

# Cross-national health care database utilization between Spain and France: results from the EPICHRONIC study assessing the prevalence of type 2 diabetes mellitus

Guillaume Moulis,<sup>1-3,\*</sup>  
Berta Ibañez,<sup>4-6,\*</sup> Aurore  
Palmaro,<sup>2,3</sup> Felipe Aizpuru,<sup>6-8</sup>  
Eduardo Millan,<sup>6,8</sup> Maryse  
Lapeyre-Mestre,<sup>2,3,9</sup> Laurent  
Sailer,<sup>1-3</sup> Koldo Cambra<sup>5,6,10</sup>

<sup>1</sup>Department of Internal Medicine, Toulouse University Hospital, Toulouse, France; <sup>2</sup>UMR 1027 INSERM, University of Toulouse, Toulouse, France; <sup>3</sup>Clinical Investigation Center 1436, Toulouse University Hospital, Toulouse, France; <sup>4</sup>Navarrabiomed, Health Department, Public University of Navarra, Pamplona, Spain; <sup>5</sup>IdiSNA, Pamplona, Spain; <sup>6</sup>Health Service Research on Chronic Patients Network (REDISSEC), Pamplona, Spain; <sup>7</sup>Research Unit Araba (BioAraba), Osakidetza-Basque Health Department, Vitoria-Gasteiz, Spain; <sup>8</sup>Healthcare Services Sub-directorate, Osakidetza-Basque Health Service, Araba, Spain; <sup>9</sup>Department of Medical and Clinical Pharmacology, Toulouse University Hospital, Toulouse, France; <sup>10</sup>Institute of Public Health and Labour Health of Navarra, Pamplona, Spain

\*These authors contributed equally to this work

**Aim:** The EPICHRONIC (EPIdemiology of CHRONIC diseases) project investigated the possibility of developing common procedures for French and Spanish electronic health care databases to enable large-scale pharmacoepidemiological studies on chronic diseases. A feasibility study assessed the prevalence of type 2 diabetes mellitus (T2DM) in Navarre and the Basque Country (Spain) and the Midi-Pyrénées region (France).

**Patients and methods:** We described and compared database structures and the availability of hospital, outpatient, and drug-dispensing data from 5.9 million inhabitants. Due to differences in database structures and recorded data, we could not develop a common procedure to estimate T2DM prevalence, but identified an algorithm specific to each database. Patients were identified using primary care diagnosis codes previously validated in Spanish databases and a combination of primary care diagnosis codes, hospital diagnosis codes, and data on exposure to oral antidiabetic drugs from the French database.

**Results:** Spanish and French databases (the latter termed Système National d'Information Inter-Régimes de l'Assurance Maladie [SNIIRAM]) included demographic, primary care diagnoses, hospital diagnoses, and outpatient drug-dispensing data. Diagnoses were encoded using the International Classification of Primary Care (version 2) and the International Classification of Diseases, version 9 and version 10 (ICD-9 and ICD-10) in the Spanish databases, whereas the SNIIRAM contained ICD-10 codes. All data were anonymized before transferring to researchers. T2DM prevalence in the population over 20 years was estimated to be 6.6–7.0% in the Spanish regions and 6.3% in the Midi-Pyrénées region with ~2% higher estimates for males in the three regions.

**Conclusion:** Tailored procedures can be designed to estimate the prevalence of T2DM in population-based studies from Spanish and French electronic health care records.

**Keywords:** epidemiology, pharmacoepidemiology, electronic health care database, cross-national study, population-based study, type 2 diabetes mellitus

## Introduction

Cross-national studies that use health care databases can be useful to compare the epidemiology of diseases and drug exposures between countries. Recently, some projects have compared national databases to identify possible common extraction models. These projects included North European databases<sup>1</sup> and European prescription databases (Pharmacoepidemiological Research on Outcomes and Therapeutics by a European Consortium [PROTECT] project).<sup>2-4</sup> This latter project demonstrated the feasibility of assessing drug exposure and pharmacovigilance signals in various national databases. Another example is the European Collaboration for Healthcare

Correspondence: Guillaume Moulis  
INSERM UMR 1027,  
Pharmacoepidemiology Unit, 37 allées  
Jules Guesde, 31000 Toulouse, France  
Tel +33 5 6114 5606  
Fax: +33 5 6114 5928  
Email moulis.g@chu-toulouse.fr

Optimization (ECHO) project,<sup>5</sup> a European Health Services Research Program that analyzed unwarranted variations in medical practice and health care outcomes, and was based on hospital databases from several European countries.

Nevertheless, health care systems and databases are often diverse and fragmented.<sup>6</sup> Hence, cross-national research initiatives are still scarce and often include only some of the available health information. Initiatives that build infrastructure for the efficient reuse of health care data for epidemiological research have been launched in recent years, such as the European Medical Information Framework (EMIF) project. This project currently collects information on around 52 million European citizens from diverse data sources from regions within six European countries (including hospital data from the Barcelona region, but no French data).<sup>7</sup>

The use of health administrative data within each European country has a longer history than cross-national data and has become common practice in public health research in both Spain and France. In Spain, one of the most widely used databases is the Minimum Basic Data Set (MBDS) with information on all hospital discharges in the National Health System.<sup>8</sup> Other important resources used in epidemiological studies in Spain are the Primary Care Electronic Medical Record System,<sup>9</sup> the Database for Pharmacoepidemiologic Research in Primary Care (Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria; BIFAP) database for pharmacovigilance studies,<sup>10</sup> the mortality register,<sup>11</sup> and the population directory database.<sup>12</sup> In France, most epidemiology or pharmacoepidemiology studies are conducted within the National Inter-Scheme Health Insurance database (Système National d'Information Inter-Régimes de l'Assurance Maladie; SNIIRAM), which links outpatient, hospital, and civil status data for the entire French population (67 million inhabitants).<sup>13,14</sup> The structure of the SNIIRAM and its data are described in the following sections. The general beneficiary sample (Echantillon Généraliste des Bénéficiaires; EGB) is a 1/97 sample from the SNIIRAM (~600,000 individuals) and is mostly used to assess frequent conditions.<sup>13,14</sup>

The EPICHRONIC (EPIde miology of CHRONIC diseases) study was part of the REFBIO project, a trans-Pyrenean cooperation network for biomedical research created to promote competitive health research.<sup>15</sup> The aim of the EPICHRONIC study was to assess the possibility of developing common procedures in French and Spanish electronic health care databases and to carry out large-scale epidemiologic studies on chronic diseases. The specific objectives were to compare the data available in these databases and to conduct a feasibility study to assess the prevalence of type 2 diabetes

mellitus in the early 2010s within the databases from three regions located on both sides of the Pyrenees.

