CHAPTER 7

A SYSTEMATIC REVIEW OF CASE-IDENTIFICATION ALGORITHMS BASED ON ITALIAN HEALTHCARE ADMINISTRATIVE DATABASES FOR THREE RELEVANT DISEASES OF THE DIGESTIVE AND GENITOURINARY SYSTEM: INFLAMMATORY BOWEL DISEASES, CELIAC DISEASE, AND CHRONIC KIDNEY DISEASE

Riccardo Di Domenicantonio,[1] Giovanna Cappai,[1] Nera Agabiti,[1] Claudia Marino,[1] Lorenzo Simonato,[2] Cristina Canova,[2] Gisella Pitter[3]
[1] Department of Epidemiology, Lazio Regional Health Service, ASL Roma 1, Rome (Italy)
[2] Department of Cardiological, Thoracic and Vascular Sciences, University of Padua, Padua (Italy)
[3] Local Health Unit "Azienda ULSS 2 Marca Trevigiana", Veneto Region, Treviso (Italy)
Corresponding author: Riccardo Di Domenicantonio; r.didomenicantonio@deplazio.it

ABSTRACT

OBJECTIVES: to identify and describe all Inflammatory Bowel Disease (IBD), Celiac Disease (CD), and Chronic Kidney Disease (CKD) case-identification algorithms by means of Italian Healthcare Administrative Databases (HADs), through a 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; exclusion criteria were the following: no details of algorithms reported, algorithm not developed in the Italian context, exclusive use of data from the death certificate register, or from general practitioner or pediatrician databases. 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.

RESULTS: the search string led to the identification of 98 articles for IBD, 42 articles for CD, and 390 for CKD. By screening the references, one paper for IBD was added. Finally, this led to 5, 9, and 8 pertinent papers respectively for IBD, CD, and CKD.

Considering the papers on IBD and CD, specific age selections were applied to focus on children and young adult populations. When a selection on age was applied for CKD, instead, it mostly considered individuals aged more than 18 years.

Three algorithms for IBD, 4 for CD, and 5 for CKD were extracted from papers and characterized. Drug prescription databases were used for both IBD and CKD algorithms, whereas the hospital discharge database and co-payment exemption database were used for IBD and CD. Pathology records and specialist visit databases were also used for CD and CKD, respectively. For each disease only one algorithm applied criteria for the exclusion of prevalent cases. External

WHAT IS ALREADY KNOWN

- **Inflammatory bowel disease (IBD), celiac disease (CD) and chronic kidney disease (CKD) are conditions causing lifelong ill health, with a large impact in terms of direct and indirect costs.**

National initiatives to ensure quality, safety, and sustainability of services for people with these conditions promote new organizational integrated-care models according to evidence-based pathways.

During the last decades, administrative databases have been increasingly used to estimate the burden of disease, assess the appropriateness of care, and plan public health policies for several conditions.

WHAT THIS PAPER ADDS

- **This review provides a comprehensive overview of the algorithms used to identify IBD, CD, and CKD in Italian administrative databases**

Despite the paucity of validated approaches, algorithms for IBD and CD can be used to perform different kinds of epidemiological studies. The same is not true for CKD, which requires improvement, mainly to detect early stage patients.

validation was performed only for Crohn's disease among IBDs, in one algorithm.
CONCLUSIONS: the results of this review indicate that case identification for IBD and CD from routinely collected data can be considered feasible and can be used to perform different kinds of epidemiological studies. The same is not true for CKD, which requires further efforts, mainly to improve the detection of early stage patients.

Keywords: algorithms, healthcare administrative databases, inflammatory bowel disease, celiac disease, chronic kidney disease

