetic drugs. After application of the algorithms, 13.9% of patients with T1DM were identified as misclassified, and were probably T2DM; 80.9% of patients with UDM were reclassified as T2DM, and 19.1% of them were misdiagnosed as DM when they probably had prediabetes. The overall prevalence of miscoding (multiple codes or UDM) was 2.2%. Finally, 55.2% of subjects with unregistered DM were classified as prediabetes, 35.7% as T2DM, 8.5% as UDM treated with insulin, and 0.6% as T1DM.

**Conclusions:** The prevalence of inappropriate codification or classification and under-registration of DM is relevant in primary care. Implementation of algorithms could automatically flag cases that need review and would substantially decrease the risk of inappropriate registration or coding.

© 2016 SEEN. Published by Elsevier España, S.L.U. All rights reserved.

## PALABRAS CLAVE

Diabetes;  
Atención primaria;  
Clasificación;  
Codificación

## ¿Se registra y clasifica correctamente la diabetes mellitus en atención primaria? Estudio poblacional en Cataluña, España

### Resumen

**Objetivo:** Evaluar la prevalencia de errores en la codificación, clasificación, diagnóstico e infrarregistro de la diabetes mellitus (DM) en asistencia primaria en Cataluña y explorar el uso de algoritmos automáticos para identificarlos.

**Métodos:** En este estudio transversal, retrospectivo utilizando una base electrónica de datos anonimizados se extrajeron los datos de pacientes con algún código de diabetes o de pacientes sin código de DM o prediabetes pero tratados con antidiabéticos (DM no registrada). Se diseñaron algoritmos de decisión para clasificar el verdadero diagnóstico de diabetes tipo 1 (DM1), diabetes tipo 2 (DM2) y diabetes indeterminada (DMI), y para clasificar a los pacientes con DM no registrada tratados con antidiabéticos.

**Resultados:** Se obtuvieron datos de un total de 376.278 sujetos con algún código ICD-10 de DM y de 8.707 pacientes sin código de DM o prediabetes tratados. Tras la aplicación de los algoritmos, un 13,9% de pacientes con DM1 se reclasificaron como DM2; un 80,9% de pacientes con DMI se reclasificaron como DM2 y un 19,1%, como prediabetes. La prevalencia global de errores de codificación (códigos múltiples o DMI) fue del 2,2%. Finalmente, el 55,2% de los sujetos con DM no registrada fueron clasificados como prediabetes, el 35,7% como DM2, el 8,5% como DMI tratada con insulina y el 0,6% como DM1.

**Conclusiones:** La prevalencia de DM inadecuadamente codificada, clasificada o infrarregistrada en asistencia primaria es relevante. La aplicación de algoritmos podría etiquetar automáticamente los casos que necesitan revisión y reducir considerablemente el riesgo de codificaciones erróneas.

© 2016 SEEN. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

## Introduction

Clinical guidelines classify diabetes mellitus based on the pathogenesis of hyperglycaemia, broadly into type 1 (T1DM; due to  $\beta$ -cell destruction), type 2 (T2DM; due to a progressive insulin secretory defect), and other types (e.g., genetic forms, gestational diabetes, drug- or chemically-induced, or other causes).<sup>1,2</sup> Moreover, guidelines give distinct recommendations for management and treatment based on classification, but they do not give clear guidance to assist physicians in classifying the condition. However, the distinction between T1DM and T2DM in clinical practice is not always obvious based on initial history, physical examination and laboratory values at first presentation. The difficulties in

the distinction between diabetes types may result in errors in patients' registries in primary care records.<sup>3</sup> Incorrect coding is categorised as misclassification (when the patient is falsely classified as a given type of DM), misdiagnosis (when the patient does not have diabetes), and miscoding (when a non-specific code – undetermined – is used and it is not possible to determine the type of diabetes).<sup>4</sup>

The prevalence of incorrect and incomplete coding and classification of diabetes in routinely collected data is difficult to quantify because of the heterogeneity among available studies, as a systematic review noted.<sup>5</sup> Still, the article highlighted the potential implications and impact, which includes inappropriate or delayed pharmacological management (e.g., prescription of insulin at T2DM onset,

or non-insulin antidiabetic drugs [NIADs] before insulin in T1DM); incorrect risk management (e.g., risk of ketoacidosis in T1DM incorrectly classified as T2DM); implications for family members (e.g., need for genetic counselling); and dubious validity and quality of care findings for researchers (e.g., inaccurate incidence and prevalence rates of DM and prescribing practices). Particularly, the inaccuracy of raw electronic data has a greater impact among patients attended in primary care, where most of them are diagnosed and managed,<sup>5,6</sup> and where incorrect coding is known to be an issue.<sup>3,7,8</sup> Finally, some patients are treated with antidiabetic drugs without any registered diabetes-related diagnosis code. This may reflect a lack of motivation, heavy workload, time constraints, or negligence in relation to a physician's responsibilities and duties with the institution in terms of a patient's health needs. These are well-known professional-related barriers to the delivery of optimal clinical practice and health care.<sup>9</sup>

Different record-based algorithms applied to raw electronically collected data have been applied and proved useful to detect and classify patients with diabetes.<sup>4,10–16</sup> The aim of the present study was to assess the quality of diabetes diagnostic data recorded electronically in primary care centres in Catalonia (Spain), and to develop and apply diagnostic algorithms to identify miscoded, misclassified and misdiagnosed diabetic records.