## Patients and methods

### Design and setting

This population-based cross-sectional study was conducted in the Midi-Pyrénées region (southern France, 3 million inhabitants) and two autonomous communities located in northern Spain: the Basque Country (2.2 million inhabitants) and Navarre (0.7 million inhabitants).

For the French part of the study, we used the SNIIRAM database restricted to Midi-Pyrénées inhabitants. In France, the national insurance system covers all citizens. Hospitalizations are reimbursed at about 80% of their costs, whereas outpatient costs, including drugs, vary between 15% and 70% of the total. The remaining are supported by mutual funds. However, there is a list of long-term diseases (LTDs) that allow full reimbursement of costs related to these conditions, and low-income patients benefit from full insurance cover without an advance payment.<sup>16</sup>

The Regional Health Systems of the Basque Country and Navarre are part of the Spanish National Health System (SNHS), which is a quasi-federal decentralized system where the regional governments of the 17 autonomous communities have full responsibility for policy making, planning, and financing at a regional level. In turn, each region is organized into basic health zones, the locus for primary care provision. Navarre consists of 57 basic health zones and the Basque Country consists of 139, each of which has a team of general practitioners, nurses, pediatricians, and other health care workers. The health care coverage provided by the SNHS is practically universal and is financed by general taxes and health services such as hospitalization, and diagnostic procedures are free of charge for all citizens. A fraction of medication costs are paid by patients within a cost-sharing scheme that is based on their employment and income status, established in Spain since July 2012.<sup>17</sup> In the Spanish regions, all patients with type 2 diabetes mellitus are managed by primary care teams, and their data are recorded in the Primary Care Electronic Medical Record System, named Atenea in Navarre and Osabide-AP in the Basque Country.

### Database comparison

We described and compared Spanish and French databases from different perspectives. First, we described and compared the database structures of all hospital, outpatient, and drug-dispensing data. Second, we described and compared the legislative regulation regarding access and linkage of these databases. Results of this comparative study were used to

define the method used to identify type 2 diabetes mellitus in these databases.

## Assessment of the prevalence of type 2 diabetes mellitus

Due to differences between databases, we could not develop a common identification process but constructed specific algorithms for each database to identify cases of type 2 diabetes mellitus.

Patients were considered to have type 2 diabetes mellitus in Navarre if, at the time of data extraction on May 15, 2014, the T90 code of the International Classification of Primary Care, version 2 (ICPC-2) was stated in the Atenea records. In the Basque Country, patients were considered to have this disease if, at the time of data extraction on February 12, 2015 (Osabide-AP), they had a diagnosis with a code beginning with 250 from the International Classification of Diseases, version 9, clinical modification (ICD-9CM) in Osabide-AP, after excluding patients with any code relating to type 1 diabetes mellitus (250.01; 250.03; 250.11; 250.13; 250.21; 250.23; 250.31; 250.33; 250.41; 250.43; 250.51; 250.53; 250.61; 250.63; 250.71; 250.73; 250.81; 250.83; 250.91; 250.93). A recent study on type 2 diabetes mellitus ICPC-2 codes in Navarre showed a sensitivity of 98%, a specificity of 99%, and a positive predictive value of 92%.<sup>18</sup>

In France, the 2011–2013 SNIIRAM data restricted to the Midi-Pyrénées region were used to estimate the prevalence of type 2 diabetes mellitus in 2012. In the SNIIRAM, there was no specific data relating to primary care diagnoses. Only the LTD diagnoses on full expenditure reimbursement are reported by the general practitioners and secondarily encoded by the health insurance physicians using the ICD-10.<sup>13,19</sup> Type 2 diabetes mellitus is one of these conditions (ICD-10 E11 code). However, LTDs are often under-recorded in practice.<sup>13,14</sup> Because all patients with type 2 diabetes mellitus cannot be identified using the corresponding LTD, we combined three sources to identify these patients in the French Midi-Pyrénées region in 2012: the LTD ICD-10 code for type 2 diabetes mellitus, the hospital ICD-10 codes (as primary, related, or associated diagnosis), and the chronic exposure to oral antidiabetic drugs (OADs), defined by at least three OADs dispensed (corresponding to 3 months of treatment) during a 6-month period. Moreover, to rule out the differential diagnoses of type 2 diabetes mellitus, we excluded patients with an LTD or hospital diagnosis code for another disease that led to hyperglycemia as well as patients with chronic exposure to systemic glucocorticoids or antiprotease drugs during the year previous to the first identification of type 2

diabetes mellitus in 2012. The full algorithm is indicated in the “Supplementary material” section.

Age-specific prevalences were estimated according to sex and region; the crude and age–sex-adjusted overall prevalences with 95% CI were estimated for each region using the European Standard Population 2013.

## Ethical considerations

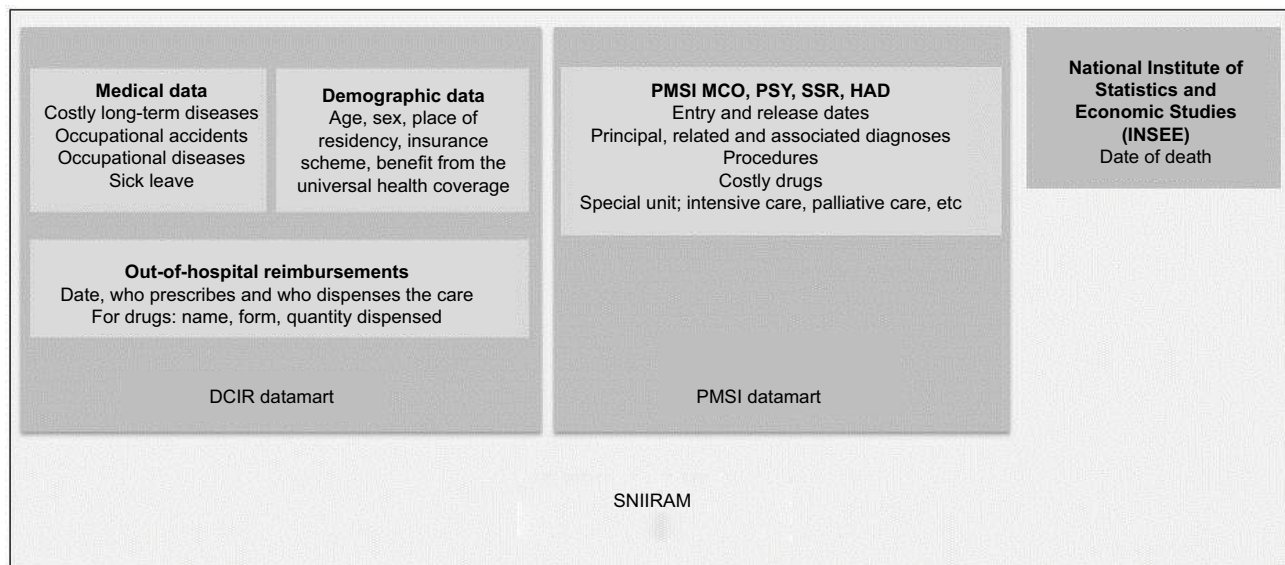
This study, which was observational in design and retrospective in nature, used data that were irreversibly anonymized prior to transfer to the research team. The study was conducted according to the amended Declaration of Helsinki, the International Guidelines for Ethical Review of Epidemiological Studies, and the Spanish and French laws on data protection and patients’ rights. The protocol for the Spanish part of the study was approved by the Ethics Committee of Navarre (Project 67/2013, session on October 30, 2013). The protocol for the French part of the study was approved by a convention with the Midi-Pyrénées Regional Directory of Health Insurance in October 2014.