RIASSUNTO

OBIETTIVI: identificare e descrivere tutti i lavori pubblicati negli ultimi 10 anni che, utilizzando flussi amministrativi sanitari (FAS) italiani, hanno elaborato almeno un algoritmo originale per l'identificazione di soggetti affetti da malattie infiammatorie croniche dell'intestino (MICI), malattia celiaca (MC) e malattia renale cronica (MRC).
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 patologie acute e croniche. La revisione, effettuata da due revisori indipendenti, mira a identificare articoli originali pubblicati tra il 2007 e il 2017 in inglese o italiano, individuati su PubMed mediante una stringa di ricerca costituita sia da testo libero che da termini MeSH, con una parte comune a tutte le patologie e una parte specifica per patologia. I lavori pertinenti sono stati classificati a seconda dell'obiettivo per cui ciascun algoritmo è stato utilizzato e si sono estratti i dati solo dagli algoritmi con obiettivi primari (**I** occorrenza di malattia, **II** selezione di coorti/popolazioni, **III** identificazione di esito). I criteri di esclusione sono stati i seguenti: assenza di una descrizione degli algoritmi riportati; sviluppo dell'algoritmo al di fuori del contesto italiano; uso esclusivo di: certificate di morte, registri di patologia, dati dei medici di medicina generali o dei pediatri di libera scelta. Le informazioni estratte per caratterizzare e confrontare gli algoritmi originali sono: i FAS utilizzati (schede di dimissione ospedaliera, prescrizioni farmaceutiche, etc.), i codici ICD-9 e ICD-10, la selezione dei farmaci secondo il sistema di classificazione ATC, i criteri di identificazione dei casi, il periodo di osservazione/follow-back, i criteri di selezione anagrafica applicati ed eventuali validazioni esterne con le relative misure di accuratezza (sensibilità, specificità, valori predittivi) riportate.
RISULTATI: la stringa di ricerca ha portato all'identificazione di 98, 42 e 390 articoli, rispettivamente per MICI, MC e MRC, con l'aggiunta dai riferimenti bibliografici di un articolo per le MICI. Alla fine del processo di selezione, sono stati identificati 5 lavori pertinenti per le MICI, 9 per la MC e 8 per la MCR. Nell'ambito degli articoli su MICI e MC sono stati applicati criteri specifici per la selezione della popolazione pediatrica e dei giovani adulti, mentre per la MRC la maggior parte dei lavori ha considerato la popolazione adulta.
Sono stati estratti dagli articoli e caratterizzati tre algoritmi per MICI, 4 per la MC e 5 per la MRC. I dati relativi alle prescrizioni farmaceutiche sono stati utilizzati sia negli algoritmi per l'identificazione delle MICI che della MRC, mentre sono stati considerati sia le esenzioni che i ricoveri per MICI e MC. I referti di anatomia patologica e le prestazioni ambulatoriali sono state utilizzate, rispettivamente, per l'identificazione della MC e della MRC. Per ciascuna patologia, solamente in un algoritmo è stato applicato un criterio per l'esclusione dei casi prevalenti. Un solo algoritmo, sviluppato per la malattia di Crohn, nell'ambito delle MICI, è stato oggetto di validazione.
CONCLUSIONE: gli elementi emersi dalla revisione indicano che l'identificazione delle MICI e della MC attraverso l'uso dei FAS può essere ritenuta affidabile e può essere utilizzata per condurre diversi tipi di studi epidemiologici. Differentemente per la MRC, è necessario lo sviluppo di ulteriori approcci, mirati principalmente a migliorare la capacità di identificazione dei pazienti con forme precoci della patologia.

Parole chiave: algoritmi, dati amministrativi sanitari, malattie infiammatorie croniche, malattia celiaca, malattia renale cronica

INTRODUCTION

Inflammatory bowel disease (IBD), celiac disease (CD), and chronic kidney disease (CKD) are chronic conditions often occurring at a young age, that have the potential to cause lifelong ill health.
The prevention, management, and care of people with these conditions have a large impact on health services in terms of direct and indirect costs.
In the last years, the Ministry of Health dedicated efforts to ensure quality, safety, and sustainability of services for people with CKD or IBD and included these conditions in the list of the major critical chronic conditions in the "Piano Nazionale Cronicità".[1] Moreover, a specific plan was published for the management and treatment of people with CKD in the "Documento di indirizzo per la malattia renale cronica",[2] which underlines both the clinical and epidemiological relevance of this disease. Both national documents encourage clinicians, researchers, and other stakeholders to promote new organizational integrated-care models according to evidence-based pathways. Until recently, CD was considered a rare disease by the Italian healthcare system; in January 2017, the Italian government approved new healthcare standards to be guaranteed by the public healthcare system, and CD was recognized as a chronic disorder with a substantial burden and the need for reliable epidemiologic tools to monitor its occurrence and outcomes.[3]

INFLAMMATORY BOWEL DISEASES

Inflammatory bowel diseases (IBDs), including Crohn's

disease and ulcerative colitis, are chronic conditions with prevalence estimated to reach 0.3% and 0.5% respectively[4] in Europe, with a peak incidence in the second or third decade of life.[5] Flares and complications of these diseases affect the quality of life and the working productivity of patients, mainly young adults, because they may require long-term treatment, hospitalization, and, in some instances, surgery.[6] As life expectancy is not significantly modified by IBD[7,8], these diseases are burdened by higher medical expenses per patient lifetime than most other chronic conditions.[9,10]

CELIAC DISEASE

CD is an immune-mediated small intestinal enteropathy triggered by the ingestion of gluten and affecting approximately 0.5%-1% of the European population.[11] According to the Annual Report to the Parliament on CD, published by the Italian Ministry of Health,[12] the overall prevalence of this disorder in Italy is 0.33%, but the figure is likely underestimated because many cases have subtle symptoms and therefore may go undiagnosed. CD can occur at any age and frequently affects children and youth, with a female predominance. It can severely impair health and quality of life due to its frequent systemic complications and comorbidities and the need for a strict life-long gluten-free diet.[13]

CHRONIC KIDNEY DISEASE

CKD is a common complex chronic condition, which may lead to kidney failure. It increases the risk of cardiovascular complications, and, when severe, is associated with debilitating symptoms.[14] CKD encompasses a variety of disorders and represents a major public health burden. An Italian epidemiological study with data at national level reports prevalence rates of 7.5% among men and 6.5% among women for the age category 35-79 years.[15] CKD is classified in 5 stages of increasing severity, with the 5th stage corresponding to renal replacement therapy (hemodialysis, peritoneal dialysis, or kidney transplantation).[2] The prevention, management, and care for people with CKD have a large impact on health services in terms of direct and indirect costs.[16,17]

The objective of this systematic review is to describe the characteristics of algorithms that have actually been used in the past 10 years, in Italy, for IBD, CD, and CKD case identification.