## Methods

### Design

This was a cross-sectional, retrospective study using an anonymised electronic general practice database (Information System for the Development of Research in Primary Care [SIDIAP]).<sup>17</sup> Briefly, the database contains data from all electronic medical records available since 2001 obtained through specific software (Electronic Clinical station in Primary Care; eCAP), and includes data from all of the 274 primary care centres of the Catalan Health Institute (ICS), which attends 80% of the total population (about 5.835 million patients) in Catalonia.

### Data extraction and variables

Data were obtained from patients who, during 2014, had a diagnosis code for diabetes mellitus by means of the International Classification of Diseases [ICD-10] codes, namely E10 (T1DM), E11 (T2DM), and E14 (undetermined DM; UDM). Moreover, we obtained data from patients receiving antidiabetic treatment but who did not have a DM or prediabetes (code R73.0; abnormal glucose) diagnosis recorded. Patients with only a code for gestational or secondary diabetes were excluded. Primary variables included were: age at time of diagnosis; HbA1c in 2014; and first HbA1c at the time of diagnosis. Secondary variables included were: age, gender, time since diagnosis; body mass index (BMI) in 2014; and BMI at the time of diagnosis. The prescribed antidiabetic treatments were extracted from prescription- and pharmacy-invoicing databases provided by the CatSalut (Catalan Health Service), which are yearly incorporated into the SIDIAP database.

Glucose lowering agents included insulin and NIADs marketed in Spain up to 2014, namely metformin, sulfonylureas, glinides, glitazones, dipeptidyl peptidase-4 inhibitors (DPP-4i), glucagon-like peptide-1 receptor agonists (GLP-1ra), alpha-glycosidase inhibitors (AGI), and sodium-glucose co-transporter 2 inhibitors (SGLT2i). Antidiabetic treatments were grouped as insulin in monotherapy, insulin plus metformin, NIADs other than metformin (alone or in combination with insulin or metformin or both), and metformin alone.

This study was approved by the Ethics Committee of the Primary Health Care University Research Institute (IDIAP) Jordi Gol.

### Decision algorithm development and output

Three different algorithms were designed and applied to sort diagnostic codes into T1DM, T2DM, prediabetes or unspecified DM treated with insulin. To establish a definite diagnosis, each algorithm started with the original code(s) recorded in the clinical history (or treatment received if no code was recorded), and further applied filters based on compatibility of prescribed antihyperglycaemic treatment, age at diagnosis, fasting plasma glucose, and HbA1c.

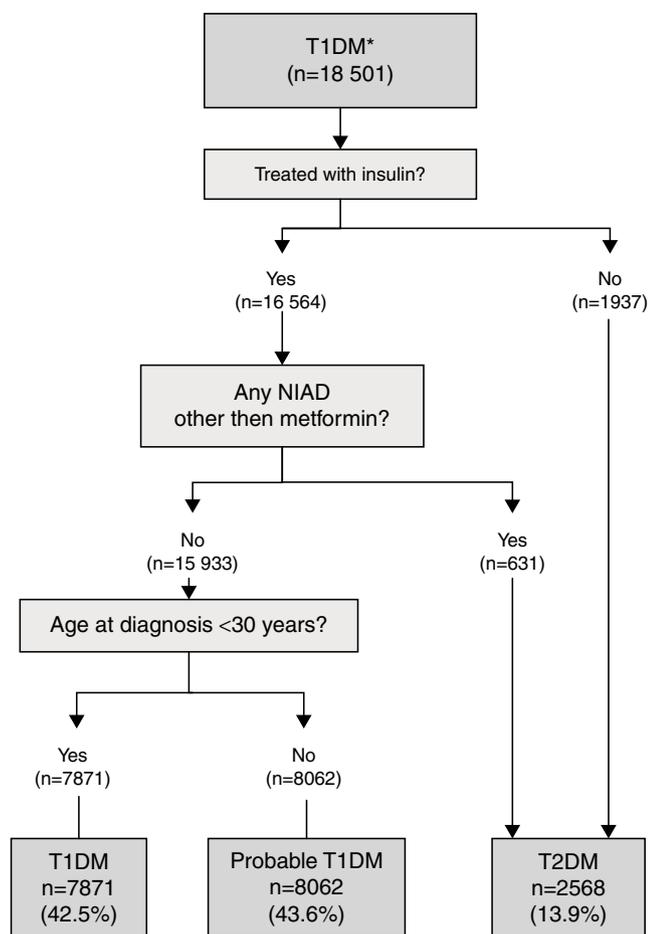
Incorrect coding was categorised as follows: (a) misclassification, when the subject was falsely classified as one type of DM (e.g., T1DM instead of T2DM); (b) misdiagnosis, when the diagnosis of diabetes was inadequate because the patient probably did not have the condition (e.g., prediabetes); (c) miscoding, when there were multiple codes or the code was non-specific (e.g., undetermined diabetes; UDM), not allowing to discriminate precisely the type of diabetes; and (d) unregistered diabetes (UnrDM), when there was no diabetes-related code recorded but the patient was being treated with antidiabetic drugs.

