CHAPTER 2

A SYSTEMATIC REVIEW OF CASE-IDENTIFICATION ALGORITHMS BASED ON ITALIAN HEALTHCARE ADMINISTRATIVE DATABASES FOR TWO RELEVANT DISEASES OF THE ENDOCRINE SYSTEM: DIABETES MELLITUS AND THYROID DISORDERS

Teresa Dalla Zuanna,[1] Gisella Pitter,[2] Cristina Canova,[1] Lorenzo Simonato,[1] Roberto Gnavi[3]
[1] Department of Cardiological, Thoracic and Vascular Sciences, University of Padua, Padua (Italy)
[2] Local Health Unit "Azienda ULSS 2 Marca Trevigiana", Veneto Region, Treviso (Italy)
[3] Epidemiology Unit, ASL TO3, Piedmont Region, Grugliasco, Turin (Italy)
Corresponding author: Teresa Dalla Zuanna; teresa.dallazuanna@studenti.unipd.it

ABSTRACT

BACKGROUND: diabetes mellitus (DM) and thyroid disorders (TDs) are two of the most prevalent and relevant endocrine disorders worldwide, and determining their occurrence and their follow-up pathways is essential. In Italy, due to the presence of a universal health care system, administrative data can be effectively used to determine these measurements. DM is an ideal model for surveillance with administrative data, due to its specific pharmacologic treatment, high rate of hospitalization, and specific care units. The identification of TDs, conversely, is more challenging: they are less frequently managed in a hospital setting, and even if the treatment is highly specific, subclinical forms often do not need any pharmacological treatment.

OBJECTIVES: to identify and to describe all DM and TD case-identification algorithms by means of Italian Healthcare Administrative Databases (HADs), through the review of papers published in the past 10 years.

METHODS: this study is part of a project that systematically reviewed case-identification algorithms for 18 acute and chronic conditions by means of HADs in Italy.

PubMed was searched for original articles, published between 2007 and 2017, in Italian or English. The search string consisted of a combination of free text and MeSH terms with a common part that focused on HADs and a disease-specific part.

All identified papers were screened by two independent reviewers. Pertinent papers were classified according to the objective for which the algorithm had been used, and only articles that used algorithms for "primary objectives" (**I** disease occurrence; **II** population/cohort selection; **III** outcome identification) were considered for algorithm extraction. The HADs used (hospital discharge records, drug prescriptions, etc.), ICD-9 and ICD-10 codes, ATC classification of drugs, follow-back periods, and age ranges applied by the algorithms have been reported. Further information on specific objective(s), accuracy measures, sensitivity analyses and the contribution of each HAD, have also been recorded.

Algorithms were divided between those identifying type 2/not specified DM and type 1 DM, and those created to identify hypo- and hyperthyroidism.

RESULTS: of the 780 articles identified for DM, 77 were included and a further 14 papers were added by screening the references. For TD, 65 articles were identified through the search string and 5 of them were included. Of the selected articles, 64% and 80% were published after 2014 for DM and TD, respectively, and 33% (for DM) and 20% (for TD) used multicentric national or international data. Forty original algorithms for DM (29 for type 2 DM/not-specified DM, and 11 for type 1 DM) and 9 for TD (6 for hypo- and 3 for

WHAT IS ALREADY KNOWN

- **Identifying the occurrence and monitoring the follow-up pathways of diabetes mellitus and thyroid disorders, two of the most prevalent and relevant endocrine disorders worldwide, is of paramount importance.**
- **During the last ten years, healthcare administrative data, which routinely collect patient-level information on healthcare services provided by the Italian universal health system, have been increasingly used for epidemiological purposes.**
- **Due to the specificity of pharmacologic treatment, high rates of hospitalization and dedicated care units, diabetes is an ideal model for surveillance with administrative data. Conversely, the identification of thyroid disorders is more challenging.**

WHAT THIS PAPER ADDS

- **This review provides a comprehensive overview of the algorithms used to identify type 1 and 2 diabetes mellitus and hypo- and hyperthyroidism in Italian administrative databases in papers published between 2007 and 2017.**
- **Italian literature on the use of administrative healthcare data for case identification of diabetes is vast, with algorithms that are quite similar to one another. A few validation studies have been recently performed with encouraging results concerning specificity and sensitivity.**
- **On the contrary, the literature concerning algorithms to identify thyroid disorders is relatively poor, and no validation studies have been included in the review.**

hyperthyroidism) were extracted. In 6 algorithms, specific selections were made so as not to include gestational diabetes. With regard to type 2 DM, the most commonly used sources were the drug prescription database (DPD, 27 cases), hospital discharge record database (HDD, 23 cases), and exemption from healthcare co-payment database (ECD, 19 cases). Other sources were the ambulatory care services database (ACD), birth register, and mortality record database (MRD). Among the 11 algorithms to identify type 1 DM, 9 used DPD, 7 ECD, and 6 HDD; in one case ACD codes were added, and all 11 algorithms but one was applied to a population of young people (always <35 years old).
With regard to TDs, 2 algorithms from one paper for hypo- and hyperthyroidism relied on DPD as the only source, the other 7 original algorithms combined DPD with HDD (5 cases), ECD (3 cases), and ACD (1 case). One paper identified autoimmune/iodine deficiency hypothyroidism by subtracting iatrogenic hypothyroidism cases (identified through records of previous procedures from HDD and ACD) from the whole hypothyroid population (identified through DPD).
External validation was performed for two algorithms for DM and none for TD. The first algorithm for DM was obtained through HDD only and its sensitivity ranged from 61% to 70%, the second had a sensitivity of 71%.
CONCLUSION. Italian literature on the use of administrative healthcare data for case identification of diabetes is vast; the proposed algorithms are quite similar to one another, and the differences between them are rarely accompanied by clinical justification. On the contrary, the literature concerning thyroid disorders is relatively poor. Further validations of the proposed algorithms, as well as their further implementation, are needed.

Keywords: algorithms, healthcare administrative data, electronic health records, diabetes mellitus, thyroid disorders

RIASSUNTO

INTRODUZIONE: il diabete mellito (DM) e le patologie tiroidee (TD) sono i due gruppi di malattie endocrine a maggiore prevalenza e di maggiore impatto a livello mondiale, ed è essenziale stimarne l'occorrenza e valutarne i percorsi terapeutici e di follow-up. In Italia, grazie alla natura universalistica del sistema sanitario nazionale, è possibile ottenere tali misurazioni utilizzando i flussi amministrativi sanitari (FAS) in maniera integrata. Il DM rappresenta la patologia ideale per una sorveglianza con dati correnti, in virtù del trattamento farmacologico specifico, degli elevati tassi di ospedalizzazione e dei servizi ambulatoriali dedicati. Al contrario, l'identificazione delle TD è più complessa: raramente richiedono l'ospedalizzazione e, nonostante il trattamento farmacologico sia molto specifico, spesso le forme subcliniche non vengono identificate perché non richiedono alcun trattamento.
OBIETTIVI: identificare e descrivere tutti i lavori pubblicati negli ultimi 10 anni che, utilizzando FAS italiani, hanno elaborato almeno un algoritmo originale per l'identificazione di pazienti affetti da DM o TD.
METODI: questo studio si inserisce all'interno di un progetto di 16 revisioni sistematiche per la valutazione dello stato dell'arte degli algoritmi per l'identificazione di 18 condizioni acute e croniche. La revisione, effettuata in doppio, mira a identificare articoli originali pubblicati tra il 2007 e il 2017 in inglese o italiano, individuati su Pubmed mediante una stringa di ricerca che combina testo libero con termini MeSH, in parte comune a tutte le patologie e in parte patologia-specifica. Gli articoli pertinenti sono stati classificati secondo l'obiettivo di utilizzo degli algoritmi e solo gli articoli con obiettivo primario (**I** stima di occorrenza; **II** identificazione di popolazioni/coorti; **III** identificazione della patologia come esito) sono stati inclusi nella revisione. Per ogni algoritmo sono state estratte informazioni sui FAS utilizzati (schede di dimissione ospedaliera – SDO, esenzioni ticket – ET, prescrizioni farmaceutiche – PF), i criteri di identificazione dei casi, il periodo di osservazione/follow-back e le fasce d'età considerate. Sono state inoltre riportate eventuali validazioni esterne. Per il DM, gli algoritmi che identificavano il DM di tipo 2 (o senza specifica) sono stati distinti da quelli che identificavano il DM di tipo 1, per la tiroide sono stati separati gli algoritmi creati per identificare ipo- ed ipertiroidismo.
RISULTATI: per il DM la stringa di ricerca ha identificato 780 articoli, di cui 77 sono stati inclusi e 14 ulteriori lavori sono stati aggiunti dalla revisione delle voci bibliografiche. Per la tiroide, sono stati inclusi 5 dei 65 lavori identificati dalla stringa. Tra i lavori selezionati, il 64% per il DM e l'80% per le TD erano stati pubblicati tra il 2014 e il 2017, e il 33% e il 20% utilizzavano dati multicentrici nazionali o internazionali. Dai lavori inclusi sono stati identificati 40 algoritmi originali per il DM (29 per il DM di tipo 2/non specificato e 11 per il DM di tipo 1) e 9 per le TD (6 per l'ipo- e 3 per l'ipertiroidismo). In 6 algoritmi per il DM erano state aggiunte selezioni per escludere i casi di diabete gestazionale. Nel caso del DM di tipo 2, i FAS più utilizzati erano le PF (27 casi), le SDO (23 casi), e le ET (19 casi). Altre fonti utilizzate erano le visite ambulatoriali e i registri di nascita e di mortalità. Tra gli 11 algoritmi identificati per il DM di tipo 1, 9 usavano le PF, 7 le ET, 6 le SDO e uno il flusso ambulatoriale. Tutti erano applicati a popolazioni di meno di 35 anni.
Per quanto riguarda le TD, 2 algoritmi dello stesso lavoro utilizzavano solo le PF per ipo- e ipertiroidismo, gli altri 7 utilizzavano una combinazione di: PF e SDO (5 casi), ET (3 casi) e in un caso flusso ambulatoriale. Un lavoro identificava l'ipotiroidismo autoimmune o da carenza di iodio sottraendo i casi di ipotiroidismo iatrogeno (identificato dai record di procedure dalle SDO e dall'ambulatoriale) dall'intera popolazione ipotiroidea, identificata con le PF; 2 algoritmi per il DM (e nessuno per TD) sono stati validati.
CONCLUSIONI: in letteratura sono stati proposti numerosi algoritmi per l'identificazione di DM con l'utilizzo di FAS, abbastanza simili fra loro e le differenze sono raramente accompagnate da giustificazione clinica. Al contrario, gli algoritmi proposti per le patologie tiroidee sono ancora pochi. Sono necessari ulteriori studi di validazione sugli algoritmi proposti e un'ulteriore implementazione degli stessi.