METHODS

All method details are available in a specific paper,[18] which reports the study protocol with complete information on the literature search (specific search string applied to retrieve administrative healthcare data papers on PubMed, 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 were allowed, to ensure an approach that could be tailored to the characteristics of the specific diseases. For all aspects not reported in the present section, we refer the reader to the aforementioned protocol paper.

The search string used to select PubMed records consisted of a part optimized to retrieve papers focused on Italian administrative healthcare data and a specific part for the condition under study, 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 crite-

IBD: ("Ulcerative colitis"[Title/Abstract] OR "crohn"[Text word] OR "IBD"[Title/Abstract] OR "Inflammatory bowel"[Title/Abstract] OR inflammatory bowel disease[MeSH Terms])

CELIAC DISEASE: ("celiac disease"[Title/Abstract] OR "coeliac disease"[Title/Abstract] OR "celiac sprue" [Title/Abstract] OR celiac disease[MeSH Terms])

CHRONIC KIDNEY DISEASE: ((((((("Renal Insufficiency"[Mesh:NoExp] OR "Renal Insufficiency, Chronic"[Mesh:NoExp] OR "Kidney Failure, Chronic"[Mesh:NoExp]) OR (CKF[Text Word] OR CKD[Text Word] OR CRF[Text Word]) OR ("chronic kidney" [Text Word] OR "chronic renal" [Text Word] OR "kidney disease"[Text Word] OR "kidney failure"[Text Word] OR "kidney insufficiency"[Text Word])))) OR (("chronic kidney disease mineral and bone disorder"[MeSH Terms] OR "diabetic nephropathies"[MeSH Terms] OR "Hypertension, Renal"[MeSH Terms] OR "Kidney Diseases, Cystic"[MeSH Terms] OR "Nephritis"[MeSH Terms] OR "Renal Insufficiency"[MeSH Terms] OR "Renal Replacement Therapy"[MeSH Terms]))) OR (("chronic kidney disease mineral and bone disorder"[MeSH Terms] OR "diabetic nephropathies"[MeSH Terms] OR "Hypertension, Renal"[MeSH Terms] OR "Kidney Diseases, Cystic"[MeSH Terms] OR "Nephritis"[MeSH Terms] OR "Renal Insufficiency"[MeSH Terms] OR "Renal Replacement Therapy"[MeSH Terms])))

Box 1. Disease-specific search strings used to select records from PubMed.

ria 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 (primary objectives): **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.

RESULTS

The search strategy led to the identification of 98 articles for IBD, 42 articles for CD, and 390 for CKD (table 1). Out of the selected articles, 90, 29, and 356 papers, respectively, were excluded by title and abstract. This brought to 8, 13, and 34 full-text reviews, resulting in 4, 9, and 8 papers considered pertinent respectively for IBD, CD, and CKD. Most article exclusions were due to the following criteria: no disease-specific algorithms reported (4 and 17 papers excluded, respectively for IBD and CKD, none excluded for CD), absence of Italian administrative healthcare data or exclusively collected from disease registers (no articles excluded for IBD, 3 and 8 papers excluded for CD and CKD, respectively). References from the selected articles allowed the identification of one more work for IBD, leading to a total of 5 pertinent papers for IBD; for CD and CKD, no additional paper was retrieved.

PERTINENT PAPERS ON IBD, CD, AND CKD

The chronological distribution of pertinent papers showed that the majority of the articles have been published in the last three years (2014-2017) (table 2). The majority of the works, for the three conditions, focused on a region-wide setting. Only 2, 1, and 3 papers were based on a national multicenter context and none on an international multicenter setting.

With the exception of one paper for IBD, all articles used administrative data (for case identification) that dated 2008 or later. In the large majority of papers, the data used for the analysis covered more than one year.

Considering the papers on IBD and CD, specific age selections were applied to focus on children and young adult populations. When a selection on age was applied for CKD, instead, it mostly considered individuals aged more than 18 years. Articles that used at least one algorithm for objectives I, II, or III, were 3, 9, and 4, respectively for IBD, CD, and CKD. The complete list and several characteristics of the papers, using algorithms for objectives I-III, can be found in tables S1-S3 (see on-line supplementary materials), along with other concomitant uses of the algorithms, for different objectives.

Prevalence across papers that estimated the occurrence of the disease (objective **I**), ranged from 0.29% (males) to 0.25% (females) for IBD (standardized rates);[19] it was between 0.07% (males) and 0.21% (females) for CD (crude rates)[20] and 0.08% for CKD (crude rates).[21] Among these works, incidence (x100,000 persons) was 22.8 for males and 19.3 among females for IBD (standardized rates), whereas 4.5 and 27.0 (x100,000 persons), respectively, among females and males for CD (crude rates) (tables S1-S3).