## Results

### Comparisons between databases

In France, outpatient data, hospital data, and health status are linkable for all beneficiaries (virtually the whole French population, 67 million inhabitants) because of individual irreversible anonymous numbering. These data are kept in a huge digital warehouse, the SNIIRAM.<sup>13,14</sup> The SNIIRAM’s simplified architecture is shown in Figure 1. The Inter-Scheme Consumption Data (Données de Consommation Inter-Régimes; DCIR) set includes administrative data and all outpatient reimbursed health expenditures. The Program for the Medicalization of Information Systems (Programme de Médicalisation des Systèmes d’Information; PMSI) includes data on inpatient care in public and private hospitals.<sup>13,14</sup> Information regarding health status is held in a separate registry by the National Institute of Statistics and Economic Studies (INSEE). Until now, only the date of death has been linked to the SNIIRAM, but the causes of death should become available soon. No detailed individual socioeconomic data are available in France except that on universal coverage for low-income patients.<sup>13,14,20</sup> The main variables useful for epidemiological and pharmacoepidemiological studies are described in Table 1.

Data for Spanish databases are generally gathered at regional level; unlike the French data, they are not routinely combined into a unique national database. They have the same architecture in the Navarre and Basque Country and correspond to similar datasets as those used in the French



**Figure 1** Simplified architecture of the SNIIRAM.

**Abbreviations:** DCIR, Données de Consommation Inter-Régimes (Inter-Scheme Consumption Data); HAD, Hospitalisation à Domicile (home hospitalization); MCO, Médecine, Chirurgie, Obstétrique (medicine, surgery, obstetrics); PMSI, Programme de Médicalisation des Systèmes d'Information (Program for the Medicalization of Information Systems); PSY, psychiatry; SNIIRAM, Système National d'Information Inter-Régime de l'Assurance Maladie (National Health Insurance Information System); SSR, Services de Suite et de Réadaptation (after care and rehabilitation).

SNIIRAM (Figure 2). One of the most exhaustive databases is the Primary Care Electronic Medical Record System, which contains outpatient data and includes demographic information (date of birth, sex, basic health zone the patient belongs to), visits to primary care services (data and type), health problems, lifestyle, detailed clinical data, laboratory results, and drug prescription data. It has some variations in structure and the coding system depending on the autonomous community and period. The ICD-9-CM, ICD-10CM version adapted to Spanish health system, and ICPC-2 are all used.

For this study, the ICPC-2 was used in Navarre and the ICD-9-MC was used in the Basque Country. The validity of the information contained in these data sets has recently begun to be assessed. Results suggest that they are valid to assess the prevalence of cardiovascular risk factors, such as type 2 diabetes mellitus,<sup>18,21</sup> although more studies are required to assess validity to conduct other types of population-based surveillance studies. Another database with outpatient information is the outpatient drug-dispensing database, with information on drugs dispensed at retail pharmacies. The electronic prescription system was introduced to primary care in the Basque Country and Navarre in 2012 and 2014, respectively. More recently (2016–2017), prescriptions for specialized care have been included from both regions. Since the introduction of this new system, data from prescriptions and drug-dispensing have been gathered into a unique database for Navarre. For the Basque Country,

this can be extracted through the Osakidetza Business Intelligence system, the business intelligence system that extracts information from the different administrative health care databases in the region. Regarding hospital data, they are recorded in the MBDS, which provides clinical and sociodemographic information on all hospital discharges in the National Health System, including diagnoses and procedures, coded according to the International Classification of Diseases (ICD-10-ES from January 2016). It is less detailed than the French PMSI (Table 1). Several studies on the validity of MBDS conducted more than 1 decade ago have identified important reliability problems,<sup>22</sup> although more recent study suggests that data quality has improved.<sup>23</sup> However, there is a need for more studies on the validity of specific MBDS codes.<sup>24</sup> In addition, there are two other data sets that can be internally linked with the aforementioned data sets at an individual level: the mortality registry, with information on date and cause of death (since 2016, coded using ICD-10-ES), and the population register, which collects individual demographic and socioeconomic data.

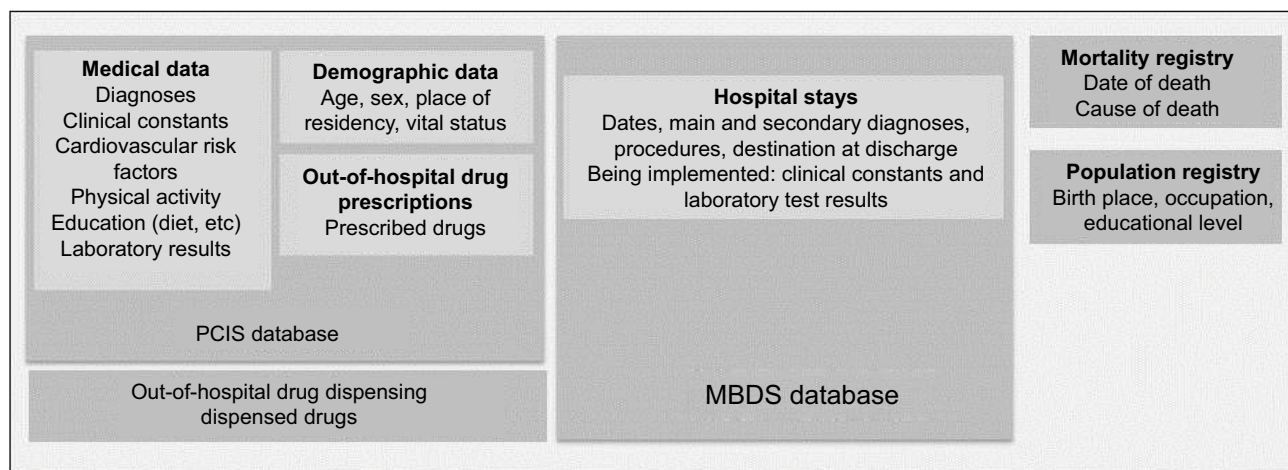
Ethical approval is required in both countries to access these data sets for research purposes. In France, all requests for SNIIRAM extractions are sent to the Health Data Institute (Institut National des Données de Santé; INDS). Two authorizations are mandatory from the National Data Protection Commission (Commission Nationale de l'Informatique et des Libertés; CNIL) regarding data protection and from an independent methodological committee (Comité d'Expertise pour