Parole chiave: algoritmi, database amministrativi sanitari, diabete mellito, ipotiroidismo, ipertiroidismo

INTRODUCTION

Diabetes mellitus (DM) and thyroid disorders are two of the most frequent endocrine disorders worldwide. Diabetes was estimated to affect 8.5% of the adult population worldwide in 2014,[1] 7.1% of men and 6.8% of women in Italy in 2013,[2] and it will most probably increase in the next few years.[3] The prevalence of overt hyperthyroidism is roughly similar in Europe and the United States (0.7% and 0.5%, respectively), while the prevalence of overt hypothyroidism in the general population ranges from between 0.2% and 5.3% in Europe and 0.3% and 3.7% in the USA.[4] Due to the high prevalence and increased occurrence of these diseases, and their impact on people's morbidity and quality of life, determining their occurrence and their follow-up pathways is particularly important.

An effective way to estimate these measurements is the use of administrative data on health care, which are easily available in Italy due to the presence of a universal healthcare system and free access to health care for all citizens. Previous published studies that attempted to identify the two groups of diseases through administrative data in Italy, however, varied considerably: diabetes, on one hand, is one of the first diseases that was monitored through administrative data[5,6] and it has been widely studied. The attempts to identify thyroid disorders, on the other hand, are more recent and few.

With regard to diabetes, its high risk of complications, both acute and chronic, and its increased risk of premature mortality[1] have made this disease one of the major public health concerns, with a high impact on health systems.[7] Monitoring diabetes is key in public health surveillance for defining the burden of disease, planning health services, evaluating strategies in disease prevention and control, and assessing outcomes.[6] Furthermore, diabetes is an ideal model for surveillance with administrative data, due to its highly specific pharmacologic treatment, high rate of hospitalization, and, at least in Italy, diabetes-specific care units for patients with confirmed diagnosis of diabetes. Moreover, since diabetic subjects have facilities for out-of-hospital care, local health authorities often keep a register of subjects with a diagnosis of diabetes for administrative purposes.[6]

Both type 2 (accounting for approximately 95% of overall diabetes) and type 1 (5%) have been identified through administrative data, and due to their different epidemiology and pathophysiology, attempts have been made to distinguish them.[7-9] Type 1 is an autoimmune disorder, usually arising during childhood or adolescence, and consists in the progressive destruction of pancreatic βcells producing insulin, and consequent need for lifelong exogenous insulin replacement therapy.[10] Type 2 diabetes usually develops after 30-40 years of age, and is caused both by familiarity and lifestyle factors, like overweight and inadequate physical activity.[1] The gradual loss of insulin production brings to progressive higher levels of glycemia, leading first to the need for oral hypoglycemic drugs, then to insulin replacement. Thyroid dysfunction is characterized by either a deficiency or an excess of thyroid hormones and is most often caused by a primary disorder of the thyroid gland.[4] Hypothyroidism commonly develops as a result of autoimmune thyroiditis (Hashimoto thyroiditis) or iodine deficiency, but can also ensue from thyroidectomy or thyroid irradiation (iatrogenic). Hyperthyroidism is mainly due to autoimmunity (Graves' disease), toxic nodular goiter or thyroid adenoma[4]. Thyroid dysfunction is considered subclinical when serum levels of thyroid-stimulating hormone are altered but thyroid hormones are within the normal range.[11,12] Structural thyroid disorders such as goiter or tumors may also be present without any functional alteration of the hormonal profile.

Thyroid hormones play a key role in metabolic homeostasis, functioning of the nervous and cardiovascular systems, reproduction, and growth. Therefore, both hypothyroidism and hyperthyroidism can exert detrimental effects on multiple body systems and have been associated with unfavorable health outcomes, especially cardiovascular events.[11,12] Given the public health impact of thyroid disorders, availability of a monitoring system through administrative data would be of great value. Like diabetes, hypothyroidism and hyperthyroidism are treated with specific drugs that can be considered as markers of disease. At variance with diabetes, they are infrequently managed in the hospital setting and cannot rely on disease-specific healthcare services. Moreover, subclin-

DIABETES: ((diabetes [Title/Abstract]) OR diabetes mellitus [MeSH Terms])

THYROID DISORDERS: Goiter[MeSH Terms] OR Hyperthyroidism[MeSH Terms] OR Hyperthyroxinemia [MeSH Terms] OR Hypothyroidism[MeSH Terms] OR Thyroiditis[MeSH Terms] OR Hyperthyroidism[title/Abstract] OR Hypothyroidism[title/Abstract] OR Thyroiditis[title/Abstract] OR "Graves disease"[title/Abstract] OR "Graves' disease"[title/Abstract] OR Hashimoto[title/Abstract] OR "thyroid dysfunction"[title/Abstract] OR "thyroid hyperfunction"[title/Abstract] OR "thyroid hypofunction"[title/Abstract] OR "thyroid disease*"[title/Abstract] OR "thyroid disorder*"[title/Abstract]

Box 1. Search strings used to select records from PubMed.

ical forms often do not need any pharmacological treatment. All these peculiarities make identification through administrative data not straightforward.
The objective of this systematic review is to describe the characteristics of algorithms that have been used in the past 10 years, in Italy, for case identification of diabetes (both type 1 and 2) and thyroid disorders.

METHODS

All method details are available in a specific paper[13] which reports the study protocol with all information on literature search (specific search string applied to retrieve administrative healthcare data papers, inclusion/exclusion criteria, and data extraction), characterization of selected papers and algorithms (strategy to identify original algorithms, algorithm objective definition). Changes from the original protocol for each disease condition are reported below.
The search string used to retrieve articles in PubMed was composed of a part shared amongst all the examined conditions, reported in the protocol paper and focusing on Italian administrative healthcare data[13] and a disease-specific part, reported in box 1.
We chose to use a single database (PubMed/Medline) for the literature search, as we believe that the types of papers to be included in the systematic review are published in journals indexed in this database. Moreover, all the bibliographic references in the identified articles are checked and relevant studies not identified by the search string are included.
Two independent researchers screened the articles and classified pertinent ones, according to the objective for which the papers' algorithms were used. Inclusion criteria for a detailed data extraction of the algorithm were instead that the article used an original case-identification algorithm for any of the following purposes: **I** disease occurrence; **II** population/cohort selection; **III** outcome identification. Papers that used "secondary objectives" (**IV** to identify the disease as comorbidity for adjustments; **V** to identify the disease as exclusion criteria for other conditions; **VI** to calculate hospitalization rates or disease-specific drug prescription rates; **VII** other objectives) are expected to apply less elaborate algorithms, such as single-source algorithms (e.g., HDD to identify chronic conditions), so they were not considered for algorithm extraction.
With regard to diabetes, only algorithms selecting type 1 and/or type 2 diabetes were considered. As specified in the methodological paper,[13] articles regarding type 1/type 2 diabetes were retrieved using a unique search string, and results selecting any or both of those types of diabetes were considered together. Extracted algorithms were then divided into two categories: those selecting subjects with type 1 diabetes, and those selecting patients with type 2 diabetes or not distinguishing between the types. For each identified algorithm, the criteria to exclude other types of diabetes were extracted and highlighted.
Furthermore, since diabetes has been widely studied, and a considerable amount of results was retrieved, a number of changes in the selection process and inclusion/exclusion criteria were made to the methods reported in the protocol.[13] First, only bibliographic references of the most pertinent and recent articles were scanned (articles with objective I and II, published in 2015-2017), to search for possible relevant studies not identified by the search string. Second, after full-text screening, two exclusion criteria were added to those described in the methodological paper: papers whose algorithm identified only certain diabetic complications (e.g., hypoglycemia), and papers with algorithms with pharmacoepidemiologic purpose including only pharmacological sources. Changes were also made in distinguishing different algorithms: first, identical algorithms with different follow-back periods were considered as the same algorithm; second, different age cut-offs were not considered criteria to distinguish two otherwise identical algorithms.
Like for diabetes, papers on thyroid disorders were retrieved using a unique search string (see box 1) which comprised both generic terms (i.e., "thyroid disease", "thyroid disorders") and more specific terms referring to different nosological entities (i.e., "Goiter", "Graves' disease", "Hashimoto", etc.). Algorithms from extracted papers were subsequently categorized according to the type of thyroid disorder they considered. Thyroid neoplasms were not specifically included in the search string since they can be identified through cancer registries and therefore are beyond the scope of the present review.

RESULTS

DIABETES AND THYROID DISORDER PAPERS IDENTIFIED WITH THE SEARCH STRATEGY

The search strategy led to the identification of 780 articles for diabetes and 65 articles for thyroid disorders (table 1). Out of the selected articles, 584 and 58 papers, respectively for diabetes and thyroid disorders, were excluded by title and abstract. This brought to 196 and 7 full-text readings, resulting in 77 papers considered pertinent for diabetes, and 5 for thyroid disorders. With regard to diabetes, most of the excluded articles used data exclusively collected from disease registers (79 papers). Other reasons for exclusion were: no disease-specific algorithm reported (16), paper with pharmacoepidemiologic purpose only (9), exclusive use of General Practitioner (GP) data (6) or death certificates (2), algorithm not defined (5) and algorithm identifying only some diabetic conditions (2). For thyroid disorders, the reasons for

exclusion after full-text readings were: no disease specific algorithm reported (1) and algorithm not defined (1). References from the selected articles allowed the identification of 14 more works for diabetes, leading to a total of 91 pertinent papers. No additional works were retrieved from the references for thyroid disorders. Of the selected papers concerning diabetes, 68 had objectives **I-III** and, among those, 35 had at least one original algorithm. All pertinent papers concerning thyroid disorders had objectives I-III, and all had at least one original algorithm.

PERTINENT PAPERS ON DIABETES AND THYROID DISORDERS

While for diabetes 11 papers had been published before 2010, all pertinent papers concerning thyroid disorders were published after this year, and most of the papers concerning both diseases were published in the years 2014-2017 (58 out of 91 and 4 out of 5 for diabetes and thyroid disorders, respectively) (table 2). For hypo- and hyperthyroidism, most of the works focused on a region-wide setting, and only one paper was based on a national multicenter context. For diabetes, the situation was more heterogeneous, with 32 papers based on a sub-regional setting, 38 on a regional one, and 20 on a national multicenter context. Only 1 paper for diabetes was based on an international multicenter setting.

Most of the papers used administrative data for case identification that dated to 2008 or later; the data used for the analysis covered more than one year in 74 cases for diabetes and in 3 cases for thyroid disorders.

