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 Pitter3
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 accuracy measures, sensitivity analyses and the contribution of each HAD, have also been recorded.

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.
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”. Moreover, a specific plan was published for the management and treatment of people with CKD in the “Documento di indiriz-
disease and ulcerative colitis, are chronic conditions with prevalence estimated to reach 0.3% and 0.5% respectively in Europe, with a peak incidence in the second or third decade of life. 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. As life expectancy is not significantly modified by IBD, these diseases are burdened by higher medical expenses per patient lifetime than most other chronic conditions.

**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. According to the Annual Report to the Parliament on CD, published by the Italian Ministry of Health, 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 lifelong gluten-free diet.

**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. 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. The 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 lifelong gluten-free diet.

**METHODS**

All method details are available in a specific paper, 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-
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; 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 supplemen-
tary 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 Antinflammatory 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.
and CHroniC Kidney diSeaSe
inflamma Tory boWel diSeaSeS, CeliaC diSeaSe,  
On the other hand, reviewed algorithms may not correct-
fected by IBD, due to the relapsing course of the disease. 
patients that were not hospitalized during the study pe-
study design. Use of co-payment exemptions may detect 
off between specificity and sensitivity according to the 
ria, with a combination of codes selected from the main 
sitivity analyses and the definition of more narrow crite-
ise developed in other Countries, generally benefiting 
hallmark of CD. Algorithm #4 exploited as an additional 
specialists and pediatricians), is not likely suited to be im-
those 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 expe-
riences developed in other Countries, generally benefiting 
from the availability of data sources not present in Italy 
(mainly information on diagnoses retrieved from outpa-
tient claims of ambulatory care, general practitioner , spe-
cialists and pediatricians), is not likely suited to be im-
plemented in the Italian context. Moreover, we chose to 
perform a comprehensive review, irrespective of the pres-
ence of algorithm accuracy measures, to provide exhaust-
ive 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. Sen-
sitivity analyses and the definition of more narrow crite-
ria, 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 pe-
riod, an event which is not negligible among those af-
fected by IBD, due to the relapsing course of the disease. 
On the other hand, reviewed algorithms may not correct-
ly discern between Crohn’s disease and ulcerative colitis, 
since evidence of both diseases in the same subject, irre-
respect of the sources, is not rare. Use of drug prescrip-
tion data was limited to only one experience which fo-
cused on prevalence estimation, based on drugs targeting 
the digestive tract, but also used for the treatment of sev-
eral conditions besides IBD. A recently published cohort 
study,22 not included in our review, reported more specif-
ic criteria to distinguish between the two diseases, defin-
ing a list of codes to characterize disease course, and ap-
plied exclusion criteria based on the use of drugs specific 
for other immune-related disease that may co-occur with 
IBD (Rheumatoid arthritis, Psoriatic arthritis, Rheuma-
toid 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 prescrip-
tion 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 ad-
ministrative databases, irrespective of the healthcare set-
ting (hospital or outpatient); therefore, it is not surpris-
ning 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 R100060 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 at-
rophy at small bowel biopsy, which is a diagnostic crite-
rion and has been shown to be a highly specific marker 
of the disorder in a validation study conducted in Swe-
den.25 Therefore, codes referring to villous atrophy in pa-
thyology reports may be a good tool to track CD. Unfor-
nately, the usefulness of this source is limited by the 
lack of availability of a regional PRD in many Italian re-
gions. During the review process, one example of pathol-
ylogy 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 ad-
ministrative database, therefore the paper was not includ-
ed among the pertinent publications for the scope of the 
present review. Recent European guidelines27,28 have in-
troduced 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 pop-
ulation. One of the extracted algorithms24 also explored 
the usefulness of a regional database containing data on 
gluten-free food prescriptions. In Italy, patients with a
<table>
<thead>
<tr>
<th></th>
<th>INFILMATORY BOWEL DISORDERS</th>
<th>COELIAC DISEASE</th>
<th>CHRONIC KIDNEY FAILURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papers identified by the string</td>
<td>98</td>
<td>42</td>
<td>390</td>
</tr>
<tr>
<td>Full-text readings</td>
<td>8</td>
<td>13</td>
<td>34</td>
</tr>
<tr>
<td>Pertinent papers</td>
<td>4</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>References added from bibliography</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total pertinent papers</strong></td>
<td><strong>5</strong></td>
<td><strong>9</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td>Papers with objectives IV+*</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Papers with objectives I-III*</td>
<td>3</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Papers (with objective I-III) with at least one original algorithm</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Papers (with objective I-III) with external validation</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Original algorithms (with objective I-III)</strong></td>
<td><strong>3</strong></td>
<td><strong>4</strong></td>
<td><strong>5</strong></td>
</tr>
</tbody>
</table>

* 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.