IBD ALGORITHMS

Out of the 5 selected papers, 3 original algorithms focused on objective **I-III** were identified (table 3A). Specific age ranges for case definition were not present and all algorithms considered all ages. Algorithm 1 in table 3A used drug prescription ATC categories A07EA and A07EC (A= Alimentary Tract and Metabolism; 07 = Intestinal Antidiarrheals; E = Intestinal Antiinflammatory Agents; A = Corticosteroids Acting Locally, C = Aminosalicylic Acid and similar agents) for case identification, drug prescription database (DPD) was the only source of data used. The other two algorithms used data from hospital discharges selecting ICD-9-CM codes 555.XX (Regional enteritis) or 556.XX (Ulcerative enterocolitis) in the principal or secondary diagnosis. Algorithm 2 in table 3A applied the exclusion of several codes among the 556.XX group and also used the exemption from healthcare co-payment database as source for case identification. Only one algorithm reported and applied an incidence case definition, adding to prevalent IBD case criteria the following criteria: **1** exclusion of all subjects identified by the hospital discharge record database (HDD) or exemption from healthcare co-payment database (ECD) in the 7 years before the study period; **2** exclusion of all subjects with health care contacts provided through IBD co-payment exemption (outpatient visits, laboratory, endoscopy and imaging examinations, drug prescriptions) in the 7 years before the study period; **3** exclusion of all cases identified by a single hospitalization with a 555.XX or 556.XX code reported only as a secondary diagnosis, with no evidence in the discharge abstract of an IBD-related diagnosis or any procedure code strongly associated with IBD. External validation was performed only for Crohn's disease in one algorithm, by means of clinical data from 5 gastroenterology centers and reporting only a sensitivity estimate of 82%.

CD ALGORITHMS

From the 9 pertinent papers on CD, 4 original algorithms were extracted, all focusing on objectives **I-III** (table 3B). No algorithm adopted a specific age range as a definition

criterion. All 4 algorithms used data from ECD (specific code RI0060 for CD), either as the only source (#1 in table 3B) or in combination with other sources. Among the latter, HDD (ICD-9-CM code 579.0: Celiac disease) was the most exploited (Algorithms #2, #3, and #4 in table 3B). Two algorithms (#2 and #4 in table 3B) also searched the Pathology Reports Database (PRD) for SNOMED codes referring to intestinal villous atrophy, a histological hallmark of CD. Algorithm #4 exploited as an additional source a regional register of gluten-free food prescriptions and performed an evaluation of internal agreement, which revealed an only partial consistency between the four different sources. None of the algorithms was externally validated using an independent source.

CKD ALGORITHMS

As regards objectives **I-III**, for the 4 algorithms considered, 1 used only DPD (ATC categories "V03AE"- Drugs for treatment of hyperkalemia and hyperphosphatemia and "B03XA"- Other antianemic preparations), all others used only HDD (ICD-9-CM diagnosis codes 584 acute renal failure, 585 CKD, 753.1 polycystic kidney disease) or ICD-9-CM procedure codes referring to hemodialysis or peritoneal dialysis (table 3C).

DISCUSSION

This work reviewed and described the characteristics of case-identification algorithms applied in the last decade in Italy for three relevant clinical conditions. We chose to focus on the national context, since evidence from experiences developed in other Countries, generally benefiting from the availability of data sources not present in Italy (mainly information on diagnoses retrieved from outpatient claims of ambulatory care, general practitioner , specialists and pediatricians), is not likely suited to be implemented in the Italian context. Moreover, we chose to perform a comprehensive review, irrespective of the presence of algorithm accuracy measures, to provide exhaustive information on characteristics and fields of utilization of all the approaches used and published in the literature based on the Italian context.

For IBD, we observed substantial agreement on the use of hospital discharge records for case identification. Sensitivity analyses and the definition of more narrow criteria, with a combination of codes selected from the main or secondary diagnosis, could lead to an optimal trade-off between specificity and sensitivity according to the study design. Use of co-payment exemptions may detect patients that were not hospitalized during the study period, an event which is not negligible among those affected by IBD, due to the relapsing course of the disease. On the other hand, reviewed algorithms may not correctly discern between Crohn's disease and ulcerative colitis, since evidence of both diseases in the same subject, irrespective of the sources, is not rare. Use of drug prescription data was limited to only one experience which focused on prevalence estimation, based on drugs targeting the digestive tract, but also used for the treatment of several conditions besides IBD. A recently published cohort study,[22] not included in our review, reported more specific criteria to distinguish between the two diseases, defining a list of codes to characterize disease course, and applied exclusion criteria based on the use of drugs specific for other immune-related disease that may co-occur with IBD (Rheumatoid arthritis, Psoriatic arthritis, Rheumatoid arthritis, Juvenile rheumatoid arthritis, Ankylosing spondylitis).