**Table 1** Main variables contained in French and Spanish health care databases

Variables	Spain	France
<b>Demographics</b>		
Date of registration of the patient in the information system	Yes	No
Date of diabetic onset	Yes (date of registration of the first diabetic episode: not necessarily means date of diabetic onset)	Necessitates algorithms to identify the first diagnosis code or the first antidiabetic drug dispensing
Sex	Yes	Yes
Birth date	Yes (year)	Yes (year)
Birth place	Not always	No
Working status	Active/retired	Active/retired
Pharmaceutical copayment category (proxy of socioeconomic class)	Yes	Beneficiary of the Couverture Médicale Universelle Complémentaire status for low-income patients; socioeconomic deprivation index available by geographical area
Living area	Basic health zone (5000–20,000 inhabitants)	Ilots Regroupés pour l'Information Statistique (IRIS): areas of 2000 inhabitants for main cities; municipalities' boundaries in the case of <2000 habitants
Education level	Compulsory, high school, university; not available in Basque Country	No
Nationality	Yes (probably biased)	No
Occupation	No	Some insurance schemes correspond to specific occupations
Date of death	Yes	Yes
Cause of death	Yes (in the mortality register, ICD-10-ES)	No (being implemented)
<b>Out-of-hospital examination, procedures, and health care</b>		
Code of the examination/care	Yes (ICPC-2 classification and PGD in Navarre, and ICD-9 in Basque Country)	Yes, using a specific national classification
Date of care	Yes	Yes
Physician who has prescribed	Yes	Yes
Health care provider	Yes	Yes
Results	Yes	No
<b>Primary care data</b>		
Long-term disabling disease	Yes (ICPC-2 classification, PGD, and ICD-9 in Basque Country)	Yes (ICD-10)
Diagnosis for each visit	Yes	No
Date of visit	Yes	Yes
Physician	Yes (anonymous identifier)	Yes (identifier and specialty)
Clinical data (blood pressure, weight, height)	Yes	No
Lifestyle data (smoking, alcohol consumption, physical activity)	Yes	No
<b>Out-of-hospital drug dispensing*</b>		
Code of the drugs	Yes, using ATC codes, except for many but not all over-the-counter drugs	Yes, using Club Inter Pharmaceutique (CIP) codes. Over-the-counter drugs are not recorded
Date of dispensing	Yes	Yes
Number of units dispensed	Yes	Yes
Physician who has prescribed	Yes	Yes
Pharmacist provider	Yes	Yes
Prescription data	Yes	No
<b>Hospital data</b>		
Dates of entry and of release	Yes	Yes
Diagnosis	Yes (ICD-9-MC at the moment of data extraction and ICD-10-ES from January 2016)	One main diagnosis ± I related diagnosis and unlimited associated diagnoses (ICD-10)
Specific departments (intensive care unit, palliative care, etc)	Yes	Yes
Procedures	Yes, date, and codes	Yes, date, and codes
Exposure to expensive drugs	No	Yes, date, and code of the drug
Exposure to non-expensive drugs	No	No

**Note:** \*In Navarre, from 2014 onward, there was a unique database for drug dispensing, which included prescriptions.

**Abbreviations:** ATC, anatomical, therapeutic, chemical classification; ICD, International Classification of Diseases; ICPC-2, International Classification of Primary Care, version 2; PGD, patient general data.



**Figure 2** Simplified architecture of Spanish databases.

**Abbreviations:** MBDS, Minimum Basic Data Set; PCIS, primary care information system.

les Recherches, les Etudes et les Evaluations dans le domaine de la Santé; CEREES) according to a new law published in 2016.<sup>25</sup> In Spain, procedures are defined at a regional level and, in all cases, the project must be approved by an ethics committee and by the Health Department of the Government of each region. In Navarre and the Basque Country, the Health Department nominates an internal coordinator for the project, who supervises the data extraction procedures and guarantees fulfillment of the law on personal data protection. Mandatory conditions are that the files are anonymized and that the data exportation process follows both the Spanish Constitutional Act 15/1999 of 13 December on personal data protection<sup>26</sup> and the law 41/2002 of 14 November, which concerns clinical information issues.<sup>27</sup>

## Prevalence of type 2 diabetes mellitus

Out of the population over 20 years of age at data extraction, the number of patients who had type 2 diabetes mellitus in Navarre and Basque Country was 32,638 and 132,455, respectively, leading to a crude prevalence of 6.62% (95% CI: 6.55–6.89) in Navarre and 7.01% (95% CI: 6.98–7.05) in the Basque Country (Table 2). In the Midi-Pyrénées region, a total of 141,669 patients were identified with type 2 diabetes mellitus. Of these, 21.9% were identified using only the outpatient drug data, 11.0% using only the LTD codes, 6.2% using only the in-hospital diagnosis codes, and the remaining 60.9% using at least two of these three sources (Figure 3). This led to a crude prevalence of 6.26% (95% CI: 6.23–6.29).

Age–sex-adjusted prevalences were 6.19 (95% CI: 6.16–6.22) for the French region, 6.84 (95% CI: 6.76–6.91) for Navarre, and 6.97 (95% CI: 6.93–7.01) for the Basque Country. There were statistically significant differences between

the regions, especially between the French region and the Basque Country (0.8% lower). The prevalence was significantly higher among males in the three regions (Table 2), and geographical differences were greater for males than females, with data 1.1% lower in the Midi-Pyrénées region than the Basque Country for males, and 0.4% for females.

Results relative to the age–sex prevalences for type 2 diabetes mellitus are given in Table 3. They increased with age and were higher in males in most age groups for all regions, with some exceptions in populations aged <35 years, where it was higher in females, especially in the French region. There were no relevant differences in the age-specific prevalence across regions, apart from the lower estimates observed in France compared to Spain for patients aged >65 years, especially in women, and the slightly higher prevalence in France for women aged <60 years.