#### T1DM algorithm

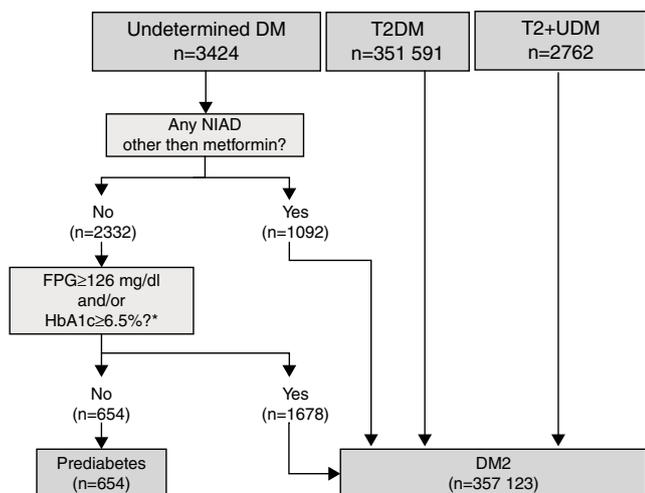
This algorithm was used for subjects with a single specific code for T1DM with and without multiple codes (miscoding): T1DM and T2DM and/or UDM (Fig. 1). Those subjects initially coded as T1DM but not treated with insulin or treated with insulin and also a NIAD other than metformin were considered misclassified and reclassified to T2DM. A definite diagnosis of T1DM was applied if the subject was treated with insulin, but not with a NIAD other than metformin (which can be prescribed with insulin in T1DM), and was younger than 30 years at the time of diagnosis. In subjects with the same pattern but older than 30 years at the time of diagnosis, T1DM was considered probable because insulin-treated T2DM patients could not be completely excluded.

#### T2DM and UDM algorithm

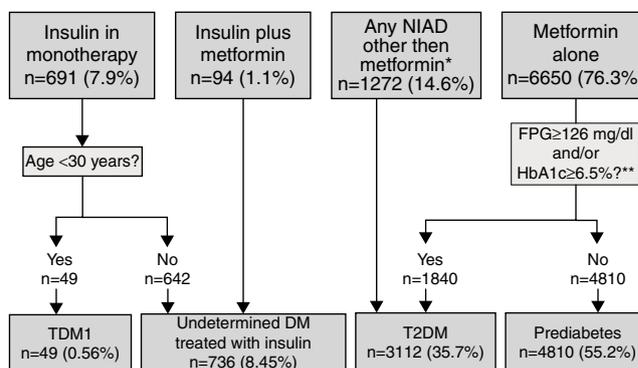
This algorithm was used for subjects with a single specific code for T2DM with and without multiple codes (T2DM plus UDM), and for subjects with a single code for UDM (Fig. 2). Within subjects coded as UDM, those not treated with a NIAD other than metformin and with glycaemic values below the threshold for diabetes diagnosis (fasting plasma glucose <126 mg/dl and/or HbA1c <6.5%) at any time (after or before the diagnosis) were considered misdiagnosed of diabetes and were classified as prediabetes. Conversely, subjects treated



**Figure 1** Algorithm for the validation of T1DM diagnosis. *Footnote:* NIADs, non-insulin antidiabetic drugs; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.



**Figure 2** Algorithm for the validation of T2DM diagnosis. *Footnotes:* BG, basal plasma glucose; DM, diabetes mellitus; NIAD, non-insulin antidiabetic drugs; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus. \*At any time (before or after onset).



**Figure 3** Algorithm for the detection of unregistered diabetes in patients without a code for diabetes or prediabetes who were treated with antidiabetic drugs. *Footnotes:* BG, basal plasma glucose; DM, diabetes mellitus; NIAD, non-insulin antidiabetic drugs; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus. \*Includes 904 patients on only NIADs other than metformin, 320 in combination with metformin, 25 in combination with insulin, and 23 on triple therapy. \*\* At any time (before or after onset).

with a NIAD other than metformin and those treated only with metformin but fulfilling diabetes diagnostic criteria values at any time were reclassified as T2DM. All subjects with a T2DM or T2DM + UDM code were considered as correct T2DM diagnosis.

**Pharmacologically treated algorithm for subjects without a DM or prediabetes code (unregistered DM) taking antidiabetics**

This algorithm considered that subjects on insulin as monotherapy and younger than 30 years at the time of diagnosis were T1DM cases (Fig. 3). Those on insulin as monotherapy who were older than 30 years at the time of diagnosis or those on insulin combined with metformin could be either T1DM or T2DM, and were therefore classified as UDM treated with insulin. Subjects treated with a NIAD other than metformin (alone or in combination with insulin or metformin or both) and subjects on metformin alone or who fulfilled diabetes diagnostic criteria at any time were classified as T2DM, while subjects on metformin alone with glycaemic values always below the threshold for diabetes diagnosis were categorised as prediabetes.

**Statistical analysis**

Demographic and clinical characteristics are presented as mean and standard deviation (SD) for continuous variables, and percentages for categorical variables. The Chi-squared test and *t*-test were used to compare qualitative and quantitative variables, respectively, with Bonferroni correction as a post hoc test to determine the significance of differences between groups at a level of  $\alpha=0.05$  for all tests. Statistical analyses were performed using the Stata Statistical Software, release 11 (StataCorp, College Station, TX, USA). Statistical significance was set at two-tailed  $p < 0.05$ .