Regarding CD, only 4 original algorithms were identified, none of which were externally validated. CD is a chronic disorder often managed in an outpatient setting, with no drug therapy available. As a consequence, drug prescription data are not useful to track the disease, while HDD may identify only the most severe or complicated cases. Thanks to the existence of a specific national exemption code, in Italy ECD allows to identify CD through administrative databases, irrespective of the healthcare setting (hospital or outpatient); therefore, it is not surprising that all four original algorithms relied on this source. Future studies will need to take into account that, with the approval of the decree on new healthcare standards, CD was moved from the category of rare diseases to that of chronic disorders and the exemption code was changed from RI00060 to 059.[3] Two algorithms published by the same research group and built on data from the region of Friuli Venezia Giulia also exploited a regional database of pathology reports.[23,24] CD is characterized by villous atrophy at small bowel biopsy, which is a diagnostic criterion and has been shown to be a highly specific marker of the disorder in a validation study conducted in Sweden.[25] Therefore, codes referring to villous atrophy in pathology reports may be a good tool to track CD. Unfortunately, the usefulness of this source is limited by the lack of availability of a regional PRD in many Italian regions. During the review process, one example of pathology reports being used for the identification of CD in the province of Varese was found,[26] however it was unclear whether the pathology archives were organized as an administrative database, therefore the paper was not included among the pertinent publications for the scope of the present review. Recent European guidelines[27,28] have introduced the possibility to avoid duodenal biopsy for the diagnosis of CD in clearly symptomatic children, which may lead PRD to underestimate CD in the pediatric population. One of the extracted algorithms[24] also explored the usefulness of a regional database containing data on gluten-free food prescriptions. In Italy, patients with a

	INFLAMMATORY BOWEL DISORDERS	COELIAC DISEASE	CHRONIC KIDNEY FAILURE
Papers identified by the string	98	42	390
Full-text readings	8	13	34
Pertinent papers	4	9	8
References added from bibliography	1	0	0
Total pertinent papers	5	9	8
Papers with objectives IV+*	2	0	4
Papers with objectives I-III*	3	9	4
Papers (with objective I-III) with at least one original algorithm	3	4	4
Papers (with objective I-III) with external validation	1	0	0
Original algorithms (with objective I-III)	**3**	**4**	**5**

* **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.

	INFLAMMATORY BOWEL DISORDERS	COELIAC DISEASE	CHRONIC KIDNEY FAILURE
Year of Publication			
2007-2010	0	0	1
2011-2013	1	1	1
2014-2017	4	8	6
Journal			
Italian	0	1	1
International	5	8	7
Setting			
Sub-regional (LHU, cities,..)	0	1	1
Regional (entire region)	3	7	4
National multicenter	2	1	3
International multicenter	0	0	0
Data time frame for the identification of the disease			
1 year	1	0	2
> 1 year	4	9	6
Use of data (even partial) following 2007 (≥2008)	4	9	6
Objective*			
I	2[21,19]	1[20]	1[21]
II	1[37]	6[24,38-42]	2[43,44]
III	0	2[23,45]	1[46]
IV	2[47,42]	0	4[48-51]
V	0	0	0
VI	0	0	0
VII	0	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 (References).

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)	Identification of cases: incident(I)-prevalent(P)	SOURCES USED IN THE ALGORITHM				CASE DEFINITION		Incidence:* criteria for the exclusion of prevalent cases (look-back time frame)	EVALUATION OF THE ALGORITHM			
				HDD ICD-9-CM code (Main diagnosis (M), Secondary diagnosis (S) Any diagnosis (A) Not reported (N))	ECD code	DPD code	Other sources (code)	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)
1	Chini, 2011[21]	I	P	–	–	A07EA, A07EC	–	DPD (1 year) >2 packages	All ages	–	Maio V et al. Using pharmacy data to identify those with chronic conditions in Emilia Romagna, Italy. J Health Serv Res Policy. 2005 Oct;10(4):232-8.	–	–	–
2	Di Domenicantonio, 2014[19]	I	I-P	555, 556 (except 556.0, 556.1, 556.4, 556.8) (A)	900.555, 900.556	–	–	HDD (10 years) OR ECD (10 years)	All ages	Biennal incidence: HDD(7yr) or ECD*(7yr) or HDD° diagn.Sec (555, 556) (-2yr) (*included outpatient visits, laboratory, endoscopy and imaging examinations, as well as those with drug prescriptions provided through IBD co-payment exception; °with no evidence in the discharge abstract of an IBD-related diagnosis or any procedure code strongly associated with IBD)	–	Only for Crohn: clinical data of five gastroenterology centres with inpatient and outpatient care settings.	Se: 82.2%	S-C
3	Meregaglia, 2015[37]	II	P	555, 556 (A)	–	–	–	HDD (1 year)	All ages	–	–	–	–	–

HDD: hospital discharge record database; **ECD:** exemption from healthcare co-payment database; **DPD:** drug prescription database; **Se:** sensitivity; **Sp:** specificity; **PPV:** positive predictive value; **NPV:** negative predictive value
* Negative values means that the criteria is assessed after the date of estimation.

Table 3A. Characteristics of Inflammatory bowel disorders case-identification algorithms published in PubMed between 2007 and 2017.