Articles that used at least one algorithm for objectives **I**, **II**, or **III**, were respectively 21, 39, and 8 for diabetes, 3, 1, and 1 for hypothyroidism, and one each for hyperthyroidism. The complete list and several characteristics of the papers, using algorithms for objectives I-III, can be found in the tables S1 and S2 (see on-line supplementary materials).

Prevalence for diabetes across papers that estimated the occurrence of the disease (objective **I**), ranged from 2.1% to 8.8% (crude rates) for type 2 diabetes, and from 0.12% to 0.2% for type 1 diabetes, with significant differences even within the same article, according to the temporal and quantitative characteristics of the algorithms used. Two works also estimated the annual incidence for type 2 diabetes at 0.55% and 0.4%, and five estimated the incidence for type 1, ranging from 13.4 to 25.2/100,000 (table S1). Among the 3 papers estimating the occurrence of thyroid diseases, one assessed the prevalence of all thyroid disorders at 4.39%, the second estimated both prevalence and annual incidence of hypothyroidism (at 3.1% and 0.69%, respectively), and the third quantified the prevalence of subclinical hypothyroidism in the pediatric population at 0.02% (table S2).

DIABETES ALGORITHMS

Out of the 68 papers focusing on objectives **I-III**, 29 original algorithms on type 2/not specified diabetes and 11 on type 1 diabetes were identified (table 3A). Among algorithms for type 2 diabetes or for unspecified diabetes, the most commonly used sources were the drug prescription database (DPD), hospital discharge record database (HDD), and exemption from healthcare co-payment database (ECD). DPD was considered in 27 out of 29 algorithms, and included codes for oral hypoglycemic drugs and insulins, in 5 cases with specific Defined Daily Doses (DDD). HDD was used in 23 algorithms, and codes with discharge diagnosis of diabetes (ICD code 250.xx or its equivalent DRG) or type 2 diabetes only were included. In 3 cases only the principal diagnosis was selected. Exemptions were part of the algorithm in 19 cases. Other sources were the ambulatory care service database (ACD), used in 5 algorithms and consisting in glycosylated hemoglobin (HbA1C) measurement (in all 5 cases) plus ACD diagnosis/exemption for diabetes (4 cases), and in one case the measurement of impaired fasting glucose (IFG). In 4 cases a clinical register was linked to the other sources to complete the algorithm, and in 2 cases both the birth register and the mortality record database (MRD) were used. Five of these algorithms used a single source (4 only DPD and 1 only HDD). Seven algorithms were specific for type 2 DM and excluded type 1 DM through one or more of the following exclusions: presence in a type 1 diabetes registry, specific exemption for type 1 DM, HDD codes specific for type 1 DM, only insulin codes in DPD, and by avoiding young people (the age cut-off was usually 30-35 years).

Among the 11 algorithms to identify type 1 diabetes, 9 used DPD (5 used all hypoglycemic drugs, while in 4 cases only insulins were selected), 7 the exemption database and 6 used HDD, in three cases specific for type 1 DM and in three cases for generic diabetes diagnosis, in any position. In one case, ACD codes were added. Three algorithms used only one source (ECD, DPD, and HDD, respectively). All these algorithms but one had either specific age range for case definition, always selecting children and young adults, or were applied to a population of young people (always under 35 years old). Four algorithms (including the one without age restriction) added other criteria to exclude type 2 diabetes: subjects were excluded if they ever took any oral hypoglycemic drug, or if they were not included in a type 1 DM register, or if they had a specific HDD, ACD, or MRD code for type 2 DM. In 6 algorithms specific selections were made not to include gestational diabetes, by excluding women with one or more of the following characteristics: being of childbearing age and receiving insulin only (without any prescription of oral hypoglycemic drug), having an obstetric code

	DIABETES MELLITUS		THYROID DISORDERS	
	Type 1	Type 2	Hypo thyroidism	Hyper thyroidism
Papers identified by the string	780		65	
Full-text readings	196		7	
Pertinent papers	77		5	
References added from bibliography	14		0	
Total pertinent papers	91		5	
Papers with objectives IV+*	23		0	
Papers with objectives I-III*	68		5	
Papers (with objective I-III) with at least one original algorithm	35		5	
Papers (with objective I-III) with external validation	2		0	
Original algorithms (with objective I-III)	**29**	**11**	**5**	**3**

* **I** to measure the occurrence of the disease; **II** to identify a population/cohort of subjects affected by the disease of interest; **III** to identify the disease as outcome; **IV** to identify the disease as comorbidity for statistical adjustments; **V** to identify the disease as exclusion criteria for other conditions; **VI** to calculate hospitalization rates or disease-specific drug prescription rates; **VII** other objectives

Table 1. Selection of papers published in PubMed between 2007 and 2017 and original algorithms included in the review according to the disease.

	DIABETES MELLITUS		THYROID DISORDERS	
	Type 1	Type 2	Hypo thyroidism	Hyper thyroidism
Year of Publication				
2007-2010	13		0	
2011-2013	20		1	
2014-2017	58		4	
Journal				
Italian	7		0	
International	84		5	
Setting				
Sub-regional (LHU, cities,..)	32		0	
Regional (entire region)	38		4	
National multicenter	20		1	
International multicenter	1		0	
Data time frame for the identification of the disease				
1 year	17		2	
> 1 year	74		3	
Use of data (even partial) following 2007 (≥2008)	60		4	
Objective*				
I	21[5-7,16,21,22,26,40-53]		3[9,16,54]	1[16]
II	39[8,14,15,23,25,55-87,116]		1[17]	1[17]
III	8[9,18,88-93]		1[18]	1[18]
IV	16[94-109]		0	0
V	1[110]		0	0
VI	1[111]		0	0
VII	5[37,112-115]		0	0

LHU: Local Health Unit
* **I** to measure the occurrence of the disease; **II** to identify a population/cohort of subjects affected by the disease of interest; **III** to identify the disease as outcome; **IV** to identify the disease as comorbidity for statistical adjustments; **V** to identify the disease as exclusion criteria for other conditions; **VI** to calculate hospitalization rates or disease-specific drug prescription rates; **VII** other objectives

Table 2. Characteristics of all pertinent papers published in PubMed between 2007 and 2017 included in the review according to the disease.

in HDD, being present in the birth register less than 10 months after diagnosis, having an exemption code specific for gestational DM.

The follow-back period for DPD and ECD codes was one year in almost all algorithms. The follow back for HDD varied from 1 to 10 years. As specified in the methods section, algorithms otherwise identical were not considered different algorithms if the follow-back frame was the unique modification. Seven algorithms for type 2, and 7 for type 1 diabetes reported an incident case definition.

External validation was performed for two algorithms only.[14,15] The first was a comparison of diabetic subjects identified only through HDD using the Tuscany Diabetes Registry (based on another algorithm),[5] the second was a composite algorithm compared with primary care medical records. The first study reported a sensitivity of 0.61-0.70 (depending on the discharging ward), while the second obtained a sensitivity of 0.71 (0.68-0.76), a specificity of 1 (0.99-1.00) and a positive predictive value (PPV) of 0.94 (0.92-0.97) (#2 and #28 in table 3A).

HYPO- AND HYPERTHYROIDISM ALGORITHMS

From the 5 papers on thyroid disorders, a total of 9 original algorithms were extracted (table 3B). Three papers[16–18] each presented two different algorithms for the identification of hypothyroidism and hyperthyroidism, while the other two papers focused on various forms of hypothyroidism. Two original algorithms from one paper relied on DPD as the only source, whilst the other 7 original algorithms combined DPD with other sources, i.e., HDD (in five cases), ECD (in three cases), and ACD (in one case).

One paper[9] developed a general algorithm (#7 in table 3B) for the identification of hypothyroidism (through DPD) and a more specific algorithm (#8 in table 3B) for retrieving iatrogenic hypothyroidism (combining DPD and records of previous thyroidectomy or thyroid irradiation from HDD and ACD); cases of autoimmune/iodine deficiency hypothyroidism were then derived by subtracting iatrogenic hypothyroidism cases from the whole hypothyroid population. Another paper pursued identification of subclinical hypothyroidism in the pediatric population (0-13 years) using a complex algorithm (#9 in table 3B) based on DPD (cumulative annual dose of levothyroxine below a predefined threshold) and ECD (as an exclusion criterion).

None of the identified algorithms was validated with an independent linked source. One paper[16] (algorithms #1,2 in table 3B) computed a cumulative prevalence estimate for all thyroid disorders (based on DPD) and then compared it with the estimates obtained from a national statistical survey and from the ECD. Since no details on ECD codes are presented in the paper, this is not considered an original algorithm in the present review. Comparison of the three sources showed very similar prevalence estimates for DPD (4.39%) and the national survey (4.54%) whilst ECD provided a much lower estimate (2.94%).

DISCUSSION

This work reviewed and described the characteristics of case-identification algorithms applied in the last decade in Italy for type 1 and type 2 diabetes and thyroid disorders. The number of algorithms proposed to identify diabetic individuals, mainly for type 2 diabetes, is considerably greater than the number retrieved for all other conditions included in this project. As already suggested, the reasons for this disparity can be the appropriateness of the diabetic condition for identification through administrative data, and the great impact of the disorder on the quality of life of individuals and public health management. The algorithms proposed for thyroid disorders, on the other hand, are few if compared with the wide occurrence of the disorder, particularly as regards hypothyroidism. Irrespective of the number of proposed algorithms, validation studies are lacking for both diseases.

ALGORITHMS FOR DIABETES

The algorithms identified for diabetes were quite similar to one another, with a predominant use of the three principal sources: DPD, HDD, and ECD. Algorithms on type 2 diabetes were often used in several papers: one of the first algorithms proposed,[5] in particular, was published in 2008, used in 16 subsequent papers, and slightly modified by other 5 (table 3A). On the contrary, algorithms on type 1 diabetes were often published and used only by the same research group. The main differences that distinguished one algorithm from another were: • the exclusion of one or more of the three main sources (particularly the eight algorithms that used only one source) or the inclusion of other less frequently utilized sources; • the exclusion of one or more types of diabetes (type 1 or 2, or gestational diabetes); • certain selections in the use of DPD, like the definition of minimum DDD, or the minimum number of yearly prescriptions to be included.

In Italy, the first attempts to identify diabetic subjects using one source were made before the first composite algorithms were proposed.[19,20] When the results provided by one-source algorithms were validated through linkage to the results of composite algorithms or other sources, however, sensitivity was over 70%,[14,20] suggesting that the combination of multiple sources might have given better results.[20] An exception seemed to be represented by algorithms using only DPD, whose occurrences have been compared to those obtained by other sources, giving similar results,[16,19] even though this type of algorithms cannot identify people with type 2 diabetes in early stages, when they do not need pharmacological treatment. Our results show that one-source-algorithms were chosen to estimate

occurrence of the disorder in selected populations (e.g., children with type 1 diabetes,[21] or aiming at identifying only pharmacologically treated diabetes),[22] but are not the most effective tool to estimate the occurrence of the disorder in the general population.