**Table 1** Demographic and clinical characteristics of 376,278 included patients who had a DM code recorded.

	T1DM N = 18,501 <sup>a</sup>	T2DM N = 356,173	UDM N = 3424
<i>Characteristic</i>			
Gender, male, n (%)	10,535 (56.9)	194,224 (54.9)	1744 (50.9)
Age in 2014, years, mean (SD)	44.8 (18.2)	69.7 (12.3)	71.2 (13.5)
Age at diagnosis, years, mean (SD)	32.4 (18.2)	60.81 (12.1)	58.9 (13.1)
Age <30 years at diagnosis, n (%)	9105 (49.2)	2800 (0.8)	92 (2.7)
Duration of diabetes, years, mean (SD)	12.3 (9.5) <sup>†</sup>	8.9 (6.0)	12.3 (5.5) <sup>†</sup>
HbA1c in 2014, %, mean (SD)	8.1 (1.5) (n = 8279)	7.1 (1.2) <sup>†</sup> (n = 272,641)	7.1 (1.3) <sup>†</sup> (n = 2223)
HbA1c at diagnosis, %, mean (SD)	9.1 (2.6) (n = 1948)	7.1 (1.7) <sup>†</sup> (n = 106,034)	6.9 (1.9) <sup>†</sup> (n = 319)
BMI in 2014, kg/m <sup>2</sup> , mean (SD)	26.5 (5.3) (n = 9100)	30.2 (5.2) (n = 235,629)	29.7 (5.2) (n = 2155)
Obesity (BMI ≥30 kg/m <sup>2</sup> ) in 2014, n (%)	2019 (22.4) (n = 9100)	117,615 (49.9) (n = 235,629)	926 (43.0) (n = 2155)
BMI at diagnosis, kg/m <sup>2</sup> , mean (SD)	24.6 (6.3) (n = 3083)	31.5 (5.3) (n = 113,886)	30.8 (5.3) (n = 780)
Obesity (BMI ≥30 kg/m <sup>2</sup> ) at diagnosis, n (%)	578 (18.7) (n = 3083)	64,837 (56.9) (n = 113,886)	396 (50.8) (n = 780)
<i>Antidiabetic treatment (N = 384,985)</i>			
No pharmacologic treatment, n (%)	1346 (7.3)	62,929 (17.7)	834 (24.4)
Metformin, n (%)	298 (1.6)	125,623 (35.3)	822 (24.0)
Any NIAD other than metformin, n (%)	87 (0.5)	20,993 (5.9) <sup>†</sup>	186 (5.4) <sup>†</sup>
Metformin + other NIADs, n (%)	206 (1.1)	71,962 (20.2) <sup>†</sup>	656 (19.2) <sup>†</sup>
Insulin as monotherapy, n (%)	14,129 (76.4)	19,982 (5.6)	314 (9.2)
Insulin + metformin, n (%)	1804 (9.8)	31,395 (8.8)	362 (10.6)
Insulin + any NIAD other than metformin, n (%)	250 (1.4)	7134 (2.0) <sup>†</sup>	81 (2.4) <sup>†</sup>
Insulin + metformin + other NIADs, n (%)	381 (2.1)	16,155 (4.5) <sup>†</sup>	169 (4.9) <sup>†</sup>

<sup>†</sup> Not statistically significant differences ( $p > 0.05$ ) between the means or percentages in the same row-Bonferroni correction was used to perform the multiple comparisons test. BMI, body mass index, NIAD, non-insulin antidiabetic drug, SD, standard deviation, T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; UDM, undetermined diabetes mellitus.

<sup>a</sup> The sample number is given between parenthesis and italics if different from the total population corresponding population.

## Results

Out of the initial 5,432,691 patients' records, we retrieved data from a total of 376,278 subjects with an ICD-10 code for DM (Table 1): 18,501 (4.9%) had at least one code for T1DM, of which 16,414 were single codes; 356,173 (94.7%) had at least one code for T2DM, of which 351,591 were single codes; 3424 (0.9%) had a single code for UDM. Overall, 2.2% of subjects were potential miscoding errors if we take into account 4849 patients with multiple codes and 3424 with UDM code. Patients in the UDM group were older than in the T1DM and T2DM groups (71.2 years vs. 44.8 and 69.7 years, respectively;  $p < 0.001$ ), and there was a lower proportion of males in the UDM group than in the other groups (50.5% vs. 56.9% and 54.9% in the T1DM and T2DM groups, respectively;  $p < 0.001$ ). The most frequent treatments were insulin alone in the T1DM group (76.4% of cases), metformin alone in the T2DM group (35.3%), and no pharmacological treatment or metformin alone in the UDM group (24% in both cases). Subsequently, we retrieved data from a total of 8707 patients who did not have any DM or prediabetes code recorded but were treated with antidiabetic drugs (UnrDM; Table 2): metformin alone was the most frequent treatment ( $n = 6650$ ;

76.3%), followed by any NIAD other than metformin (alone or in association with metformin or insulin) ( $n = 1272$ ; 14.6%), insulin as monotherapy ( $n = 691$ ; 7.9%), and insulin with metformin ( $n = 94$ ; 1.1%).

## Algorithm application for T1DM, T2DM and UDM codes

When the 18,501 patients initially having at least one code for T1DM were categorised based on the proposed TD1M algorithm (Fig. 1), 42.5% ( $n = 7871$ ) of the cases were classified as definite T1DM. Moreover, 43.6% ( $n = 8062$ ) of cases were on insulin, not taking any NIADs other than metformin and were >30 years at the time of diagnosis, and were considered as a probable T1DM since it cannot be discarded that the group included insulin-treated T2DM patients. Finally, 13.9% of cases ( $n = 2568$ ) were found to be misclassified because they were not on insulin or they were on insulin in combination with NIADs other than metformin, and were thus reclassified as T2DM.