<table>
<thead>
<tr>
<th></th>
<th>INFILMATORY BOWEL DISORDERS</th>
<th>COELIAC DISEASE</th>
<th>CHRONIC KIDNEY FAILURE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year of Publication</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2010</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2011-2013</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2014-2017</td>
<td>4</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td><strong>Journal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>International</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-regional (LHU, cities,..)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Regional (entire region)</td>
<td>3</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>National multicenter</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>International multicenter</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Data time frame for the identification of the disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>&gt; 1 year</td>
<td>4</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Use of data (even partial) following 2007 (&gt;2008)</td>
<td>4</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td><strong>Objective</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>221.19</td>
<td>120</td>
<td>121</td>
</tr>
<tr>
<td>II</td>
<td>137</td>
<td>624-38-42</td>
<td>243,44</td>
</tr>
<tr>
<td>III</td>
<td>0</td>
<td>223,45</td>
<td>146</td>
</tr>
<tr>
<td>IV</td>
<td>247,42</td>
<td>0</td>
<td>448-51</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VI</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VII</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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).
<table>
<thead>
<tr>
<th>Algorithm ID</th>
<th>Author(s) Year of Publication</th>
<th>Objective</th>
<th>Population cohort selection</th>
<th>Outcome identification</th>
<th>Identification of cases: Incident (i) - Prevalent (p)</th>
<th>Sources used in the algorithm</th>
<th>Case definition</th>
<th>Evaluation of the algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chini, 2011</td>
<td>I</td>
<td>P</td>
<td>–</td>
<td>–</td>
<td>HDD ECD-9-CM code (Main diagnosis (M), Secondary diagnosis (S)) Any diagnosis (A) Not reported (N)</td>
<td>ECD code DPD code Other sources (code)</td>
<td>Algorithm*</td>
</tr>
<tr>
<td>2</td>
<td>Di Domenicantonio, 2014</td>
<td>I</td>
<td>P</td>
<td>555, 556 (except 556.0, 556.1, 556.4, 556.8) (A)</td>
<td>900.555, 900.556</td>
<td>HDD (10 years) OR ECD (10 years)</td>
<td>–</td>
<td>All ages</td>
</tr>
<tr>
<td>3</td>
<td>Meregaglia, 2015</td>
<td>II</td>
<td>P</td>
<td>555, 556 (A)</td>
<td>–</td>
<td>HDD (1 year)</td>
<td>All ages</td>
<td>–</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Algorithm ID</th>
<th>Author year of publication of the original articles with the same algorithm</th>
<th>Objective (I disease occurrence, II population-cohort selection, III outcome identification)</th>
<th>Identification of cases: incident-prevalent (following articles with the same algorithm)</th>
<th>SOURCES USED IN THE ALGORITHM</th>
<th>CASE DEFINITION</th>
<th>EVALUATION OF THE ALGORITHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Angeli, 2012; Bianchi, 2016</td>
<td>I-P</td>
<td>Celiac R0060</td>
<td>HDD: ICD-9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))</td>
<td>ECD code</td>
<td>Algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Canova, 2014; Canova, 2015; Canova, 2016; Canova, 2016a; Canova, 2016b; Canova, 2017</td>
<td>III I 579.0 (A)</td>
<td>Celiac R0060</td>
<td>HDD: ICD-9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))</td>
<td>ECD code</td>
<td>Algorithm</td>
</tr>
<tr>
<td>3</td>
<td>Fortunato, 2014</td>
<td>II P 579.0 (A)</td>
<td>Celiac R0060</td>
<td>HDD (11 years, 2001-2011) OR ECD (1 year, 2010)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Pitter, 2017</td>
<td>II I 579.0 (A)</td>
<td>Celiac R0060</td>
<td>HDD: ICD-9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))</td>
<td>ECD code</td>
<td>Algorithm</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Algorithm ID</th>
<th>Author, year of publication</th>
<th>Objective (I disease identification, II disease cohort selection, III outcome identification)</th>
<th>SOURCES USED IN THE ALGORITHM</th>
<th>CASE DEFINITION</th>
<th>Incidence:* criteria for the exclusion of prevalent cases (look-back time frame)</th>
<th>EVALUATION OF THE ALGORITHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chini, 2011</td>
<td>I</td>
<td>HDD: IC9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))</td>
<td>DPD (1 year) ≥2 package</td>
<td>All ages</td>
<td>Algorithm derived from a previous published one (reference)</td>
</tr>
<tr>
<td>2</td>
<td>Degli Esposti, 2016</td>
<td>III</td>
<td>HDD: IC9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))</td>
<td>HDD (at least 15 months of follow up) 18+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>Roggeri, 2017</td>
<td>II</td>
<td>HDD: IC9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))</td>
<td>HDD (-1 year) or ACD (-1 year)</td>
<td>All ages</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>Degli Esposti, 2017</td>
<td>II</td>
<td>HDD: IC9-CM code (Main diagnosis (M) Any diagnosis (A) Not reported (N))</td>
<td>HDD (3 years) 18+</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

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.

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