In a few algorithms, other sources were added to the three main ones, mainly obtained from laboratories (diagnostic tests performed and, in some cases, their results) or from clinical registries. Laboratory tests are recorded in administrative databases and could be added, in the future, to other local experiences. Furthermore, they have been easily and effectively used to monitor adherence to an adequate follow-up of diabetic people.[15,23] The results from the laboratory source (e.g., HbA1C or IFG value) could be very sensitive in identifying diabetic and even pre-diabetic subjects, and could make it possible to adopt the WHO diabetes definition.[24] Their introduction in some local experiences[7,25] showed that biochemistry results had the highest percentages of cases reported only by this source. However, since these laboratory data are not part of any national or regional administrative data flow, it would be hard to spread these local experiences to the national context. The linkage of the three main sources used in most algorithms can identify a good portion of people affected by the disease, as is demonstrated through capture-recapture models, which estimated a completeness of ascertainment of around 80%.[26] Nevertheless, the inclusion of less frequently used administrative sources, like laboratory tests or diagnostic procedures, could lead to more sensitive algorithms.

The selection or exclusion of some types of diabetes in the algorithm has been attempted by a few authors, mainly through the selection of specific HDD codes, or a combination of a DPD selection and the age groups included. Through these selections, results can be more precise, although the estimations would not change considerably, due to the predominant prevalence of type 2 diabetes (type 1 diabetes has been measured to account for around 4%-5% of total diabetic subjects).[7,8]

All these small differences between the algorithms are rarely accompanied by any clinical justification, nor is the choice of a specific algorithm supported by a validation of the algorithm itself.

With the exclusion of the validation of single sources,[14] described above, only one study included in the review validated the results obtained by an algorithm (#28 in table 3B), linking them to GP records,[15] with a very good PPV (94%) but suboptimal sensitivity (71%). To the best of our knowledge, only another paper validated an algorithm on diabetes in the Italian context, but it was published recently and we could not include it in the review.[27] This study validated the most commonly used algorithm for diabetes[5] (#4 in table 3A), linking its results to self-reported diagnosis of diabetes, with a sensitivity of 90.9%, a specificity of 97.4%, and a PPV of 70.9%. An attempt was also made to estimate the accuracy of the algorithm compared to the laboratory data of HbA1C >5.7%, but the sensitivity was considerably lower than the one obtained with the self-reported diagnosis. The results obtained through these studies are encouraging and support the hypothesis that administrative data are effective in identifying diabetic subjects. Nevertheless, since many different algorithms have been proposed, with small variations between each other, further efforts should be made in validation studies to identify the best algorithm for diabetes.

When considering the international experiences on this topic, many approaches have been proposed since the early Nineties to identify diabetic subjects with administrative data in various Countries around the world,[28-30] and their validity has been widely demonstrated and reviewed.[31-34] Some Countries have built national disease registries on composite sources, that enable them to have updated occurrence, morbidity, and mortality measures, and to monitor the disease follow-up.[35,36] Algorithms used in other Countries, however, can often rely on sources that are not available in Italy: information on diagnoses retrieved from GPs, specialists, and pediatricians, outpatient claims of ambulatory care and laboratory results available at a national level.[36,37] At the same time, the exemption from healthcare co-payment database is not available in other Countries.[37] Since evidence from experiences developed in other Countries is not likely suited to be implemented in the Italian context, we chose to focus on Italy.

ALGORITHMS FOR THYROID DISORDERS

With regard to thyroid disorders, all identified algorithms used the DPD, which is undoubtedly the reference source for detecting affected subjects, due to the specificity of the drugs used to treat these diseases. HDD and ECD were also included in some algorithms. Of interest is also the attempt in one algorithm to distinguish between the different conditions leading to hypothyroidism: the identification in the HDD and ACD databases of individuals who had undergone a procedure of thyroidectomy or irradiation as a marker for iatrogenic hypothyroidism, defining all other individuals as having autoimmune/iodine deficiency hypothyroidism.[9]

No validation study was included in the review. One paper only compared the estimates with those published by an external source, the survey performed in 2004-2005 by the National Institute of Statistics,[16] which provided the self-referred diagnosis of hypo- and hyperthyroidism. The study showed comparable estimates obtained from the survey and the DPD, while ECD gave a lower preva-

lence. The same recent paper that validated the algorithm for diabetes,[27] published too recently to be included in the review, validated one of the algorithms for hypothyroidism with a linkage to self-referred diagnosis, used as gold standard[9] (#7 in table 3B). The results showed a good specificity of the algorithm (91,8%) but a poor sensitivity (47,8%) and PPV (8,1%). For thyroid disorders a different gold standard, such as the altered results of laboratory tests, could be more affordable, and the same study found a significantly higher PPV (but a slightly lower sensitivity) when the algorithm was compared to this measure. The low sensitivity could be explained by two mutually acceptable reasons: the algorithm underestimates the prevalence because it does not detect subclinical hypothyroidism, and alterations of thyroid hormones can be related to acute conditions that don't necessarily lead to the diagnosis of hypothyroidism.

One of the limitations this systematic review could have encountered is related to the sensitivity of the search string: the string was made to maximize the inclusion of articles that used algorithms for objectives **I**, **II**, or **III**. Articles that relied on single-source algorithms might have been missed. In any case, 22 articles were selected for diabetes based on a single source (and 8 included original algorithms), while for thyroid disorders no papers were identified through the screening of the references.

The availability of administrative sources in Italy is wide, and comparable to other advanced international settings in which disease registries are available. The Italian Ministry of Health, through the national prevention plan[38] recommends the development of disease registries. Transposing local experiences to a wider, national setting could provide a useful instrument not only to estimate the occurrence of diseases, but also to monitor the quality of the health policies that are in place.

While the experiences of validation studies for diabetes are few but encouraging, the only attempt made to validate a thyroid disorder algorithm gave unsatisfactory results. For both diseases, but in particular for thyroid disorders, there is a need for an affordable gold standard, to make the comparison as reliable as possible. GP databases could enable this, since they have already been used for the validation of an algorithm for diabetes[15], and have recently been shown to be a valid instrument to identify people with specific diseases, including type 2 diabetes.[39]

CONCLUSION

Italian literature on the use of administrative healthcare data for case identification of diabetes is vast; the proposed algorithms are quite similar to one another, and the minor differences between them are rarely accompanied by clinical justification. On the contrary, the literature concerning thyroid disorders is relatively poor. The attempt made by Di Domenicantonio et al.,[27] which estimated the accuracy of the algorithms for both groups of diseases through their validation, is the first step to understand which algorithm could be the most appropriate, especially in relation to a specific research. Further validations of the proposed algorithms, as well as their further implementation, could be of great value.

Conflict of interest disclosure: none reported.

Funding disclosure: the "Algoritmi" Project was partially funded by the Department of Cardio-Thoraco-Vascular Sciences and Public Health (University of Padua) within the Projects BIRD (Integrated Budget for Research in Departments) in the year 2017.

Algorithm ID#	Author, year of publication of the original algorithm (following articles with the same algorithm)	Objective (I disease occurrence, II population-cohort selection, III outcome identification)	Type 1 - Type 2 diabetes mellitus	Identification of cases: incident (I)-prevalent (P)	SOURCES USED IN THE ALGORITHM				
					HDD ICD-9-CM code (Main diagnosis (M), Any diagnosis (A) Not reported (N))	ECD code	DPD code	ACD code	Other sources (code)
1	Brocco, 2007[6]	I	Type 2	I-P	–	Yes, not defined	A10A, A10B	–	–
2	Forastiere, 2007[88] Seghieri, 2011[14]	III	Type 2	P	250.xx (A)	–	–	–	–
3	Giarrizzo, 2007[40]	I	Type 2	P	250.xx (A)	Yes, not defined	–	Exemption for DM in ambulatory-care service (ACD); HbA1C measurement (LAB)	–
4	Gnavi, 2008b[5] Gnavi, 2008a[26] Gnavi, 2009[55] Giorda, 2012[23] Gini, 2013[43] Visca, 2013[61] Buja, 2014[46] Scalone, 2014[47] Gini, 2014[64] Anello, 2015[91] Fedeli, 2015[49] Pagano, 2015[71] Policardo, 2015[72] Barletta, 2016[74] Policardo, 2017a[83] Policardo, 2017b[84] Profili, 2017[85]	I	Type 2	P	250.xx (A)	13.250	A10A, A10B	–	–
5	Bruno, 2008[8] Bruno, 2012[58]	II	Type 2	P	250.xx (A)	Yes, not defined	A10A, A10B	–	Regional DM1 Clinical Register
6	Lonati, 2008[41]	I	Type 2	p	ICD9: 250.xx (M) OR DRG: 294 e 295	013* (and previous 0024)	A10A (at least 10% DDD), A10B (at least 30% DDD)	–	–
7	Nordio, 2009[56]	II	Type 2	I-P	–	–	A10A, A10B	–	–
8	Baviera, 2011[57]	II	Type 2	P	250.xx (M)	13.250	A10* (at least 30% DDD)	–	–
9	Chini, 2011[16]	I	Type 2	P	–	–	A10AB, A10AC, A10AD, A10AE, A10BA, A10BB, A10BD, A10BG, A10BX	–	–
10	Magoni, 2011[89]	III	Type 2	P	ICD9: 250.xx (A) OR DRG: 294 e 295	Yes, not defined	A10A (at least 10% DDD), A10B (at least 30% DDD)	–	Diagnosis of DM at admission in the Registers of residential psychiatric facilites
11	Marchesini, 2011[22] Marchesini, 2014[45]	I	Type 2	I-P	–	–	A10A, A10B	–	–
12	Monesi, 2011[42] Baviera,2014a[63] Monesi,2014[67]	I	Type 2	I-P	DRG 294, 295	013.250	A10* (at least 30% DDD)	–	–
13	De Berardis, 2012a[59]	II	Type 2	P	250.xx (A)	–	A10A, A10B	–	–
14	Fantini, 2012[60]	II	Type 2	P	250.xx (A)	–	(insulin or oral antidiabetic drugs)	–	Register "Progetto diabete"
15	Degli Esposti, 2013[25]	II	Type 2	P	250.xx (A)	–	A10	Exemption for DM in ACD Impairing fasting glucose (IFG)>126 mg-dL; Measurement of HbA1C (LAB)	–
16	Valent, 2013[14] Valent,2017a[86] Valent,2017b[87] Valent,2017c[93]	I	Type 2	P	250.xx (A)	P20, 013	A10A, A10B	–	–