Epidemiol Prev 2019; 43 (4) Suppl 2:88-98. doi: 10.19191/EP19.4.S2.P088.095

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)	Identification of cases: incident-prevalent	SOURCES USED IN THE ALGORITHM				CASE DEFINITION		Incidence: criteria for the exclusion of prevalent cases (look-back time frame)	EVALUATION OF THE ALGORITHM			
				HDD ICD-9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))	ECD code	DPD code	Others sources (code)	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)
1	Angeli, 2012[20] Bianchi, 2016[45]	I	I-P	–	Celiac: RI0060	–		ECD	–	–	–	–	–	–
2	Canova, 2014[23] Canova, 2015[39] Canova, 2016a[40] Canova, 2016b[41] Canova, 2017[42]	III	I	579.0 (A)	Celiac: RI0060	–	PRD: D6218, M58, M58005, M58006, M58007	HDD OR ECD OR PRD	–	First event after 12 months of age	Ludvigsson, BMC Gastroenterol 2009	–	–	–
3	Fortunato, 2014[38]	II	P	579.0 (A)	Celiac: RI0060	–		HDD (11 years, 2001-2011) OR ECD (1 year, 2010)	–	–	–	–	–	–
4	Pitter, 2017[24]	II	I	579.0 (A)	Celiac: RI0060	–	PRD: D6218, M58, M58005, M58006, M58007; AFIR: 4AA2D*, 4AA2B39, 4AA2B41, 4AA2B42, 4AA2B47	HDD OR ECD OR PRD OR AFIR	–	First event	Ludvigsson, BMC Gastroenterol 2009; Canova, Am J Epidemiol 2014	–	–	S-C

HDD: hospital discharge record database; **ECD:** exemption from healthcare co-payment database; **DPD:** drug prescription database; **PRD:** pathology report database; **AFIR:** Regional Register including gluten-free food prescription (Assistenza Farmaceutica Integrativa Regionale); **Se:** sensitivity; **Sp:** specificity; **PPV:** positive predictive value; **NPV:** negative predictive value

Table 3B. Characteristics of Coeliac disease 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)	Identification of cases: incident(I)-prevalent(P)	SOURCES USED IN THE ALGORITHM				CASE DEFINITION		Incidence:* criteria for the exclusion of prevalent cases (look-back time frame)	EVALUATION OF THE ALGORITHM		
				HDD ICD-9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))	ECD code	DPD code	Others sources (code)	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*)
1	Chini, 2011[21]	I	P	–	–	V03AE; B03XA	–	DPD (1 year) >2 package	All ages	–	Maio V et al. Using pharmacy data to identify those with chronic conditions in Emilia Romagna, Italy. J Health Serv Res Policy. 2005 Oct;10(4):232-8.	–	–
2	Degli Esposti, 2016[46]	III	P	584, 585	–	–	–	HDD (at least 15 months of follow up)	18+	–	–	–	–
3	Roggeri, 2017[43]	II	I	38.95, 54.98 (procedures)	–	–	ACD: 39.95.5, 39.95.6; 39.95.7; 39.95.8, 39.95.9, 39.95.6, 39.95.1; 39.95.2; 39.95.3, 39.95.4, 39.95.2, 39.95A, 54.98.1, 54.98.2	HDD (-1 year) or ACD (-1 year)	All ages	HDD (2yr)>0 or ACD (2yr)>0	–	–	–
4	Degli Esposti, 2017[44]	II	P	753.1 (A)	–	–	–	HDD (3 years)	18+	–	–	–	–

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

* Negative values means that the criteria is assessed after the date of estimation.

Table 3C. Characteristics of Chronic kidney failure case-identification algorithms published in PubMed between 2007 and 2017.

certified diagnosis of CD can obtain vouchers from the public healthcare service to buy gluten-free food with a predefined monthly cap. All gluten-free dietary products have to be included in a registry of the Italian Ministry of Health and are identified by specific ATC/GMP codes. Therefore, the database of gluten-free food prescriptions may be useful to identify patients with CD and possibly to monitor adherence to the dietary regimen and related outcomes. However, in the above-mentioned paper the regional database of gluten-free food prescriptions was only partially overlapping with the ECD source, therefore its reliability should be further investigated.

All the above-mentioned considerations highlight that more research is needed to develop a suitable and accurate strategy to identify CD cases from administrative databases. Many different possible sources are available, each with advantages and limitations that should ideally be quantified in a validation study against a clinical independent source. The existence of CD registries held by referral clinical centers may serve this purpose.[29-31]

Regarding CKD, this review highlights the paucity of works conducted in Italy to identify people with CKD on the basis of administrative data. The 4 papers, pertinent to objectives I-III, used different sources of data and different algorithms. None of these 4 papers were able to identify the whole CKD population: two of them identified only the most severe cases of CKD on the basis of specific drugs and dialysis procedures, one studied a single form of CKD, i.e., polycystic kidney disease; only one paper analyzed both CKD and acute renal failure, as outcomes. CKD has multiple etiopathogenesis and clinical forms, with increasing levels of severity. While more severe disease can be captured by selecting specific drugs and procedures – like in two studies included in this review – the mild and moderate stages of disease need to be assessed through a more complex integrated multisource algorithm, but such experiences was not detected by our review. Only one paper, from the Lazio region, estimated the prevalence of CKD, but for the highest severity levels, based on the list of drug agents used; the value of prevalence is in line with the data from the Lazio Regional Dialysis and Transplant database.[32,33] Strategies need to be developed in Italy to identify the complete burden of CKD using administrative data, in order to monitor temporal and geographic variation in the epidemiology of the disease, evaluate quality of care for CKD patients, and support the implementation of a new organizational integrated-care model.[1,2] In some Italian regions, registries on dialysis do exist, and there is a nation-wide coordination effort to describe and monitor this stage of the disease.[34,35] Moreover, in 2017, the Italian Ministry of Health promoted the institution of a national CKD registry – including all stages – to describe the epidemiology, monitor the quality of care and health outcomes, and prevent the incidence of the highest severity level of this disease.[36] This substantiates the interest of the Italian Ministry of Health for CKD population and care; however, no data is yet available from this initiative.