The proposed T2DM algorithm included 351,591 patients with only a code for T2DM, 3424 subjects with a single code

**Table 2** Demographic and clinical characteristics of 8707 patients who were on an antidiabetic treatment but no code for DM or prediabetes recorded (unregistered diabetes).

	Metformin alone <i>N</i> = 6650 <sup>a</sup>	Any NIAD other than metformin <sup>b</sup> <i>N</i> = 1272	Insulin plus metformin <i>N</i> = 94	Insulin alone <i>N</i> = 691
Gender, male, <i>n</i> (%)	2838 (42.7)	532 (41.8)	42 (44.7)	182 (26.3)*
Age in 2014, years, mean (SD)	60.1 (7.7)	65.7 (16.9)	61.4 (16.8)	50.0 (19.8)*
Age <30 in 2014, years, <i>n</i> %	472 (7.1)*	36 (2.8)	2 (2.1)	49 (7.1)*
HbA1c in 2014, %, mean (SD)	6.2 (0.8) ( <i>n</i> = 3042)	6.5 (1.5) ( <i>n</i> = 464)	7.1 (1.9)* ( <i>n</i> = 46)	6.2 (1.3) ( <i>n</i> = 200)
HbA1c, ≤7%, <i>n</i> (%)	2747 (90.3)*	358 (77.2)	28 (60.9)*	157 (78.5)
HbA1c 7.1–8%, <i>n</i> (%)	182 (6.0)*	48 (10.3)	8 (17.4)*	20 (10.0)
HbA1c 8.1–10%, <i>n</i> (%)	87 (2.9)*	37 (8.0)	6 (13.0)*	19 (9.5)
HbA1c >10%, <i>n</i> (%)	26 (0.9)*	21 (4.5)	4 (8.7)	4 (2.0)
BMI in 2014, kg/m <sup>2</sup> , mean (SD)	32.0 (6.4)* ( <i>n</i> = 3384)	30.4 (5.9) ( <i>n</i> = 618)	30.6 (7.1) ( <i>n</i> = 42)	28.8 (5.1)* ( <i>n</i> = 368)
Normal weight (BMI <25 kg/m <sup>2</sup> ), <i>n</i> (%)	374 (11.1)*	116 (18.8)	10 (23.8)	92 (25.0)*
Overweight (BMI 25.0–29.9 kg/m <sup>2</sup> ), <i>n</i> (%)	1012 (29.9)	196 (31.7)	13 (31.0)	134 (36.4)*
Obesity (BMI ≥30 kg/m <sup>2</sup> ), <i>n</i> (%)	1998 (59.0)*	306 (49.5)	19 (45.2)	142 (38.6)*

BMI, body mass index, NIAD, non-insulin antidiabetic drug, SD, standard deviation.

<sup>a</sup> The sample number is given between parenthesis and italics if different from the corresponding total population.

<sup>b</sup> Includes 904 patients on only NIADs other than metformin, 320 in combination with metformin, 25 in combination with insulin, and 23 on triple therapy.

\* Statistically significant differences ( $p < 0.05$ ) in comparison to the remaining results in the same row. Bonferroni correction was used to perform the multiple comparisons test.

for UDM, and 2762 patients with both codes (Fig. 2). After applying the algorithm to subjects with a single code of UDM diabetes, 80.9% of them ( $n = 2770$ ) were classified as T2DM because they were treated with a NIAD other than metformin (31.9% of cases), or because although they were treated only with metformin, their glycaemic values were above the threshold for diabetes diagnosis (fasting plasma glucose  $\geq 126$  mg/dl and/or HbA1c  $\geq 6.5\%$ ; 49% of cases). Finally, 19.1% of cases were considered misdiagnosed because they were not pharmacologically treated and their glycaemic values were always under the diagnosis threshold for diabetes, and were therefore reclassified as prediabetes.

### Algorithm application for subjects pharmacologically treated without a DM or prediabetes code (UnrDM)

Among patients on insulin as monotherapy ( $n = 691$ ), 49 of them (7%) were younger than 30 years at the time of diagnosis, and were considered as a definite T1DM diagnosis, but the remaining ones (>30 years at the time of diagnosis), together with subjects on insulin and metformin, could also be patients with T2DM, and were therefore classified as UDM treated with insulin (Fig. 3). Among 6650 subjects treated with metformin alone, 72.3% of them did not have abnormal glycaemic values, and were thus classified as prediabetes. Finally, a total of 3112 patients were classified as T2DM, and these included 27.7% of patients on metformin alone but with high glycaemic values ( $n = 1840$ ), and 1272 patients treated with NIADs other than metformin (904 patients on only NIADs other than metformin, 320 in combination with