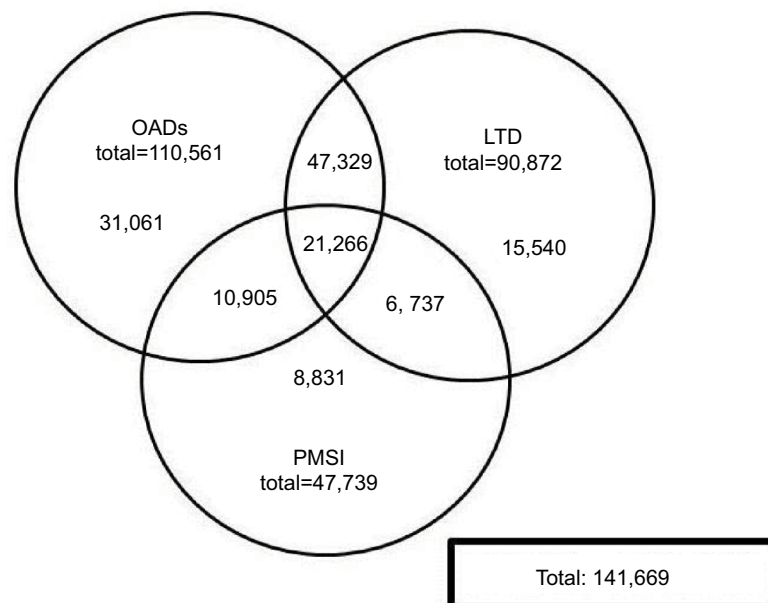
## Discussion

This study shows that base-specific procedures that share common grounds, but account for the particularities of each database, need to be developed to conduct cross-national epidemiological studies using French and Spanish electronic health care databases. The database-specific algorithms developed in this study to identify type 2 diabetes mellitus provided prevalence estimates between 6% and 7% in all three regions. They were 2% higher in males than females in all three regions and up to 20% higher in people aged >75 years, especially males.

This study demonstrates the feasibility of such approaches, provided that in-depth comparative analysis of database structures and contents in relation to the pathology under consideration is carried out beforehand. Indeed, using the

**Table 2** Prevalence of type 2 diabetes mellitus in Navarre, Basque Country, and Midi-Pyrénées expressed as percentages with 95% CIs

	Basque Country (n=1,888,830)	Navarre (n=493,443)	Midi-Pyrénées (n=2,260,948)
Crude prevalence			
Females	6.14 (6.09, 6.19)	5.60 (5.51, 5.98)	5.51 (5.47, 5.55)
Males	7.94 (7.88, 7.99)	7.48 (7.38, 7.86)	7.08 (7.03, 7.13)
Total	7.01 (6.98, 7.05)	6.62 (6.55, 6.89)	6.26 (6.23, 6.29)
Age-adjusted prevalence by sex			
Females	5.51 (5.46, 5.55)	5.34 (5.25, 5.43)	5.07 (5.03, 5.11)
Males	8.43 (8.37, 8.49)	8.33 (8.21, 8.45)	7.32 (7.26, 7.37)
Age–sex-adjusted prevalence	6.97 (6.93, 7.01)	6.84 (6.76, 6.91)	6.19 (6.16, 6.22)

**Figure 3** Combination of three sources to identify 141,669 prevalent French patients with type 2 diabetes mellitus.

**Note:** Numbers indicate the numbers and percentages of patients with type 2 diabetes mellitus identified using each source alone and combinations of sources.

**Abbreviations:** LTD, long-term disease; PMSI, Programme de Médicalisation des Systèmes d'Information (Program for the Medicalization of Information Systems); OADs, oral antidiabetic drugs.

**Table 3** Prevalence of type 2 diabetes mellitus according to age and sex expressed as percentages

Age group (years)	Basque Country		Navarre		Midi-Pyrénées	
	Males	Females	Males	Females	Males	Females
20–24	0.11	0.11	0.08	0.10	0.07	0.16
25–29	0.18	0.17	0.12	0.17	0.16	0.38
30–34	0.31	0.28	0.34	0.29	0.39	0.51
35–39	0.66	0.44	0.66	0.43	0.77	0.78
40–44	1.34	0.75	1.54	0.93	1.44	1.18
45–49	2.89	1.39	3.06	1.58	2.92	2.04
50–54	5.64	2.64	5.88	2.67	5.40	3.63
55–59	9.68	4.91	9.92	4.75	9.06	5.76
60–64	14.77	7.67	14.42	7.01	13.39	8.22
65–69	19.57	11.66	17.94	10.53	16.73	10.97
70–74	22.78	15.64	23.47	15.20	18.92	13.06
75–79	24.55	18.56	22.41	17.64	20.40	14.41
80–84	24.66	20.54	25.38	19.64	19.63	14.77
85–89	23.16	20.42	22.29	20.80	17.02	12.98
≥90	16.43	15.89	19.14	18.80	14.65	11.33

same algorithm in different databases has been suggested to cause a major risk of error.<sup>28</sup> Similarly, the EMIF project<sup>7</sup> has shown that different strategies need to be adopted to identify a particular condition (such as type 2 diabetes mellitus) when handling distinct sources of health data that have heterogeneous characteristics.

The procedure designed to identify patients with type 2 diabetes mellitus in the Spanish regions was based on the specific codes for this disease in the Primary Care Electronic Medical System. This procedure is similar to that developed by Roberto et al<sup>7</sup> for databases from a primary care setting and is similar to that used by Vinagre et al<sup>29</sup> in the Primary Care Electronic Medical System of Catalonia. The codes used in our study to identify type 2 diabetes mellitus have been validated in the Navarre database,<sup>18</sup> and also in other regions, such as Madrid.<sup>30</sup> In the Basque Country, the quality of the codification of diagnoses in the electronic health care

records for primary care is also high.<sup>31</sup> Nevertheless, studies on the quality and exhaustiveness of this type of data source in Spain are still needed.

In the French region, the procedure used to identify these patients was based on information from the SNIIRAM database, which is now frequently used for epidemiological purposes at a national level.<sup>13,19</sup> The information available in the SNIIRAM is based on an irreversibly anonymous identifier, which makes validation studies difficult.<sup>13,19</sup> To reinforce the identification of patients with diabetes mellitus, we used a combination of variables, which is a common process for studies in the SNIIRAM.<sup>13,19</sup> The results were highly consistent with the Spanish results.