CONCLUSION

The results of this review indicate that for IBD and CD, case identification from routinely collected data can be considered feasible and can be used to perform different kinds of epidemiological studies, however more research is needed to identify the most accurate algorithms, ideally through validation studies. The same is not true for CKD, which requires further efforts, mainly to improve the detection of early stage patients.

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.

REFERENCES

1. Ministero della Salute. Piano Nazionale della Cronicità. Anno 2016. Available from: http://www.salute.gov.it/imgs/C_17_pubblicazioni_2584_allegato.pdf
2. Berloco P, Brizzi F, Canu C, et al. Documento di indirizzo per la malattia renale cronica. Roma, Ministero della Salute, 2014. Available from: http://www.salute.gov.it/imgs/C_17_pubblicazioni_2244_allegato.pdf (last accessed: December 20, 2018).
3. Decreto del Presidente del Consiglio di Ministri; 12.01.2017. Definizione e aggiornamento dei livelli essenziali di assistenza. GU Serie Generale n. 65 del 18.03.2017. Available from: http://www.trovanorme.salute.gov.it/norme/dettaglioAtto?id=58669&completo=true (last accessed: December 20, 2018).
4. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2017;390(10114):2769-78.
5. Loftus EV, Schoenfeld P, Sandborn WJ. The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. Aliment Pharmacol Ther 2002;16(1):51-60.
6. Høivik ML, Moum B, Solberg IC, et al. Work disability in inflammatory bowel disease patients 10 years after disease onset: results from the IBSEN Study. Gut 2013;62(3):368-75.
7. Sonnenberg A. Time trends of mortality from Crohn's disease and ulcerative colitis. Int J Epidemiol 2007;36(4):890-99.
8. Jess T, Winther KV, Munkholm P, Langholz E, Binder V. Mortality and causes of death in Crohn's disease: follow-up of a population-based cohort in Copenhagen County, Denmark. Gastroenterology 2002;122(7):1808-14.
9. Bodger K. Cost of illness of Crohn's disease. Pharmacoeconomics 2002;20(10):639-52.
10. Hay JW, Hay AR. Inflammatory bowel disease: costs-of-illness. J Clin Gastroenterol 1992;14(4):309-17.
11. Mustalahti K, Catassi C, Reunanen A, et al. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. Ann Med 2010;42(8):587-95.