CASE DEFINITION				EVALUATION OF THE ALGORITHM			
Algorithm	Age range (as definition criteria)	Exclusion of other types of diabetes (GEST= gestational DM, DM1= type1 Diabetes, DM2= type 2 diabetes)	Incidence: criteria for the exclusion of prevalent cases (look-back time frame)	Algorithm derived from a previous published one (reference)	Source used for external validation	Accuracy measures of external validation (Se, Sp, PPV, NPV*)	Sensitivity analysis (S) / contribution or coherence of the sources (C)
DPD (6 months) OR ECD. IF ONLY DPD, then ≥2DPD	NOT women 15-44 years	GEST: NOT Women 15-44 years 1≥ DPD for Insulin, OR temporary ET	1 year	–	–	–	–
HDD	–	–	–	–	In Seghieri 2011: Validation of HDD with Tuscany Diabetes Registry (based on Gnavi2008b algorithm)	Se: 60.5%-70.2% (depending on discharging ward)	–
HDD OR ECD OR LAB OR ACD	–	–	–	–	–	–	(C)
≥2 DPD at different times (year) OR ECD (in the year or previous 3 years) OR HDD (in the year or previous 4 years)	–	–	–	–	–	–	(C)
≥2 DPD (year) OR HDD (year) OR ECD (year)	if type of diabetes not defined, >30 years	DM1: NOT in DM1 Register AND, (if not defined) >30 years	–	From Gnavi 2008b	–	–	–
DPD OR ED OR HDD (yearly)	–	–	–	–	–	–	(C)
DPD	≥30 years	DM1: NOT <30 years AND DPD insulin ≥1 year	1 year	–	–	–	–
DPD OR ECD OR HDD (yearly)	–	–	–	–	–	–	–
DPD, at least 3 packages (year)	–	–	–	–	–	–	–
DPD OR ECD OR HDD, OR Other	–	–	–	From Lonati 2008	–	–	–
DPD (year)	–	–	1 year	–	–	–	–
DPD OR ECD OR HDD (year)	–	–	1-8 years (all the previous years)	–	–	–	(C)
≥2 DPD (year) OR HDD	–	–	–	–	–	–	–
HDD (2 years) OR DPD (year) OR being included in "Progetto diabete"	–	–	–	–	–	–	–
HDD OR DPD OR AMB OR LAB, NOT IF ONLY 1 HbA1C (year)	≥45 years	–	–	–	–	–	–
HDD (10 years) OR ECD OR ≥3 DPD (1 year)	–	–	–	–	–	–	–

Table 3A. Characteristics of diabetes mellitus case-identification algorithms published in PubMed between 2007 and 2017.

Algorithm ID#	Author, year of publication of the original algorithm (following articles with the same algorithm)	Objective (I disease occurrence, II population-cohort selection, III outcome identification)	Type 1 - Type 2 diabetes mellitus	Identification of cases: incident (I)-prevalent (P)	SOURCES USED IN THE ALGORITHM				
					HDD ICD-9-CM code (Main diagnosis (M), Any diagnosis (A) Not reported (N))	ECD code	DPD code	ACD code	Other sources (code)
17	Valent, 2013[44]	I	Type 2	P	250.xx (A)	P20, 013	A10A, A10B	HbA1C measurement (90.28.1) (LAB)	–
18	Ballotari, 2014[7] Ballotari 2017a[78]	I	Type 2	P	250.xx, 357.2x, 362.0x, 366.41, 648.0x (A) NOT MDC14	013.250	A10A, A10B	At least 1 value of HbA1C > 6,5% (LAB) Diabetes diagnosis in the ACD medical record	MRD (E10-E14) Birth Register (CEDAP)
19	Baviera, 2014b[62] Baviera, 2017a[81] Baviera, 2017b[80] Marcellusi, 2016[77] Pagano, 2016[116]	II	Type 2	P	250.xx (A)	013.250	A10*	–	–
20	Corrao, 2014[90]	III	Type 2	I	250.xx (M)	–	A10*	–	–
21	Giorda, 2014a[65] Giorda,2015[70]	II	Type 2	P	250.x1, 250.x3 (A)	–	A10A, A10B	–	–
22	Giorda, 2014b[66]	II	Type 2	P	250.xx (A)	13.250	A10A, A10B	–	–
23	Ballotari, 2015a[48] Ballotari, 2015b[68] Ballotari, 2017b[79]	I	Type 2	P	250.xx, 357.2x, 362.0x, 366.41, 648.0x (A) NOT MDC14	013.250	A10A, A10B	At least 1 value of HbA1C > 6,5% (LAB) Diabetes diagnosis in the ACD medical record	MRD (E10-E14) Birth Register (CEDAP)
24	Franchini, 2015[69]	II	Type 2	P	250.xx (A)	013	A10A (at least 10% DDD), A10B (at least 30% DDD)	–	–
25	Valent, 2015[73] Gini, 2016[76]	II	Type 2	P	250.xx (A)	P20, 013	A10A, A10B	–	–
26	Seghieri, 2016[92]	III	Type 2	I	–	–	A10A, A10B	–	–
27	D'Ovidio, 2017[82]	II	Type 2	P	–	Yes, not defined	A10A, A10B	–	–
28	Gini, 2017[15]	II	Type 2	P	250.x0; 250.x2 (A)	*250	A10A, A10B	–	–
29	Giorda, 2017[9]	III	Type 2	P	250.xx (A)	13.250	A10A, A10B	–	–
30	Bruno, 2008[8] Bruno, 2012[58]	II	Type 1	P	250.xx (A)	Yes, not defined	A10A, A10B	–	Regional DM1 Clinical Register

CASE DEFINITION				EVALUATION OF THE ALGORITHM			
Algorithm	**Age range (as definition criteria)**	**Exclusion of other types of diabetes (GEST= gestational DM, DM1= type1 Diabetes, DM2= type 2 diabetes)**	**Incidence: criteria for the exclusion of prevalent cases (look-back time frame)**	**Algorithm derived from a previous published one (reference)**	**Source used for external validation**	**Accuracy measures of external validation (Se, Sp, PPV, NPV*)**	**Sensitivity analysis (S) / contribution or coherence of the sources (C)**
(HDD (10 years) OR ECD OR ≥3 DPD (1 year)) OR (≥2 ACD (1 year)) OR (DPD (1 year) OR ACD (1 year))	–	–	–	–	–	–	–
HDD (1 year) OR ECD (1 year) OR LAB (1 year) OR ≥2 DPD OR ACD OR MRD (MRD will contribute for incident cases only)	if DM type not defined, then ≥19 years	DM1: NOT HDD 250.x1 OR 250.x3, NOT AMB IF DM1, IF not defined, THEN NOT <19 years GEST: NOT HDD MDC14 OR presence in CEDAP less than 10 months after 1 diagnosis of DM	–	–	–	–	(C) (S)
DPD OR ECD OR HDD	–	–	–	–	–	–	–
≥3 DPD OR HDD	–	–	No DPD OR HDD in any position, in the 3 years before the index event	–	–	–	–
DPD	–	DM1: NOT HDD 250.x1 OR 250.x3	–	–	–	–	–
≥2 DPD at separate times (1 year) OR ECD (in the year or previous 3 years) OR HDD (in the year or previous 4 years)	>35 years	GEST: NOT ≤35 years AND NOT IF 36-40 AND only insulin	–	Gnavi 2008	–	–	–
HDD (1 year) OR ECD (1 year) OR LAB (1 year) OR ≥2 DPD OR AMB OR MRD (MRD will contribute for incident cases only)	–	GEST: NOT HDD MDC14 OR CEDAP presence in CEDAP less than 10 months after 1 diagnosis of DM	–	Ballotari 2014	–	–	–
DPD OR ECD OR HDD (1 year)	–	–	–	–	–	–	–
HDD (10 years) OR ECD OR ≥3 DPD (1 year)	–	DM1: NOT DPD=ONLY A10A	–	Valent 2013	–	–	–
≥2 DPD, 1=in the first 30 days after index discharge, 2=6 following months	–	–	NOT in the DIABETES REGISTRY (see Gini 2013) before index discharge (for 6 years)	–	–	–	–
ECD OR ≥2 DPD	–	DM1: NOT DPD=ONLY A10A	–	–	–	–	–
HDD (4 years) OR ECD (3 years) OR ≥2 DPD (A10A) (2 years) OR ≥2 DPD (A10B) 2 years	–	–	–	Gnavi 2008	In: Gini R. "A Validation Odyssey: From big data to local intelligence", 2016 primary care medical records	Se: 0.71 (0.68-0.76) Sp: 1.00 (0.99-1.00) PPV: 0.94 (0.92-0.97)	–
≥2 DPD at different times (year) OR ECD (in the year or previous 3 years) OR HDD (in the year or previous 4 years)	≥35 if ≥2 A10A; NOT Women 20-45 years AND ≥1 DPD	DM1: NOT ET=DM1 OR if not in ET: OR (SDO=250.x1 OR 250.x3 in the previous 5 years), OR (2≥A10A in the previous 12 months AND ≤35). GEST: women 20-45 years AND ECD=Gestational DM)	–	Gnavi 2008	–	–	–
≥2 DPD OR HDD (year) OR ECD AND (in the DM1 register OR, if not defined, <30 years)	if type of diabetes not defined, <30 years	DM2: in DM1 register OR if not defined, <30 years	–	–	–	–	–

Table 3A. Characteristics of diabetes mellitus case-identification algorithms published in PubMed between 2007 and 2017.

Algorithm ID#	Author, year of publication of the original algorithm (following articles with the same algorithm)	Objective (I disease occurrence, II population-cohort selection, III outcome identification)	Type 1 - Type 2 diabetes mellitus	Identification of cases: incident (I)-prevalent (P)	SOURCES USED IN THE ALGORITHM				
					HDD ICD9-CM code (Main diagnosis (M), Any diagnosis (A) Not reported (N))	ECD code	DPD code	ACD code	Other sources (code)
31	Ballotari, 2014[47]	I	Type 1	P	250.xx, 357.2x, 362.0x, 366.41, 648.0x (A) NOT MDC14	013.250	A10A, A10B	At least 1 value of HbA1C > 6,5% (LAB) Diabetes diagnosis in the ACD medical record	MRD (E10-E14) Birth Register (CEDAP)
32	Vichi, 2014[21]	I	Type 1	I	250.x1, 250.x3 (A)	–	–	–	–
33	Marigliano, 2015[50]	I	Type 1	I-P	–	Yes, not defined	–	–	–
34	Bianchi, 2016[75]	II	Type 1	I	–	013	A10A (at least 2000 U-year)	–	–
35	Bruno, 2016[51]	I	Type 1	I-P	–	–	A10A	–	–
36	Canova, 2016[18]	III	Type 1	I	250.xx (A)	–	A101A	–	–
37	Fortunato, 2016[52]	I	Type 1	I	250.x1; 250.x3 (A)	013	A10A	–	–
38	Valent, 2016[53]	II	Type 1	I-P	250.xx (A)	013	A10A, A10B	–	–
39	D'Ovidio, 2017[82]	II	Type 1	P	-	Yes, not defined	A10A, A10B	–	–
40	Giorda, 2017[9]	III	Type 1	P	250.x1, 250.x3 (A)	Yes, not defined	A10A, A10B	–	–

HDD: hospital discharge record database; **ECD:** exemption from healthcare co-payment database; **DPD:** drug prescription database; **ACD:** ambulatory care-services database; **LAB:** laboratory database; **MRD:** mortality registry database; **DDD:** Defined Daily Dose; **HbA1C:** glycosylated haemoglobin; **Se:** sensitivity; **Sp:** specificity; **PPV:** positive predictive value; **NPV:** negative predictive value

Table 3A. Characteristics of diabetes mellitus case-identification algorithms published in PubMed between 2007 and 2017.