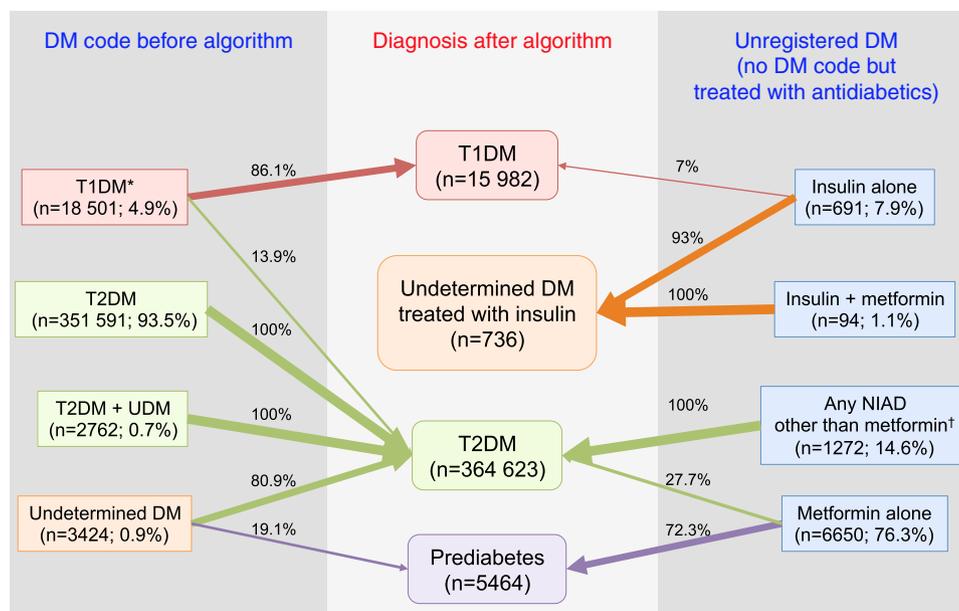
metformin, 25 with insulin, and 23 patients on triple therapy).

In summary, after the application of the designed algorithms to all subsets, 13.9% of patients with T1DM were found to be misclassified and to actually be T2DM cases, 80.9% of patients with UDM were reclassified as T2DM, and 19.1% of them were classified as probably having prediabetes (Fig. 4). Overall, the prevalence of miscoding because of patients with multiple codes or initially coded as UDM in our sample was 2.2%. Finally, 55.2% of subjects with UnrDM (i.e., lacking a DM code but pharmacologically treated with antidiabetic drugs) could be classified as prediabetes, 35.7% as T2DM, 0.6% as T1DM, and in 8.5% of cases the DM type could not be discriminated and were coded as UDM treated with insulin.

### Discussion

This study showed the application of automated algorithms to flag and reclassify cases inappropriately coded in electronic records from primary care. The overall prevalence of coding errors in our database was 5.2%. The most common errors were under-registration (no DM-related code but treated with antidiabetic drugs; 43.1% of all errors) and miscoding (i.e., multiple DM codes or UDM code; 41% of all errors). Misclassification was less frequent (i.e., T1DM instead of T2DM; 12.7% of all errors), and misdiagnosis (i.e., UDM probably prediabetes) was the least frequent error (2.2% of them).

Clinical features that have been shown to be the most discriminatory to differentiate T1DM and T2DM are age at diagnosis and time to insulin, while BMI, usually used in



**Figure 4** Distribution of diagnoses before and after the application of the algorithms. *Footnotes:* DM, diabetes mellitus; NIAD, non-insulin antidiabetic drugs; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus. \*Includes codes for only T1DM, and multiple codes: T1DM + undetermined DM, T1DM + T2DM, and T1DM + T2DM + undetermined DM. † Includes 904 patients on only NIADs other than metformin, 320 in combination with metformin, 25 in combination with insulin, and 23 on triple therapy.

clinical practice as a surrogate marker, does not seem to add much to these 2 other criteria.<sup>18</sup> Moreover, the use of algorithms that include surrogate markers to confirm or refute the diagnosis of diabetes, and to help differentiate between diabetes types has been proved useful when used with data from electronic records in primary care.<sup>19,20</sup> One study reported that HbA1c levels >6.5% have a high positive predictive value (PPV) to confirm a DM diagnosis; the combination of age <30 years, insulin treatment, and BMI <30 kg/m<sup>2</sup> had a high PPV for T1DM; and age >45 years, BMI >30 kg/m<sup>2</sup> and hypertension a high PPV for T2DM.<sup>19</sup> Another study, conducted in insulin-treated diabetic patients, reported that the most discriminatory variables to predict diabetes type were time from diagnosis to insulin treatment (optimal cut-off 12 months), and age at diagnosis (optimal cut-off ≤39 years), while BMI (optimal cut-off ≤23.1 kg/m<sup>2</sup>) was less discriminatory.<sup>20</sup> Following the same scheme, the discriminatory variables that we included in our algorithms were therapy, age at onset, and glycaemic levels, but not BMI.

The prevalence of T2DM incorrectly labelled as T1DM (misclassified diabetes) in our study was 13.9%. This is in line with the previous 10–30% prevalence of misclassifications reported in the UK primary care setting through the use of algorithms,<sup>4,10–13,16</sup> and also in line with the 22.7% reported in a Medicaid paediatric T1DM population misclassified as T2DM in the US, predominantly in primary care settings.<sup>14</sup> Contributing factors to this high rate of T1DM misclassification may be the increasing prevalence of obesity in children and adolescents, and the decreasing age of onset for T2DM, which during the last decades tends to overlap with age at onset of T1DM, all together blurring the differential diagnosis.<sup>21,22</sup> From previous studies, people with T2DM misclassified as T1DM tend to be older than those correctly classified, have lower HbA1c levels, are more likely to

achieve good glycaemic control,<sup>12,16</sup> and have a higher BMI at diagnosis.<sup>20</sup>