The prevalences estimated in this study are in line with those provided in the seventh edition of the IDF diabetes atlas,<sup>32</sup> which showed an estimated prevalence for the 2015 adult Spanish population aged 20–79 years of 10.4 (age-adjusted on 2001 World Health Organization global structure of population: 7.7) and 7.4 (age-adjusted: 5.3) for the French population. Navarra and the Basque Country have among the lowest diabetes prevalences in Spain, a finding that could be related to the lower prevalence of obesity compared to other Spanish regions.<sup>33</sup> The higher prevalence in Spanish regions compared to France is similar to those reported in the literature.<sup>32</sup>

Our results showed a similar effect for age and sex in the three regions, with a gradual increment in prevalence, particularly in those aged 55–75 years, and a constantly higher prevalence in males than females of about 2–3 points. This is similar to previous studies that have demonstrated an increased prevalence of type 2 diabetes mellitus with increasing age, particularly in males.<sup>34–38</sup> The higher prevalence in young women in France could be because, although the corresponding LTD and hospital codes related to gestational diabetes were removed, some of these patients may have been captured.

The main strengths of this study are that it was based on recent health data from >5.9 million people from three regions of Spain and France and that it provides an exhaustive exploration of common and different fields from the main electronic health care databases used in research in both countries.

One of the main limitations of assessing type 2 diabetes mellitus was that it was based on registered data, so that patients who had the disease but had not been diagnosed were not included. A national population-based survey study conducted in Spain estimated an overall global incidence of

diabetes of 13.8%, of which about 6.0% of the population had unknown diabetes.<sup>33</sup> Another study, conducted in the Basque Country,<sup>39</sup> showed an overall prevalence of 10.6%; of this proportion, 4.3% were not aware they had diabetes. As pointed out earlier, a second limitation is that the identification of patients with type 2 diabetes mellitus was not validated in all the sources used in this study. However, our results are consistent with the published literature, which is reassuring considering the risk of misdiagnosis. Of note, the procedure used in Spain allowed the identification of patients with type 2 diabetes mellitus treated by diet only (not receiving any antidiabetic medication) whereas, in France, those treated by diet only but without a specific LTD or hospital code could not be captured. However, 22.0% of the identified patients in the French database had no data on OAD prescriptions (Figure 3), whereas a French study estimated that, in 2010, 11% of patients were treated by diet only and 12% by insulin only.<sup>40</sup> Consequently, we may have captured most patients treated by diet only in France, despite no visit data from general practitioners. A third limitation is that there was 2 years difference between the Spanish and French data extraction due to accessibility issues. However, there is no reason to expect large differences in prevalences within this 2-year period.

## Conclusion

These results provide more evidence on the recently stated need to develop a common public health research infrastructure at a European level to facilitate the reuse of health administrative databases for research purposes.<sup>41</sup> The potential of health administrative data to advance developments in public health is being widely recognized, but important study is needed to overcome major obstacles regarding accessibility, legal issues, record linkage, integration, and data validation. Our study shows the possibility and benefits of reusing these data transnationally for research purposes to better understand the epidemiology of type 2 diabetes mellitus. We were able to achieve comparable estimates of prevalences “between” the Navarre, the Basque Country, and the Midi-Pyrénées regions, but with a slight gradient from less to more prevalence in the Midi-Pyrénées and the Basque Country. This could be the first step toward more combined studies across regions and countries that are based on monitoring patients with type 2 diabetes mellitus using health administrative data sets. These studies could focus, for instance, on identifying the risk factors for the major complications or assess the risk–benefit ratios for new drugs.



## Acknowledgments

We thank the Caisse régionale d'Assurance Maladie, particularly Dr Robert Bourrel, and the Regional Health Service of Navarre, particularly Javier Baquedano and Marian Nuin, for data extraction. This study was supported by the POCTEFA Programme (REFBIO EFA 237/11), Instituto de Salud Carlos III (grant PI15/02196), Spanish thematic network REDIS-SEC (grant RD12/0001 and RD16/0001 from the Instituto de Salud Carlos III, Spanish Ministry of Health and co-financed by the European Regional Development Fund), and by the Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT620-13).

## Disclosure

The authors report no conflicts of interest in this work.