CASE DEFINITION			Incidence: criteria for the exclusion of prevalent cases (look-back time frame)	EVALUATION OF THE ALGORITHM			
Algorithm	Age range (as definition criteria)	Exclusion of other types of diabetes (GEST= gestational DM, DM1= type1 Diabetes, DM2= type 2 diabetes)		Algorithm derived from a previous published one (reference)	Source used for external validation	Accuracy measures of external validation (Se, Sp, PPV, NPV*)	Sensitivity analysis (S) / contribution or coherence of the sources (C)
HDD (1 year) OR ECD (1 year) OR LAB (year) OR ≥2 DPD OR AMB OR MRD (MRD will contribute for incident cases only)	if DM type not defined, then ≤19 years	DM2: NOT (HDD 250.x0 OR 250.x2), NOT AMB if DM2, NOT DPD=A10B, NOT CM=E11, IF not defined, THEN NOT ≥19 years GEST: NOT HDD MDC14 OR CEDAP less than 10 months after the first DM diagnosis	–	–	–	–	(C) (S)
HDD	0-4 years	(syndromic diabetes, secondary diabetes, other non-type 1 diabetes): NOT <180 days, NOT HDD 775.1, 204.0+V58.1, 253.3, 588.1, 775.0, 996.85, 277.0, 271.4, 271.0, 759.81, 334.3, 758.6, V42.7, 206.0+V58.1, 790.29, 138.75, 577.8, 157.2, 255.3, 194.0, 237.5, 255.0, 277.87, 577.8, 330.0, 284.09	4 years	–	–	–	–
ECD	0-18 years	–	ECD in 2008-2013 not present before 2008.	–	–	–	–
ECD OR DPD (1 year)	1-17 years	–	12 months without DPD	–	–	–	–
≥2 DPD (1 year) and continuously treated the following year	0-29 years	–	1 year	–	–	–	–
HDD OR DPD	<18 years	–	Birth Cohorts: diagnoses only after the index date	–	–	–	–
HDD (10 years) OR ECD OR DPD (10 years)	<18 years	–	5 years	–	–	–	(C) (S)
HDD (10 years in Valent 2015) OR ECD OR ≥3 DPD (1 year)	0-18 years	DM2: included only if DPD=ONLYA10A	1 year	–	–	–	–
ECD OR ≥2 DPD	–	DM2: NOT IF DPD= A10B OR A10X	–	–	–	–	–
ECD OR HDD (5 years) OR ≥2 DPD (1 year)	<35 years	–	–	–	–	–	–

Algorithm ID#	Author, year of publication of the original algorithm (following articles with the same algorithm)	Objective (I disease occurrence, II population-cohort selection, III outcome identification)	Hypo- Hyperthyroidism	Identification of cases: incident (I)-prevalent (P)	SOURCES USED IN THE ALGORITHM			
					HDD ICD9-CM code (Main diagnosis (M), Any diagnosis (A) Not reported (N))	ECD code	DPD code	Other sources (code)
1	Chini, 2011[16]	I	Hyperthyroidism	P	–	–	H03BB, H03BC	–
2	Chini, 2011[16]	I	Hypothyroidism	P	–	–	H03AA	–
3	Maccagnano, 2016[17]	II	Hyperthyroidism	P	242.0, 242.1, 242.2, 242.3 (M)	035	H03B, H03BA, H03BA01, H03BA02, H03BA03, H03BB, H03BB01, H03BB52, H03BC, H03BC01, H03BX, H03BX01, H03BX02, H03CA	–
4	Maccagnano, 2016[17]	II	Hypothyroidism	P	243, 244.1, 244.2, 244.3, 244.8 (M)	027	H03AA01, H03AA02, H03AA03, H03AA04, H03AA05	–
5	Canova, 2016[18]	III	autoimmune hyperthyroidism	I	242.x (A)	–	H03BB	–
6	Canova, 2016[18]	III	autoimmune hypothyroidism	I	245.x (A)	–	H03AA	–
7	Giorda, 2017[9]	I	Hypothyroidism	I-P	–	–	H03AA01, H03AA02, H03AA03	–
8	Giorda, 2017[9]	I	Iatrogenic hypothyroidism	I-P	06.2, 06.3, 06.4, 92.28	–	H03AA01, H03AA02, H03AA03	ACD: 06.2, 06.3, 06.4, 92.28
9	Greggio, 2017[54]	I	Subclinical pediatric hypothyroidism	P	–	027, 056, 048	H03AA01	–

HDD: hospital discharge record database; **ECD:** exemption from healthcare co-payment database; **DPD:** drug prescription database; **ACD:** ambulatory care-services database; **Se:** sensitivity; **Sp:** specificity; **PPV:** positive predictive value; **NPV:** negative predictive value

Table 3B. Characteristics of thyroid disorders case-identification algorithms published in PubMed between 2007 and 2017.

CASE DEFINITION		Incidence: criteria for the exclusion of prevalent cases (look-back time frame)	EVALUATION OF THE ALGORITHM			
Algorithm	Age range (as definition criteria)		Algorithm derived from a previous published one (reference)	Source used for external validation	Accuracy measures of external validation (Se, Sp, PPV, NPV*)	Sensitivity analysis (S)/ contribution or coherence of the sources (C)
≥3 DPD (1 year)	–	–	Von Korff M, J Clin Epidemiol 1992	–	–	–
≥3 DPD (1 year)	–	–	Von Korff M, J Clin Epidemiol 1992	–	–	–
HDD (1year) OR [DPD (1year) AND ECD (1year)]	–	–	–	–	–	–
HDD (1year) OR [DPD (1year) AND ECD (1year)]	–	–	–	–	–	–
HDD OR DPD	–	excluding cases who matched the disease definition before the index date (starting from birth)	Marrie RA, Neuroepidemiology 2012	–	–	–
HDD OR DPD	–	excluding cases who matched the disease definition before the index date (starting from birth)	Marrie RA, Neuroepidemiology 2012	–	–	–
≥2 DPD (1year)	–	excluding cases which had received at least one DPD in the previous year (H03AA01, H03AA02, H03AA03)	–	–	–	–
≥2 DPD (1year) AND [HDD (previous 5 years) OR ACD (previous 5 years)]	–	Excluding cases which had received at least one DPD in the previous year (H03AA01, H03AA02, H03AA03)	–	–	–	–
DPD (low dose) NOT ECD	0-13 years	–	–	–	–	–