The rate of misdiagnosed patients in our study, namely subjects coded as UDM but actually more likely to have prediabetes than T2DM was 19.1%, which is line within the wide range of observed misdiagnoses reported in other studies (6.1–34.5%).<sup>4,10,11,13,16</sup> Misdiagnosed patients have been reported to have similar age or BMI at diagnosis than actual T2DM patients, but tend to have lower HbA1c values at diagnosis, levels that remain lower over time, and thus they essentially mimic good glycaemic control.<sup>11,13</sup> Moreover, we found that 72.3% of subjects without a code for DM but treated with metformin alone were reclassified as prediabetes, which is of concern and emphasises the fact that those subjects might be receiving unnecessary medication, and although they should be closely followed in case of changes in their glycaemic levels, they could be managed with lifestyle modification as the first option.<sup>23</sup> In fact, neither metformin nor other antidiabetic drugs are currently authorised for the treatment of prediabetes in Europe, so its use in these patients is off-label.

The prevalence of miscoded diabetes (i.e., patients with multiple codes or UDM) was 2.2% in our study, which is lower than the rates observed in other studies, ranging from 4.9% to 10%,<sup>4,10,11,13</sup> and much lower than the 47.8% reported in another study.<sup>16</sup> However, if we also consider subjects uncoded for DM but treated with antidiabetics because they are not registered as having the disease,<sup>16</sup> then the rate of miscoding in our study would rise up to 4.4%. It has been reported that patients with vague or non-specific codes have poorer glycaemic control,<sup>12,16</sup> which may indicate that they are at risk of suboptimal management because of the lack of a clear identification that would otherwise allow the adoption of specific clinical guidelines or recommendations for their DM type. Finally, in 8.5% of patients without a code

for DM but treated with insulin alone or insulin with NIADS and who were older than 30 years, it was not possible to discriminate between T1DM and T2DM, and patients were labelled as UDM treated with insulin. This group may include real T1DM patients, insulin-treated T2DM patients, and even latent autoimmune diabetes of the adult (LADA), although for the latter type there are no current specific international guidelines for management and treatment. Time to insulin treatment since diagnosis would have been helpful in order to discriminate between DM types, but unfortunately the SIDIAP database started in 2006 and this criterion could not be included in the algorithms for people diagnosed before this year. Some have suggested that, when facing diagnostic ambiguity, the diagnosis should be reviewed, the patient should be referred to a specialist, or complementary tests such as measurement of insulin secretion should be performed (C-peptide and islet autoantibody levels).<sup>20</sup> However, these are expensive techniques and not routinely performed in some primary care settings.

The main advantage of the present study is the use of a large database of computerised data collected automatically from primary care centres representative of the health care practices in our country. The main limitation is the data source itself, which is subject to errors in data recording, missing values, or incomplete clinical histories. For instance, the number of patients with recorded HbA1c values was in some cases low. Since this variable was used to establish a definite diagnosis in the algorithms for T2DM and/or UDM codes and unregistered diabetes, we cannot discard that some patients with diabetes have been actually classified as prediabetes. Moreover, and as previously pointed out, the algorithms rely upon clinicians coding and prescribing decisions, and it cannot minimise the future risk of incorrect coding if a wrong diagnosis is made and the patient is treated accordingly.<sup>15</sup> Conversely, we cannot discard that in some cases of misclassification (T1DM instead of T2DM for patients treated only with insulin) both the patient and the physician know the correct diagnosis and treatment is actually appropriate. In addition, we did not consider the possibility that patients coded as T2DM or T2DM + UDM could be misclassified. This is because more than 95% of diabetic patients are T2DM, and the objective of the algorithms is to help clinicians classify infrequent cases, namely a T1DM code, a doubtful code (i.e., only UDM code), multiple codes or antidiabetic treatment without a diabetes or prediabetes code. Finally, we did not explore the presence of false negative cases, namely subjects likely to have undiagnosed diabetes based on existing altered blood glucose recordings, and estimated to be around 1% in studies applying algorithms to primary care electronic registries in the US and the UK,<sup>24,25</sup> or subjects who do not have a diabetes diagnostic code nor diabetes-related treatment recorded. Therefore, it is probable that the overall prevalence of coding errors in our study was underestimated.

In conclusion, we detected a high prevalence of problems with misclassification, misdiagnosis, miscoding and under-registration of DM in primary care in Catalonia. This may be in part due to patients who were difficult to characterise at the time of diagnosis, but it also points to a lack of accuracy in distinguishing diabetes from no diabetes, and incomplete coding when only treatment but not disease is recorded. This has consequences at the level of quality of

disease management, but also at the epidemiological level when the database is used for research on the epidemiology and natural history of the disease. The embedment of algorithms in the primary care centre's computer systems could automatically flag cases that would need review because of potential miscoding (i.e., absent, vague, unspecific or multiple DM codes), misclassification (T2DM instead of T1DM or vice versa), under-registration (patients treated with antidiabetic drugs but without diagnosis) and misdiagnosis (no current criteria for DM), and largely reduce the risk of inappropriate coding.

## Authors' contributions

MM-C and DM wrote the manuscript and contributed equally to this study; JR managed the database, performed the statistical analyses and contributed to the discussion; and JF-N, MM-C, BB and DM designed and conducted the study, reviewed/edited the manuscript and contributed to the discussion. MM-C had full access to all data in the study and takes responsibility for the integrity of data and the accuracy of the data analysis.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

We acknowledge Mònica Gratacòs and Amanda Prowse for providing support in the manuscript preparation and editing. CIBER of Diabetes and Associated Metabolic Diseases (CIBER-DEM) is an initiative included in Plan Nacional de I + D + I and cofinanced by Instituto de Salud Carlos III, Subdirección General de Evaluación and Fondo Europeo de Desarrollo Regional (FEDER).