## References

- Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdóttir AB, Sørensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol*. 2010;106(2):86–94.
- Huerta C, Abbing-Karahagopian V, Requena G, et al. Exposure to benzodiazepines (anxiolytics, hypnotics and related drugs) in seven European electronic healthcare databases: a cross-national descriptive study from the PROTECT-EU Project. *Pharmacoepidemiol Drug Saf*. 2016;25(suppl 1):56–65.
- Klungel OH, Kurz X, de Groot MCH, et al. Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project. *Pharmacoepidemiol Drug Saf*. 2016;25(suppl 1):156–165.
- PROTECT. *Drug Consumption Databases in Europe. Countries Summary*; 2015. Available from: <http://www.imi-protect.eu/documents/DUinventoryCOUNTRIESFeb2015.pdf>. Accessed April 16, 2017.
- Gutacker N, Bloor K, Cookson R, et al. Hospital surgical volumes and mortality after coronary artery bypass grafting: using international comparisons to determine a safe threshold. *Health Serv Res*. 2017;52(2):863–878.
- Auffray C, Balling R, Barroso I, et al. Making sense of big data in health research: Towards an EU action plan. *Genome Med*. 2016;8(1):71.
- Roberto G, Leal I, Sattar N, et al. Identifying cases of type 2 diabetes in heterogeneous data sources: strategy from the EMIF project. *PLoS One*. 2016;11(8):e0160648.
- Librero J, Ibañez B, Martínez-Lizaga N, Peiró S, Bernal-Delgado E; Spanish Atlas of Medical Practice Variation Research Group. Applying spatio-temporal models to assess variations across health care areas and regions: Lessons from the decentralized Spanish National Health System. *PLoS One*. 2017;12(2):e0170480.
- Aizpuru F, Latorre A, Ibañez B, et al. Variability in the detection and monitoring of chronic patients in primary care according to what is registered in the electronic health record. *Fam Pract*. 2012;29(6):696–705.
- Erviti J, Alonso A, Gorricho J, López A. Oral bisphosphonates may not decrease hip fracture risk in elderly Spanish women: a nested case-control study. *BMJ Open*. 2013;3(2):e002084.
- Borrell C, Mari-Dell'olmo M, Serral G, Martínez-Beneito M, Gotsens M; MEDEA Members. Inequalities in mortality in small areas of eleven Spanish cities (the multicenter MEDEA project). *Health Place*. 2010;16(4):703–711.
- Regidor E, Vallejo F, Granados JAT, Viciano-Fernández FJ, de la Fuente L, Barrio G. Mortality decrease according to socioeconomic groups during the economic crisis in Spain: a cohort study of 36 million people. *Lancet*. 2016;388(10060):2642–2652.
- Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc J-L, Sailler L. French health insurance databases: What interest for medical research? *Rev Médecine Interne Fondée Par Société Natl Française Médecine Interne*. 2015;36:411–417.
- Tuppin P, de Roquefeuil L, Weill A, Ricordeau P, Merlière Y. French national health insurance information system and the permanent beneficiaries sample. *Rev Épidémiologie Santé Publique*. 2010;58:286–290.
- Refbio. Pyrenees biomedical network. *Refbio Pyrenees Biomed Netw*. Available from: <https://refbio.eu/en/>. Accessed June 8, 2018.
- Tuppin P, Drouin J, Mazza M, Weill A, Ricordeau P, Allemand H. Hospitalization admission rates for low-income subjects with full health insurance coverage in France. *Eur J Public Health*. 2011;21:560–566.
- Puig-Junoy J, Rodríguez-Feijóo S, Lopez-Valcarcel BG. Paying for formerly free medicines in Spain after 1 year of co-payment: changes in the number of dispensed prescriptions. *Appl Health Econ Health Policy*. 2014;12(3):279–287.
- Moreno-Iribas C, Sayon-Orea C, Delfrade J, et al. Validity of type 2 diabetes diagnosis in a population-based electronic health record database. *BMC Med Inform Decis Making*. 2017;17(1):34.
- Palmaro A, Moulis G, Despas F, Dupouy J, Lapeyre-Mestre M. Overview of drug data within French health insurance databases and implications for pharmacoepidemiological studies. *Fundam Clin Pharmacol*. 2016;30(6):616–624.
- Rey G, Jouglu E, Fouillet A, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997 - 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health*. 2009;9:33.
- Ramos R, Balló E, Marrugat J, et al. Validity for use in research on vascular diseases of the SIDIAP (information system for the development of research in primary care): the EMMA study. *Rev Espanola Cardiol Engl Ed*. 2012;65:29–37.
- Calle JE, Saturno PJ, Parra P, et al. Quality of the information contained in the minimum basic data set: results from an evaluation in eight hospitals. *Eur J Epidemiol*. 2000;16(11):1073–1080.
- Yetano J, Izarzugaza I, Aldasoro E, Ugarte T, López-Arbeloa G, Aguirre U. Calidad de las variables administrativas del Conjunto Mínimo Básico de Datos de Osakidetza-Servicio Vasco de Salud. *Rev Calid Asist Organo Soc Espanola Calid Asist*. 2008;23:216–221. Spanish.
- Medrano IH, Guillán M, Masjuan J, Cánovas AA, Gogorcena MA. Reliability of the minimum basic dataset for diagnoses of cerebrovascular disease. *Neurología*. 2017;32:74–80.
- Legifrance [webpage on the Internet]. LOI n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. *J Off Répub Fr*. 2016. Available from: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031912641&categorieLien=id>. Accessed April 16, 2017. French.
- España [webpage on the Internet]. Ley orgánica 15/1999, de 13 de Diciembre, de Protección de datos de carácter personal. *Boletín Oficial del Estado*. 1999;298:43088–43099. Available from: <https://www.boe.es/buscar/doc.php?id=BOE-A-1999-23750>. Accessed April 16, 2017. Spanish.
- España [webpage on the Internet]. Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. *Boletín Oficial del Estado*. 2002;274:40126–40132. Available from: <https://www.boe.es/buscar/act.php?id=BOE-A-2002-22188>. Accessed April 16, 2017. Spanish.
- Moore TJ, Furberg CD. Electronic health data for postmarket surveillance: a vision not realized. *Drug Saf*. 2015;38(7):601–610.
- Vinagre I, Mata-Cases M, Hermosilla E, et al. Control of glycemia and cardiovascular risk factors in patients with type 2 diabetes in primary care in Catalonia (Spain). *Diabetes Care*. 2012;35(4):774–779.
- De Burgos-Lunar C, Salinero-Fort MA, Cárdenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC Med Res Methodol*. 2011;11:146.

31. Orueta J, Urraca J, Berraondo I, Darpon J. ¿Es factible que los médicos de primaria utilicen CIE-9-MC? Calidad de la codificación de diagnósticos en las historias clínicas informatizadas. *Gac Sanit.* 2006;20:194–201. Spanish.
32. Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract.* 2011;94:311–321.
33. Soriguer F, Goday A, Bosch-Comas A, et al. Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. *Diabetologia.* 2012;55(1):88–93.
34. Gourdy P, Ruidavets JB, Ferrieres J, et al. Prevalence of type 2 diabetes and impaired fasting glucose in the middle-aged population of three French regions - The MONICA study 1995-97. *Diabetes Metab.* 2001;27(3):347–358.
35. Bringer J, Fontaine P, Detournay B, Nachit-Ouinekh F, Bami G, Eschwege E. Prevalence of diagnosed type 2 diabetes mellitus in the French general population: the INSTANT study. *Diabetes Metab.* 2009;35(1):25–31.
36. Bonaldi C, Vernay M, Roudier C, et al. A first national prevalence estimate of diagnosed and undiagnosed diabetes in France in 18- to 74-year-old individuals: the French Nutrition and Health Survey 2006/2007. *Diabet Med.* 2011;28(5):583–589.
37. Valverde JC, Tormo M-J, Navarro C, et al. Prevalence of diabetes in Murcia (Spain): a Mediterranean area characterised by obesity. *Diabetes Res Clin Pract.* 2006;71(2):202–209.
38. Tamayo-Marco B, Faure-Nogueras E, Roche-Asensio MJ, Rubio-Calvo E, Sánchez-Oriz E, Salvador-Oliván JA. Prevalence of diabetes and impaired glucose tolerance in Aragón, Spain. *Diabetes Care.* 1997;20:534–536.
39. Aguayo A, Urrutia I, González-Frutos T, et al. Prevalence of diabetes mellitus and impaired glucose metabolism in the adult population of the Basque Country, Spain. *Diabet Med.* 2017;34: 662–666.
40. Haute Autorité de Santé. *Epidémiologie et coût du diabète de type 2 en France.* 2013. Available from: [https://www.has-sante.fr/portail/upload/docs/application/pdf/2013-03/argumentaire\\_epidemiologie.pdf](https://www.has-sante.fr/portail/upload/docs/application/pdf/2013-03/argumentaire_epidemiologie.pdf). Accessed April 16, 2017. French.
41. Burgun A, Bernal-Delgado E, Kuchinke W, et al. Health data for public health: towards new ways of combining data sources to support research efforts in Europe. *Yearb Med Inform.* 2017;26: 235–240.