REFERENCES

1. World Health Organization, eds. Global Report on Diabetes. Geneva, Switzerland: WHO; 2016.
2. Gnavi R, Migliardi A, Maggini M, Costa G. Prevalence of and secular trends in diagnosed diabetes in Italy: 1980-2013. Nutr Metab Cardiovasc Dis 2018;28(3): 219-25.
3. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Research and Clinical Practice 2014;103(2):137-49.
4. Taylor PN, Albrecht D, Scholz A, et al. Global epidemiology of hyperthyroidism and hypothyroidism. Nature Reviews Endocrinology 2018;14(5):301-16.
5. Gnavi R, Karaghiosoff L, Balzi D, et al. Diabetes prevalence estimated using a standard algorithm based on electronic health data in various areas of Italy. Epidemiol Prev 2008;32(3) Suppl:15-21.
6. Brocco S, Visentin C, Fedeli U, et al. Monitoring the occurrence of diabetes mellitus and its major complications: the combined use of different administrative databases. Cardiovasc Diabetol 2007;6:5.
7. Ballotari P, Ranieri SC, Vicentini M, et al. Building a population-based diabetes register: an Italian experience. Diabetes Res Clin Pract 2014;103(1):79-87.
8. Bruno G, Karaghiosoff L, Merletti F, et al. The impact of diabetes on prescription drug costs: the population-based Turin study. Diabetologia 2008;51(5):795-801.
9. Giorda CB, Carna P, Romeo F, Costa G, Tartaglino B, Gnavi R. Prevalence, incidence and associated comorbidities of treated hypothyroidism: an update from a European population. Eur J Endocrinol 2017;176(5):533-42.
10. Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ. Epidemiology of Type 1 Diabetes. Endocrinol Metab Clin North Am 2010;39(3):481-97.
11. De Leo S, Lee SY, Braverman LE. Hyperthyroidism. Lancet 2016;388(10047): 906-18.
12. Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. Lancet 2017;390 (10101):1550-62.
13. Canova C, Simonato L, Barbiellini Amidei C, et al. A systematic review of case-identification algorithms for 18 conditions based on Italian healthcare administrative databases: a study protocol. Epidemiol Prev 2019;43(4) Suppl 2:8-16.
14. Seghieri C, Francesconi P, Cipriani S, et al. Gender effect on the relation between diabetes and hospitalization for heart failure. Exp Clin Endocrinol Diabetes 2012;120(1):51-55.
15. Gini R, Schuemie MJ, Pasqua A, et al. Monitoring compliance with standards of care for chronic diseases using healthcare administrative databases in Italy: Strengths and limitations. Manzoli L, ed. PLoS One 2017;12(12):e0188377.
16. Chini F, Pezzotti P, Orzella L, Borgia P, Guasticchi G. Can we use the pharmacy data to estimate the prevalence of chronic conditions? a comparison of multiple data sources. BMC Public Health 2011;11:688.
17. Maccagnano G, Notarnicola A, Pesce V, Mudoni S, Tafuri S, Moretti B. The Prevalence of Fragility Fractures in a Population of a Region of Southern Italy Affected by Thyroid Disorders. Biomed Res Int 2016;2016:6017165.
18. Canova C, Pitter G, Ludvigsson JF, et al. Celiac Disease and Risk of Autoimmune Disorders: A Population-Based Matched Birth Cohort Study. J Pediatr 2016;174: 146-52.
19. Maio V, Yuen E, Rabinowitz C, et al. Using pharmacy data to identify those with chronic conditions in Emilia Romagna, Italy. J Health Serv Res Policy 2005;10(4):232-38.
20. Gnavi R, Karaghiosoff L, Dalmasso M, Bruno G. The Piedmont Diabetes Register: opportunities and limits for epidemiological purposes. Epidemiol Prev 2006;30(1):59-64.
21. Vichi M, Iafusco D, Galderisi A, Stazi MA, Nistico L. An easy, fast, effective tool to monitor the incidence of type 1 diabetes among children aged 0-4 years in Italy: the Italian Hospital Discharge Registry (IHDR). Acta Diabetol 2014;51(2):287-94.
22. Marchesini G, Forlani G, Rossi E, Berti A, De Rosa M; ARNO Working Group. The direct economic cost of pharmacologically-treated diabetes in Italy-2006. The ARNO observatory. Nutr Metab Cardiovasc Dis 2011;21(5):339-46.
23. Giorda C, Picariello R, Nada E, et al. The Impact of Adherence to Screening Guidelines and of Diabetes Clinics Referral on Morbidity and Mortality in Diabetes. Atkin SL, ed. PLoS One 2012;7(4):e33839.
24. World Health Organization, International Diabetes Federation. Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia: Report of a WHO/IDF Consultation. Geneva: WHO; 2006. Available from: http://www.who.int/diabetes/publications/diagnosis_diabetes2006/en/ (last accessed: 26.11.2018).
25. Esposti LD, Saragoni S, Buda S, Sturani A, Esposti ED. Glycemic control and diabetes-related health care costs in type 2 diabetes; retrospective analysis based on clinical and administrative databases. Clinicoecon Outcomes Res 2013;5:193-201.
26. Gnavi R, Karaghiosoff L, Costa G, Merletti F, Bruno G. Socio-economic differences in the prevalence of diabetes in Italy: the population-based Turin study. Nutr Metab Cardiovasc Dis 2008;18(10):678-82.
27. Di Domenicantonio RD, Cappai G, Cascini S, et al. Validation of algorithms for the identification of subjects with chronic disease using health information systems. Epidemiol Prev 2018;42(5-6):316-25.
28. Morris AD, Boyle DI, MacAlpine R, et al. The diabetes audit and research in Tayside Scotland (darts) study: electronic record linkage to create a diabetes register. BMJ 1997;315(7107):524-58.
29. Lipscombe LL, Hux JE. Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995-2005: a population-based study. Lancet 2007;369(9563): 750-56.
30. Costa G, Salmaso S, Cislaghi C (eds). Misurare in sanità pubblica: registri e sorveglianza. Situazione attuale e prospettive. Roma: Istituto Superiore di Sanità. 2014. Available from: http://www.iss.it/publ/index.php?lang=1&id=2849&tipo=5 (last accessed: 10.12.2018).
31. Saydah SH, Geiss LS, Tierney E, Benjamin SM, Engelgau M, Brancati F. Review of the performance of methods to identify diabetes cases among vital statistics, administrative, and survey data. Ann Epidemiol 2004;14(7):507-16.
32. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic Review and Meta-Analysis of Validation Studies on a Diabetes Case Definition from Health Administrative Records. PLoS One 2013;8(10):e75256.
33. Shiff NJ, Jama S, Boden C, Lix LM. Validation of administrative health data for the pediatric population: a scoping review. BMC Health Serv Res 2014;14(1):236.
34. Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. BMJ Open 2016;6(8):e009952.
35. Public Health Agency of Canada. Diabetes in Canada: Facts and figures from a public health perspective. Ottawa 2011. Available from: http://www.phac-aspc.gc.ca/cd-mc/publications/diabetes-diabete/facts-figures-faits-chiffres-2011/pdf/facts-figures-faits-chiffres-eng.pdf
36. Gudbjornsdottir S, Cederholm J, Nilsson P, Eliasson B; Steering Committee of the Swedish National Diabetes Register. The national diabetes register in Sweden: an implementation of the St. Vincent Declaration for Quality Improvement in Diabetes Care. Diabetes Care 2003;26(4):1270-76.
37. 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.
38. Ministero della Salute. Piano Nazionale della Prevenzione 2014-2018.
39. Gini R, Schuemie MJ, Mazzaglia G, et al. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study. BMJ Open 2016;6(12):012413.
40. Giarrizzo ML, Pezzotti P, Silvestri I, Lallo DD. Estimating prevalence of diabetes mellitus in a Lazio province, Italy, by capture-recapture models. Epidemiol Prev 2007;31(6):333-39.
41. Lonati F, Scarcella C, Indelicato A, et al. Brescia Local Health Autority Population Database: a method based on current data for monitoring chronic diseases and management. Epidemiol Prev 2008;32(3):137-44.
42. Monesi L, Baviera M, Marzona I, et al. Prevalence, incidence and mortality of diagnosed diabetes: evidence from an Italian population-based study: Prevalence, incidence and mortality of diabetes. Diabet Med 2012;29(3):385-92.
43. Gini R, Francesconi P, Mazzaglia G, et al. Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey. BMC Public Health 2013;13:15.
44. Valent F, Tillati S, Zanier L. Prevalence and comorbidities of known diabetes in northeastern Italy. J Diabetes Investig 2013;4(4):355-60.
45. Marchesini G, Bernardi D, Miccoli R, et al. Under-treatment of migrants with diabetes in a universalistic health care system: the ARNO Observatory. Nutr Metab Cardiovasc Dis 2014;24(4):393-99.
46. Buja A, Gini R, Visca M, et al. Need and disparities in primary care management of patients with diabetes. BMC Endoc Disord 2014;14:56.
47. Scalone L, Cesana G, Furneri G, et al. Burden of diabetes mellitus estimated with a longitudinal population-based study using administrative databases. PloS One 2014;9(12):e113741.
48. Ballotari P, Caroli S, Ferrari F, et al. Differences in diabetes prevalence and inequalities in disease management and glycaemic control by immigrant status: a population-based study (Italy). BMC Public Health 2015;15:87.
49. Fedeli U, Casotto V, Ferroni E, Saugo M, Targher G, Zoppini G. Prevalence of diabetes across different immigrant groups in North-eastern Italy. Nutr Metabol Cardiovasc Dis 2015;25(10):924-30.
50. Marigliano M, Tadiotto E, Morandi A, et al. Epidemiology of type 1 diabetes mellitus in the pediatric population in Veneto Region, Italy. Diabetes Res Clin Pract 2015;107(3):e19-21.