12. Ministero della Salute. Relazione annuale al Parlamento sulla celiachia - Anno 2016. Available from: http://www.salute.gov.it/portale/documentazione/p6_2_2_1.jsp?lingua=italiano&id= 2689 (last accessed: January 18, 2019).
13. Lebwohl B, Sanders DS, Green PHR. Coeliac disease. Lancet Lond Engl 2018; 391(10115):70-81.
14. Ketteler M, Block GA, Evenepoel P, et al. Executive summary of the 2017 KDIGO Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD) Guideline Update: what's changed and why it matters. Kidney Int 2017;92(1):26-36.
15. Viscogliosi G, De Nicola L, Vanuzzo D, et al. Mild to moderate chronic kidney disease and functional disability in community-dwelling older adults. The Cardiovascular risk profile in Renal patients of the Italian Health Examination Survey (CARHES) study. Arch Gerontol Geriatr 2019;80:46-52.
16. Dodd R, Palagyi A, Guild L, Jha V, Jan S. The impact of out-of-pocket costs on treatment commencement and adherence in chronic kidney disease: a systematic review. Health Policy Plan 2018;33(9):1047-54.
17. Vanholder R, Annemans L, Brown E, et al. Reducing the costs of chronic kidney disease while delivering quality health care: a call to action. Nat Rev Nephrol 2017;13(7):393-409.
18. 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.
19. Di Domenicantonio R, Trotta F, Cascini S, et al. Population-based cohort study on comparative effectiveness and safety of biologics in inflammatory bowel disease. Clin Epidemiol 2018;10:203-13.
20. Angeli G, Pasquini R, Panella V, Pelli MA. An epidemiologic survey of celiac disease in the Terni area (Umbria, Italy) in 2002-2010. J Prev Med Hyg 2012;53(1):20-23.
21. 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.
22. Di Domenicantonio R, Trotta F, Cascini S, et al. Population-based cohort study on comparative effectiveness and safety of biologics in inflammatory bowel disease. Clin Epidemiol 2018;10:203-13.
23. Canova C, Zabeo V, Pitter G, et al. Association of maternal education, early infections, and antibiotic use with celiac disease: a population-based birth cohort study in northeastern Italy. Am J Epidemiol 2014;180(1):76-85.
24. Pitter G, Gnavi R, Romor P, Zanotti R, Simonato L, Canova C. Assessment of an algorithm to identify paediatric-onset celiac disease cases through administrative healthcare databases. Epidemiol Prev 2017;41(2):102-08.
25. Ludvigsson JF, Brandt L, Montgomery SM, Granath F, Ekbom A. Validation study of villous atrophy and small intestinal inflammation in Swedish biopsy registers. BMC Gastroenterol 2009;9:19.
26. Elli L, Contiero P, Tagliabue G, Tomba C, Bardella MT. Risk of intestinal lymphoma in undiagnosed coeliac disease: results from a registered population with different coeliac disease prevalence. Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver 2012;44(9):743-47.
27. Husby S, Koletzko S, Korponay-Szabó IR, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. J Pediatr Gastroenterol Nutr 2012;54(1):136-60.
28. Murch S, Jenkins H, Auth M, et al. Joint BSPGHAN and Coeliac UK guidelines for the diagnosis and management of coeliac disease in children. Arch Dis Child 2013;98(10):806-11.
29. Silano M, Volta U, Mecchia AM, et al. Delayed diagnosis of coeliac disease increases cancer risk. BMC Gastroenterol 2007;7:8.
30. Volta U, Vincentini O, Quintarelli F, Felli C, Silano M, Collaborating Centres of the Italian Registry of the Complications of Celiac Disease. Low risk of colon cancer in patients with celiac disease. Scand J Gastroenterol 2014;49(5):564-68.
31. Zingone F, West J, Auricchio R, et al. Incidence and distribution of coeliac disease in Campania (Italy): 2011-2013. United Eur Gastroenterol J 2015;3(2):182-89.
32. Registro Regionale Dialisi e Trapianto Lazio - RRDTL. Available from: http://www.deplazio.net/it/registro-dialisi-e-trapianto (last accessed: December 20, 2018).
33. Bossola M, Marino C, Di Napoli A, et al. Functional impairment and risk of mortality in patients on chronic hemodialysis: results of the Lazio Dialysis Registry. J Nephrol 2018;31(4):593-602.
34. Registro Italiano dialisi e trapianto (RIDT). Available from: https://ridt.sinitaly.org/ (last accessed: December 20, 2018).
35. Società Italiana di Nefrologia Pediatrica. Available from: https://www.sinepe.it/ (last accessed: May 22, 2019).
36. Decreto del Presidente del Consiglio dei Ministri, 03.03.2017. Identificazione dei sistemi di sorveglianza e dei registri di mortalità, di tumori e di altre patologie. GU Serie Generale n. 109, 12.05.2017. Available from: http://www.gazzettaufficiale.it/eli/id/2017/05/12/17A03142/sg (last accessed: December 20, 2018).
37. Meregaglia M, Banks H, Fattore G. Hospital Burden and Gastrointestinal Surgery in Inflammatory Bowel Disease Patients in Italy: A Retrospective Observational Study. J Crohns Colitis 2015;9(10):853-62.
38. Fortunato F, Martinelli D, Prato R, Pedalino B. Results from ad hoc and routinely collected data among celiac women with infertility or pregnancy related disorders: Italy, 2001-2011. ScientificWorldJournal 2014;2014:614269.
39. Canova C, Pitter G, Ludvigsson JF, et al. Coeliac disease and asthma association in children: the role of antibiotic consumption. Eur Respir J 2015;46(1):115-22.
40. Canova C, Pitter G, Ludvigsson JF, et al. Risks of hospitalization and drug consumption in children and young adults with diagnosed celiac disease and the role of maternal education: a population-based matched birth cohort study. BMC Gastroenterol 2016;16:1.
41. 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.e1.
42. Canova C, Pitter G, Zanier L, Zanotti R, Simonato L, Ludvigsson JF. Inflammatory Bowel Diseases in Children and Young Adults with Celiac Disease. A Multigroup Matched Comparison. Inflamm Bowel Dis 2017;23(11):1996-2000.
43. Roggeri A, Roggeri DP, Zocchetti C, et al. Healthcare costs of the progression of chronic kidney disease and different dialysis techniques estimated through administrative database analysis. J Nephrol 2017;30(2):263-69.
44. Degli Esposti L, Veronesi C, Perrone V, Buda S, Santoro A. Healthcare resource consumption and cost of care among patients with polycystic kidney disease in Italy. Clin Outcomes Res 2017;9:233-39.
45. 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.
46. Degli Esposti L, Desideri G, Saragoni S, Buda S, Pontremoli R, Borghi C. Hyperuricemia is associated with increased hospitalization risk and healthcare costs: Evidence from an administrative database in Italy. Nutr Metab Cardiovasc Dis 2016;26(10):951-61.
47. Degli Esposti L, Sangiorgi D, Perrone V, et al. Adherence and resource use among patients treated with biologic drugs: findings from BEETLE study. Clin Outcomes Res 2014;6:401-07.
48. 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.
49. Conti V, Biagi C, Melis M, et al. Acute renal failure in patients treated with dronedarone or amiodarone: a large population-based cohort study in Italy. Eur J Clin Pharmacol 2015;71(9):1147-53.
50. 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.
51. Fedeli U, De Giorgi A, Gennaro N, et al. Lung and kidney: a dangerous liaison? A population-based cohort study in COPD patients in Italy. Int J Chron Obstruct Pulmon Dis 2017;12:443-50.