## References

1. Association AD. 2. Classification and diagnosis of diabetes. *Diabetes Care*. 2016;39:S13–22.
2. World Health Organisation. Diabetes fact sheet. Updated March 2016. Available at: <http://www.who.int/mediacentre/factsheets/fs312/en/> [accessed May 2016].
3. Rollason W, Khunti K, de Lusignan S. Variation in the recording of diabetes diagnostic data in primary care computer systems: implications for the quality of care. *Inform Prim Care*. 2009;17:113–9.
4. de Lusignan S, Khunti K, Belsey J, Hattersley A, van Vlymen J, Gallagher H, et al. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. *Diabetes Med*. 2010;27:203–9.
5. Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT, Khunti K. Incorrect and incomplete coding and classification of diabetes: a systematic review. *Diabetes Med*. 2010;27:491–7.
6. Khunti K, Ganguli S. Who looks after people with diabetes: primary or secondary care. *J R Soc Med*. 2000;93:183–6.
7. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam Pract*. 2006;23:253–63.

8. Shephard E, Stapley S, Hamilton W. The use of electronic databases in primary care research. *Fam Pract*. 2011;28:352–4.
9. Cochrane LJ, Olson CA, Murray S, Dupuis M, Tooman T, Hayes S. Gaps between knowing and doing: understanding and assessing the barriers to optimal health care. *J Contin Educ Health Prof*. 2007;27:94–102.
10. Royal College of General Practitioners. NHS Diabetes. Coding, classification and diagnosis of diabetes. A review of the coding, classification and diagnosis of diabetes in primary care in England with recommendations for improvement. 201. Available at: [http://www.sdrn.org.uk/sites/sdrn.org.uk/files/nhs\\_diagnosis\\_classification\\_report.pdf](http://www.sdrn.org.uk/sites/sdrn.org.uk/files/nhs_diagnosis_classification_report.pdf) [accessed May 2016].
11. de Lusignan S, Sadek N, Mulnier H, Tahir A, Russell-Jones D, Khunti K. Miscoding, misclassification and misdiagnosis of diabetes in primary care. *Diabetes Med*. 2012;29:181–9.
12. Sadek AR, van Vlymen J, Khunti K, de Lusignan S. Automated identification of miscoded and misclassified cases of diabetes from computer records. *Diabetes Med*. 2012;29:410–4.
13. Hassan Sadek N, Sadek AR, Tahir A, Khunti K, Desombre T, de Lusignan S. Evaluating tools to support a new practical classification of diabetes: excellent control may represent misdiagnosis and omission from disease registers is associated with worse control. *Int J Clin Pract*. 2012;66:874–82.
14. Tripathi A, Rizvi AA, Knight LM, Jerrell JM. Prevalence and impact of initial misclassification of pediatric type 1 diabetes mellitus. *South Med J*. 2012;105:513–7.
15. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care*. 2013;36:914–21.
16. Seidu S, Davies MJ, Mostafa S, de Lusignan S, Khunti K. Prevalence and characteristics in coding, classification and diagnosis of diabetes in primary care. *Postgrad Med J*. 2014;90:13–7.
17. Bolibar B, Fina Aviles F, Morros R, et al. SIDIAP database: electronic clinical records in primary care as a source of information for epidemiologic research. *Med Clin (Barc)*. 2012;138:617–21.
18. Shields BM, Peters JL, Cooper C, et al. Can clinical features be used to differentiate type 1 from type 2 diabetes? A systematic review of the literature. *BMJ Open*. 2015;5:e009088.
19. Bagheri A, Sadek A, Chan T, Khunti K, de Lusignan S. Using surrogate markers in primary electronic patient record systems to confirm or refute the diagnosis of diabetes. *Inform Prim Care*. 2009;17:121–9.
20. Hope SV, Wienand-Barnett S, Shepherd M, et al. Practical Classification Guidelines for Diabetes in patients treated with insulin: a cross-sectional study of the accuracy of diabetes diagnosis. *Br J Gen Pract*. 2016;66:e315–22.
21. Kaufman FR, Shaw J. Type 2 diabetes in youth: rates, antecedents, treatment, problems and prevention. *Pediatr Diabetes*. 2007;8 Suppl. 9:4–6.
22. Jones KL. Role of obesity in complicating and confusing the diagnosis and treatment of diabetes in children. *Pediatrics*. 2008;121:361–8.
23. Mata-Cases M, Artola S, Escalada J, et al. Consensus on the detection and management of prediabetes. Consensus and Clinical Guidelines Working Group of the Spanish Diabetes Society. *Endocrinol Nutr*. 2015;62:e23–36.
24. Holt TA, Gunnarsson CL, Cload PA, Ross SD. Identification of undiagnosed diabetes and quality of diabetes care in the United States: cross-sectional study of 11.5 million primary care electronic records. *CMAJ Open*. 2014;2:E248–55.
25. Holt TA, Stables D, Hippisley-Cox J, O'Hanlon S, Majeed A. Identifying undiagnosed diabetes: cross-sectional survey of 3.6 million patients' electronic records. *Br J Gen Pract*. 2008;58:192–6.