51. Bruno G, Gruden G, Songini M. Incidence of type 1 diabetes in age groups above 15 years: facts, hypothesis and prospects for future epidemiologic research. Acta Diabetol 2016;53(3):339-47.
52. Fortunato F, Cappelli MG, Vece MM, et al. Incidence of Type 1 Diabetes among Children and Adolescents in Italy between 2009 and 2013: The Role of a Regional Childhood Diabetes Registry. J Diabetes Res 2016;2016:7239692.
53. Valent F, Candido R, Faleschini E, et al. The incidence rate and prevalence of pediatric type 1 diabetes mellitus (age 0-18) in the Italian region Friuli Venezia Giulia: population-based estimates through the analysis of health administrative databases. Acta Diabetol 2016;53(4):629-35.
54. Greggio NA, Rossi E, Calabria S, et al. Subclinical hypothyroidism in paediatric population treated with levothyroxine: a real-world study on 2001-2014 Italian administrative data. Endocr Connect 2017;6(6):367-74.
55. Gnavi R, Picariello R, la Karaghiosoff L, Costa G, Giorda C. Determinants of Quality in Diabetes Care Process: The population-based Torino Study. Diabetes Care 2009;32(11):1986-92.
56. Nordio M, Antonucci F, Feriani M, Inio A, Marchini P. Reliability of administrative databases in epidemiological research: the example of end-stage renal disease requiring renal replacement therapy in patients with diabetes. G Ital Nefrol 2009;26 Suppl 45:S7-11.
57. Baviera M, Monesi L, Marzona I, et al. Trends in drug prescriptions to diabetic patients from 2000 to 2008 in Italy's Lombardy Region: a large population-based study. Diabetes Res Clin Pract 2011;93(1):123-30.
58. Bruno G, Picariello R, Petrelli A, et al. Direct costs in diabetic and non diabetic people: the population-based Turin study, Italy. Nutr Metab Cardiovasc Dis 2012;22(8):684-90.
59. Berardis GD, D'Ettorre A, Graziano G, et al. The burden of hospitalization related to diabetes mellitus: a population-based study. Nutr Metab Cardiovasc Dis 2012;22(7):605-12.
60. Fantini MP, Compagni A, Rucci P, Mimmi S, Longo F. General practitioners' adherence to evidence-based guidelines: a multilevel analysis. Health Care Manage Rev 2012;37(1):67-76.
61. Visca M, Donatini A, Gini R, et al. Group versus single handed primary care: a performance evaluation of the care delivered to chronic patients by Italian GPs. Health Policy 2013;113(1-2):188-98.
62. Baviera M, Cortesi L, Tettamanti M, et al. Changes in prescribing patterns and clinical outcomes in elderly diabetic patients in 2000 and 2010: analysis of a large Italian population-based study. Eur J Clin Pharmacol 2014;70(8):965-74.
63. Baviera M, Santalucia P, Cortesi L, et al. Sex differences in cardiovascular outcomes, pharmacological treatments and indicators of care in patients with newly diagnosed diabetes: Analyses on administrative database. Eur J Int Med 2014;25(3):270-75.
64. Gini R, Schuemie MJ, Francesconi P, et al. Can Italian healthcare administrative databases be used to compare regions with respect to compliance with standards of care for chronic diseases? PloS One 2014;9(5):e95419.
65. Giorda CB, Picariello R, Nada E, et al. Incretin therapies and risk of hospital admission for acute pancreatitis in an unselected population of European patients with type 2 diabetes: a case-control study. Lancet Diabetes Endocrinol 2014; 2(2):111-15.
66. Giorda CB, Picariello R, Nada E, et al. Comparison of direct costs of type 2 diabetes care: different care models with different outcomes. Nutr Metab Cardiovasc Dis 2014;24(7):717-24.
67. Monesi L, Tettamanti M, Cortesi L, et al. Elevated risk of death and major cardiovascular events in subjects with newly diagnosed diabetes: findings from an administrative database. Nutr Metab Cardiovasc Dis 2014;24(3):263-70.
68. Ballotari P, Ranieri SC, Luberto F, et al. Sex differences in cardiovascular mortality in diabetics and nondiabetic subjects: a population-based study (Italy). Int J Endocrinol 2015;2015:914057.
69. Franchini M, Pieroni S, Fortunato L, Molinaro S, Liebman M. Poly-pharmacy among the elderly: analyzing the co-morbidity of hypertension and diabetes. Curr Pharm Des 2015;21(6):791-805.
70. Giorda CB, Picariello R, Tartaglino B, et al. Hospitalisation for heart failure and mortality associated with dipeptidyl peptidase 4 (DPP-4) inhibitor use in an unselected population of subjects with type 2 diabetes: a nested case-control study. BMJ Open 2015;5(6):007959.
71. Pagano E, Petrelli A, Picariello R, Merletti F, Gnavi R, Bruno G. Is the choice of the statistical model relevant in the cost estimation of patients with chronic diseases? An empirical approach by the Piedmont Diabetes Registry. BMC Health Serv Res 2015;15:582.
72. Policardo L, Seghieri G, Francesconi P, et al. Gender difference in diabetes-associated risk of first-ever and recurrent ischemic stroke. J Diabetes Complications 2015;29(5):713-17.
73. Valent F. Diabetes mellitus and cancer of the digestive organs: An Italian population-based cohort study J Diabetes Complications 2015;29(8):1056-61.
74. Barletta V, Profili F, Gini R, et al. Impact of Chronic Care Model on diabetes care in Tuscany: a controlled before-after study. Eur J Public Health 2016;27(1):8-13.
75. Bianchi M, Cartabia M, Clavenna A, et al. Serological screening for celiac disease in a northern Italian child and adolescent population after the onset of type 1 diabetes: a retrospective longitudinal study of a 7-year period. Eur J Gastroenterol Hepatol 2016;28(6):696-701.
76. Gini A, Bidoli E, Zanier L, et al. Cancer among patients with type 2 diabetes mellitus: A population-based cohort study in northeastern Italy. Cancer Epidemiol 2016;41:80-87.
77. Marcellusi A, Viti R, Sciattella P, et al. Economic aspects in the management of diabetes in Italy. BMJ Open Diabetes Res Care 2016;4(1):e000197.
78. Ballotari P, Venturelli F, Greci M, Rossi PG, Manicardi V. Sex Differences in the Effect of Type 2 Diabetes on Major Cardiovascular Diseases: Results from a Population-Based Study in Italy. Int J Endocrinol 2017;2017:6039356.
79. Ballotari P, Vicentini M, Manicardi V, et al. Diabetes and risk of cancer incidence: results from a population-based cohort study in northern Italy. BMC Cancer 2017;17(1):703.
80. Baviera M, Avanzini F, Marzona I, et al. Cardiovascular complications and drug prescriptions in subjects with and without diabetes in a Northern region of Italy, in 2002 and 2012. Nutr Metab Cardiovasc Dis 2017;27(1):54-62.
81. Baviera M, Roncaglioni MC, Tettamanti M, et al. Diabetes mellitus: a risk factor for seizures in the elderly-a population-based study. Acta Diabetol 2017;54(9):863-70.
82. D'Ovidio F, d'Errico A, Carna P, Calvo A, Costa G, Chio A. The role of pre-morbid diabetes on developing Amyotrophic Lateral Sclerosis. Eur J Neurol 2018;25(1):164-70.
83. Policardo L, Seghieri G, Anichini R, Francesconi P. Effect of statins on hospitalization risk of bacterial infections in patients with or without diabetes. Acta Diabetol 2017;54(7):669-75.
84. Policardo L, Seghieri G, Francesconi P, Anichini R, Franconi F, Prato SD. Gender difference in diabetes related excess risk of cardiovascular events: When does the "risk window" open? J Diabetes Complications 2017;31(1):74-79.
85. Profili F, Bellini I, Zuppiroli A, Seghieri G, Barbone F, Francesconi P. Changes in diabetes care introduced by a Chronic Care Model-based programme in Tuscany: a 4-year cohort study. Eur J Public Health 2017;27(1):14-19.
86. Valent F, Tonutti L, Grimaldi F. Does diabetes mellitus comorbidity affect in-hospital mortality and length of stay? Analysis of administrative data in an Italian Academic Hospital. Acta Diabetol 2017;54(12):1081-90.
87. Valent F. Falls requiring visit to emergency room in a population-based cohort of diabetic patients in Italy. J Inj Violence Res 2017;9(2):83-90.
88. Forastiere F, Stafoggia M, Tasco C, et al. Socioeconomic status, particulate air pollution, and daily mortality: differential exposure or differential susceptibility. Am J Ind Med 2007;50(3):208-16.
89. Magoni M, Scarcella C, Vassallo F, et al. The evolving burden of HIV infection compared with other chronic diseases in northern Italy. HIV Med 2011;12(3):129-37.
90. Corrao G, Ibrahim B, Nicotra F, et al. Statins and the Risk of Diabetes: Evidence From a Large Population-Based Cohort Study. Diabetes Care 2014;37(8):2225-32.
91. Anello P, Cestari L, Canova C, et al. Investigation on the health effects of traffic-related air pollution from Mestre motorway (Veneto Region, Northern Italy). Epidemiol Prev 2015;39(4):251-60.
92. Seghieri G, Policardo L, Profili F, Francesconi P, Anichini R, Prato SD. Hospital incidental diagnosis of diabetes: A population study. J Diabetes Complications 2016;30(3):457-61.
93. Valent F, Busolin A, Boscutti G. Inception and utility of a renal replacement registry using administrative health data in North-East Italy. J Ren Care 2017;43(2):121-27.
94. Gnavi R, Migliardi A, Demaria M, Petrelli A, Caprioglio A, Costa G. Statins prescribing for the secondary prevention of ischaemic heart disease in Torino, Italy. A case of ageism and social inequalities. Eur J Public Health 2007;17(5):492-96.
95. Corrao G, Zambon A, Parodi A, et al. Do socioeconomic disparities affect accessing and keeping antihypertensive drug therapy? Evidence from an Italian population-based study. J Hum Hypertens 2009;23(4):238-44.
96. Agabiti N, Belleudi V, Davoli M, et al. Profiling hospital performance to monitor the quality of care: the case of COPD. Eur Respir J 2010;35(5):1031-38.
97. Corrao G, Parodi A, Nicotra F, et al. Better compliance to antihypertensive medications reduces cardiovascular risk. J Hypertens 2011;29(3):610-18.
98. Esposti LD, Sangiorgi D, Arca M, Vigna GB, Budal S, Esposti ED. Achievement of therapeutic target in subjects on statin treatment in clinical practice. Results of the STAR (Statins Target Assessment in Real practice) study. Monaldi Arch Chest Dis 2011;76(4):160-67.
99. Berardis GD, Lucisano G, D'Ettorre A, et al. Association of aspirin use with major bleeding in patients with and without diabetes. JAMA 2012;307(21):2286-94.
100. Macchia A, Romero M, D'Ettorre A, Mariani J, Tognoni G. Temporal trends of the gaps in post-myocardial infarction secondary prevention strategies of co-morbid and elderly populations vs. younger counterparts: an analysis of three successive cohorts between 2003 and 2008. Eur Heart J 2012;33(4):515-22.

101. Sousa LP de, Burba I, Ruperto C, Lattuada L, Barbone F, Chiara AD. Vitamin K antagonists in patients with nonvalvular atrial fibrillation: appropriateness and quality of treatment in an Italian cohort. J Cardiovasc Med (Hagerstown) 2013;14(7):534-40.
102. Bonifazi M, Franchi M, Rossi M, et al. Long term survival of HER2-positive early breast cancer treated with trastuzumab-based adjuvant regimen: a large cohort study from clinical practice. Breast 2014;23(5):573-78.
103. Vinceti M, Malagoli C, Rothman KJ, et al. Risk of birth defects associated with maternal pregestational diabetes. Eur J Epidemiol 2014;29(6):411-18.
104. Fabiani M, Bella A, Rota MC, et al. A/H1N1 pandemic influenza vaccination: A retrospective evaluation of adverse maternal, fetal and neonatal outcomes in a cohort of pregnant women in Italy. Vaccine 2015;33(19):2240-47.
105. Monaldi B, Bologna G, Costa GG, et al. Adherence to statin treatment following a myocardial infarction: an Italian population-based survey. Clinicoecon Outcomes Res 2015;7:273-80.
106. Maggioni AP, Orso F, Calabria S, et al. The real-world evidence of heart failure: findings from 41 413 patients of the ARNO database. Eur J Heart Fail 2016; 18(4):402-10.
107. Baldi I, Azzolina D, Berchialla P, Gregori D, Scotti L, Corrao G. Comorbidity-adjusted relative survival in newly hospitalized heart failure patients: A population-based study. Int J Cardiol 2017;243:385-88.
108. Borsari L, Malagoli C, Ballotari P, et al. Validity of hospital discharge records to identify pregestational diabetes in an Italian population. Diabetes Res Clin Pract 2017;123:106-11.
109. Terzano C, Colamesta V, Unim B, et al. Chronic obstructive pulmonary disease (COPD) exacerbation: impact of comorbidities on length and costs during hospitalization. Eur Rev Med Pharmacol Sci 2017;21(16):3680-89.
110. Trifiro G, Parrino F, Pizzimenti V, et al. The Management of Diabetes Mellitus in Patients with Chronic Kidney Disease: A Population-Based Study in Southern Italy. Clin Drug Investig 2016;36(3):203-12.
111. Antonelli-Incalzi R, Ancona C, Forastiere F, Belleudi V, Corsonello A, Perucci CA. Socioeconomic status and hospitalization in the very old: a retrospective study. BMC Public Health 2007;7:227.
112. Guaraldi G, Zona S, Menozzi M, et al. Cost of noninfectious comorbidities in patients with HIV. ClinicoEcon Outcomes Res 2013;5:481-88.
113. Pintaudi B, Lucisano G, Pellegrini F, et al. The long-term effects of stillbirth on women with and without gestational diabetes: a population-based cohort study. Diabetologia 2015;58(1):67-74.
114. Notarnicola A, Maccagnano G, Tafuri S, Moretti L, Laviola L, Moretti B. Epidemiology of diabetes mellitus in the fragility fracture population of a region of Southern Italy. J Biol Regul Homeost Agents 2016;30(1):297-302.
115. Lombardo FL, Maggini M, De Bellis A, Seghieri G, Anichini R. Lower Extremity Amputations in Persons with and without Diabetes in Italy: 2001-2010. PLoS One 2014;9(1):e86405.
116. Pagano E, De Rosa M, Rossi E, et al. The relative burden of diabetes complications on healthcare costs: The population-based CINECA-SID ARNO Diabetes Observatory. Nutr Metab Cardiovasc Dis 2016;26(10):944-50.