## Supplementary material

### Algorithm for the identification of prevalent type 2 diabetes mellitus patients in the SNIIRAM

Algorithm for the identification of the prevalent type 2 diabetes mellitus patients in 2012 in the Midi-Pyrénées region:

1. Extraction of the Midi-Pyrénées SNIIRAM data of the years 2011, 2012 and 2013.
  2. Definition of the patients:
    - #1 Long-term disease
      - ongoing in 2012 OR starting in 2012
      - AND with the E11.X ICD-10 code (Type II diabetes mellitus)
    - #2 In-hospital diagnosis code in the PMSI database as principal diagnosis OR related diagnosis OR associated diagnosis
      - with an hospital stay entry date in 2011 OR in 2012
      - AND with the E11.X ICD-10 code (Type II diabetes mellitus)
    - #3 At least 3 out-of-hospital dispensing of oral anti-diabetic drugs (ATC code beginning by A10B) with traditional packaging during the period 2011-2012
    - #4 At least 2 out-of-hospital dispensing of oral antidiabetic drugs (ATC code beginning by A10B) with large packaging during the period 2011-2012
    - #5 At least 1 out-of-hospital dispensing of oral antidiabetic drugs (ATC code beginning by A10B) with large packaging during the period 2011-2012 and 1 or 2 out-hospital dispensing of oral antidiabetic drugs (ATC code beginning by A10B) with traditional packaging during the period 2011-2012
    - #6 Long-term disease
      - ongoing in 2012 OR starting in 2012 or starting during the six months following the first event among #1, #2, #3 and #4
      - AND with an ICD-10 code for a disease responsible for secondary diabetes mellitus, that is:
        - E05.X (thyrotoxicosis [hyperthyroidism])
        - OR E24.X (Cushing syndrome)
        - OR E22.0 (acromegaly and pituitary gigantism) OR E22.9 (hyperfunction of pituitary gland, unspecified)
        - OR E83.1 (disorders of iron metabolism: hemochromatosis, excluding anemia by iron deficiency (D50) and sideroblastic (D64.0-D64.3))
- OR M14.5 (arthropathies in other endocrine, nutritional and metabolic disorders: in acromegaly and pituitary gigantism, hemochromatosis, hypothyroidism or thyrotoxicosis [hyperthyroidism])
  - OR K86.0 (alcohol-induced chronic pancreatitis), OR K86.1 (other chronic pancreatitis), OR K86.8 (other specified diseases of pancreas), OR K86.9 (disease of pancreas, unspecified), OR K90.3 (pancreatic steatorrhea)
  - OR O24.1 (pre-existing diabetes mellitus, non-insulin-dependent), OR O24.9 (diabetes mellitus arising in pregnancy)
- #7 In-hospital diagnosis code in the PMSI database as principal diagnosis OR related diagnosis OR associated diagnosis
    - with an hospital stay entry date during the year before the first event among #1, #2, #3 and #4 OR during the six months following the first event among #1, #2, #3 and #4
    - AND with an ICD-10 code for a disease responsible for secondary diabetes mellitus, that is:
      - E05.X (thyrotoxicosis [hyperthyroidism])
      - OR E24.X (Cushing syndrome)
      - OR E22.0 (acromegaly and pituitary gigantism) OR E22.9 (hyperfunction of pituitary gland, unspecified)
      - OR E83.1 (disorders of iron metabolism: hemochromatosis, excluding anemia by iron deficiency (D50) and sideroblastic (D64.0-D64.3))
      - OR M14.5 (arthropathies in other endocrine, nutritional and metabolic disorders: in acromegaly and pituitary gigantism, hemochromatosis, hypothyroidism or thyrotoxicosis [hyperthyroidism])
      - OR K86.0 (alcohol-induced chronic pancreatitis), OR K86.1 (other chronic pancreatitis), OR K86.8 (other specified diseases of pancreas), OR K86.9 (disease of pancreas, unspecified), OR K90.3 (pancreatic steatorrhea)
      - OR O24.1 (pre-existing diabetes mellitus, non-insulin-dependent), OR O24.9 (diabetes mellitus arising in pregnancy)

- #8 At least 3 out-of-hospital dispensing of systemic glucocorticoids (ATC code beginning H02AB) during the year before the first event among #1, #2, #3 and #4
- #9 At least 3 out-of-hospital dispensing of antiproteases (ATC code beginning J05AE) during the year before the first event among #1, #2, #3 and #4

### Definition

(#1 OR #2 OR #3 OR #4 OR #5) AND NOT (#6 OR #7 OR #8 OR #9)

### Remarks regarding this algorithm

- 3, #4 and #5 exclude insulin dispensing; indeed these drugs are not specific of type 2 diabetes mellitus, and therefore cannot be included in this algorithm. Consequently, this will fail in identifying the patients with insulin dispensing and without oral antidiabetic drug dispensing nor long-term disease or hospitalization with type 2 diabetes mellitus diagnosis codes. However, this situation seems improbable.
- Because of data extractions (years 2011-2013), if the first event among #1, #2, #3 and #4 occurred in 2011, #6, #7 and #8 cannot be searched during the full year before.

### Comparison of this algorithm compared to other French algorithms

- The definition (not validated) used by the CNAMTS to identify prevalent diabetic patients (whatever the type of

diabetes mellitus) is (#1 or #3 or #4), extended to all diabetes mellitus diagnosis codes and insulins (unpublished data).

- The definition used by the CNAMTS to identify diabetes mellitus (whatever the type of diabetes mellitus) as comorbidity in the SNIIRAM for Charlson's score calculation is (#1 or #2 or #3 or #4), extended to all diabetes mellitus diagnosis codes and insulins, recorded during the year before index date.<sup>1</sup>
- The definition used in most studies identifying diabetes mellitus patients in the SNIIRAM is (#3 or #4).<sup>2</sup> Indeed, the population of interest is frequently restricted to treated patients in epidemiologic studies conducted in the SNIIRAM.<sup>3</sup>

Our algorithm combines all the sources of information regarding diabetes mellitus in the SNIIRAM and might be more accurate to estimate its prevalence.

### References

1. Bannay A, Chaignot C, Blotière P-O, et al. Score de Charlson à partir des données du Sniiram chaînées au PMSI : faisabilité et valeur pronostique sur la mortalité à un an. *Rev Epidemiol Sante Publique*. 2013;61:S9. French.
2. Weill A, Païta M, Tuppin P, et al. Benfluorex and valvular heart disease: a cohort study of a million people with diabetes mellitus. *Pharmacoepidemiol Drug Saf*. 2010;19(12):1256–1262.
3. Kusnik-Joinville O, Weill A, Ricordeau P, et al. Treated diabetes in France in 2007: a prevalence rate close to 4% and increasing geographic disparities. *Bull Epidemiol Hebd*. 2008;43:409–413.

### Clinical Epidemiology

#### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress