

# Testing of Coding Algorithms for Inflammatory Bowel Disease Identification, as Indication for Use of Biological Drugs, Using a Claims Database from Southern Italy

Ylenia Ingrasciotta<sup>1,2</sup>, Valentina Isgrò<sup>1</sup>, Saveria Serena Foti<sup>2</sup>, Valentina Ientile<sup>3</sup>, Andrea Fontana<sup>4</sup>, Luca L'Abbate<sup>3</sup>, Roberto Benoni<sup>1</sup>, Elena Sofia Fiore<sup>1</sup>, Michele Tari<sup>5</sup>, Angela Alibrandi<sup>6</sup>, Gianluca Trifirò<sup>1,2</sup>

<sup>1</sup>Department of Diagnostics and Public Health, University of Verona, Verona, Italy; <sup>2</sup>Academic Spin-off "INSPIRE – Innovative Solutions for Medical Prediction and Big Data Integration in Real World Setting" – Azienda Ospedaliera Universitaria "G. Martino", Messina, Italy; <sup>3</sup>Department of Biomedical and Dental Sciences and Morphofunctional Imaging, University of Messina, Messina, Italy; <sup>4</sup>Unit of Biostatistics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni, Rotondo, Italy; <sup>5</sup>Caserta-I Local Health Service, Caserta, Italy; <sup>6</sup>Department of Economics, University of Messina, Messina, Italy

Correspondence: Ylenia Ingrasciotta, Department of Diagnostics and Public Health, University of Verona, P. le L.A. Scuro 10, Verona, 37134, Italy, Tel +39 045 802 7679, Email ylenia.ingrasciotta@univr.it

**Background:** Inflammatory bowel diseases (IBDs), Crohn's disease (CD) and ulcerative colitis (UC), are chronic diseases that have been increasingly treated with biological drugs in recent years. Newly developed coding algorithms for IBD identification using claims databases are needed to improve post-marketing surveillance of biological drugs.

**Objective:** To test algorithms to identify CD and UC, as indication for use of biological drugs approved for IBD treatment, using a claims database.

**Methods:** Data were extracted from the Caserta Local Health Unit database between 2015 and 2018. CD/UC diagnoses reported by specialists in electronic therapeutic plans (ETPs) were considered as gold standard. Five algorithms were developed based on ICD-9-CM codes as primary cause of hospital admissions, exemption from healthcare service co-payment codes and drugs dispensing with only indication for CD/UC. The accuracy was assessed by sensitivity (Se), specificity (Sp), positive (PPV) and negative predicted values (NPV) along with computation of the Youden Index and F-score.

**Results:** In the study period, 1205 subjects received at least one biological drug dispensing approved for IBD and 134 (11.1%) received  $\geq 1$  ETP with IBD as use indication. Patients with CD and CU were 83 (61.9%) and 51 (38.1%), respectively. Sensitivity of the different algorithms ranged from 71.1% (95% CI: 60.1–80.5) to 98.8 (95% CI: 93.5–100.0) for CD and from 64.7% (95% CI: 50.1–77.6) to 94.1 (95% CI: 83.8–98.8) for UC, while specificity was always higher than 91%. The best CD algorithm was "Algorithm 3", based on hospital CD diagnosis code OR CD exemption code OR [IBD exemption code AND dispensing of non-biological drugs with only CD indication] (Se: 98.8%; Sp: 97.2%; PPV: 84.5%, NPV: 99.8%), achieving the highest diagnostic accuracy (Youden Index=0.960). The best UC algorithm was "Algorithm 3", based on specific hospital UC diagnosis code OR UC exemption code OR [IBD exemption code AND golimumab dispensing] OR dispensing of non-biological drugs with only UC indication (Se: 94.1%; Sp: 91.6%; PPV: 50.0%; NPV: 99.4%), and achieving the highest diagnostic accuracy (Youden Index=0.857).

**Conclusion:** In a population-based claims database, newly coding algorithms including diagnostic and exemption codes plus specific drug dispensing yielded highly accurate identification of CD and UC as distinct indication for biological drug use.

**Keywords:** biological drugs, algorithm, healthcare database, ulcerative colitis, Crohn's disease

## Introduction

Inflammatory bowel diseases (IBDs), including Crohn's disease (CD) and ulcerative colitis (UC), are chronic diseases of the gastrointestinal tract associated with increased morbidity and reduced quality of life.<sup>1</sup> In some cases, CD and UC may not be easily differentiated from a clinical and histopathological point of view (ie IBD unspecified).<sup>2</sup>

The incidence and prevalence of CD and UC are increasing worldwide:<sup>3</sup> from a few cases observed at the beginning of the last century to the current estimates of 2.4 million patients in Europe.<sup>4–7</sup> In Europe, a prevalence of 0.3% and 0.5% for CD and UC, respectively, has been estimated.<sup>8</sup> In Italy, epidemiological data about IBD are limited, mostly because a national register has never been implemented, and most of the available studies are not updated.<sup>9</sup>

According to recent guidelines, several drugs are available for the management of IBDs, such as conventional drugs (ie aminosalicylates), corticosteroids acting locally (ie budesonide, beclomethasone), systemic steroids, immunosuppressants, biological drugs (ie tumour necrosis factor (TNF)- $\alpha$  inhibitors, anti-integrins, anti-interleukins 12/23) and a more recent small-molecule drug, tofacitinib (janus kinase inhibitor).<sup>10,11</sup>

Considering the expensive costs and the safety issues associated with biological drug treatments, healthcare databases may be useful tools for monitoring biological drug use in the real-world setting.<sup>12–15</sup> Biological drugs approved for the treatment of IBDs are also approved for the treatment of other immune-mediated inflammatory diseases (IMIDs), such as rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, psoriasis, juvenile idiopathic arthritis, uveitis and hidradenitis suppurativa.

Despite claims databases having been extensively used for research purposes, the lack of information on the exact indications of use related to dispensed drugs is a relevant criticism. As such, it is essential to identify the exact indication for use of biological drugs through validated coding algorithms as patients eligible for those drugs are often affected by concomitant immune-mediated inflammatory diseases for which biological drugs are approved.<sup>16</sup>

Several validation studies aimed at detecting a reliable diagnosis of IBD within different claims databases have been conducted.<sup>17–20</sup> A validation study conducted in Ontario (Canada), using the Ottawa Hospital Data Warehouse (OHDW), identified adult patients with IBD by linking information extracted from charts to healthcare administrative data and compared the accuracy of several algorithms. The authors reported that the best accurate algorithm to identify patients from 18 to 64 years at IBD diagnosis date considered five physician contacts or hospitalizations related to IBD within 4 years (sensitivity (Se), 76.8%; specificity (Sp), 96.2%; positive predictive value (PPV), 81.4%; negative predictive value (NPV), 95.0%); while, in patients aged  $\geq 65$  years at IBD diagnosis date, adding an IBD-related medication as a pharmacy claim, accuracy was improved.<sup>17</sup> Specifically, an Italian study using the Lazio region claims database tested an algorithm to identify IBD patients, showing that CD/UC patients could be identified using a specific diagnosis code as a primary/secondary cause of hospital admissions OR specific exemption from healthcare service co-payment code;<sup>20</sup> however, both studies<sup>17,20</sup> did not focus on the identification of CD/UC as indication for use of biological drugs specifically.

Moreover, in general, some studies tested algorithms identifying IBD patients and do not distinguish patients with Crohn's disease from those with ulcerative colitis. Moreover, due to the differences across various country-specific healthcare systems, coding algorithms for disease identification should be adapted to the specific national and/or loco-regional settings.<sup>21,22</sup>

The aim of the present study was to test coding algorithms for CD/UC identification as indication for use of biological drugs approved for the treatment of IBDs, using claims database from Caserta Local Health Unit (LHU) in the period 2015–2018.

## Methods

### Data Source

Fully anonymized data were extracted from the claims database of Caserta Local Health Unit (LHU), covering a total population of more than 1 million inhabitants, during the years 2015–2018. Collected data included: 1) demographic database, including information about the date of birth, gender, date of registration in the regional healthcare system, and where applicable, date and cause of death, coded with International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) codes; 2) out-patient pharmacy database, including data on the date of drug dispensing,

number of dispensed packages, active substance and brand name; 3) hospital discharge database, including information on the date of hospital admission and discharge, diagnosis-related group (DRG), main diagnosis and up to five secondary diagnoses, main procedure and up to five secondary procedures, coded with ICD-9-CM; and 4) exemptions from healthcare service co-payment database, including coded information about chronic diseases or socioeconomic factors. In Italy, biological drugs are fully reimbursed by the National Health Service and for each biological drug prescription, specialists have to fill a therapeutic plan, which indicates the exact drug name, number of dispensed packages, dosing regimen and exact indication for use. Electronic therapeutic plans (ETPs) were available in the Caserta LHU. All databases can be linked through an anonymous subject identifier. Caserta record linkage database, as well as specifically ETPs, have been shown to provide accurate and reliable information for pharmacoepidemiology research, as documented elsewhere.<sup>23–27</sup> The dispensed drugs were coded using the Anatomical Therapeutic Clinical (ATC) classification system and the Italian Marketing Authorization Code (AIC), while comorbidities and indications for use were coded through the ninth revision of the International Classification of Diseases – Clinical Modification (ICD-9-CM).

## Study Drugs

The exposure of interest was the biological drug treatment. All biological drugs (both originators and biosimilars) available and approved in Italy for the treatment of Crohn's disease or ulcerative colitis during the study period were included: infliximab, adalimumab, golimumab, vedolizumab and ustekinumab. The Italian Medicines Agency approved ustekinumab for the treatment of Crohn's disease in September 2018 and for the treatment of UC in October 2019. [Supplement Table 1](#) shows all the drugs included in this study.

## Gold-Standard Cohort Definition

From the source population of Caserta LHU, all subjects with at least one biological drug dispensing approved for the treatment of IBDs during the years 2015–2018 were identified. Of these, biological drug users with at least one ETP related to biological drug prescriptions were retrieved. Gold standard cases were defined as subjects with at least one CD or UC diagnosis registered in the ETP during the study period. Patients receiving ETPs reporting unspecified IBD diagnosis (not specifically CD or UC) as indication for use were excluded from the cohort.

## Identifying Algorithms for IBD Patient Identification

Five specific coding algorithms for the distinct identification of CD and UC, as indication for use, were identified based on clinical expertise and based on algorithms previously validated in other countries ([Table 1](#); [Supplement Table 2](#)).<sup>20,22,28</sup> IBD-related ICD-9-CM codes (CD: 555.xx; UC: 556.xx) as primary causes of hospital admissions,

**Table 1** Coding Algorithms Developed to Identify Patients with CD or UC from the Claims Database of Caserta LHU During the Years 2015–2018

Algorithms	Crohn's Disease	Comment of the Local Expert
1	ICD-9-CM: 555x (primary) <b>OR</b> Exemptions from healthcare service co-payment: 009.555 OR 555	CD algorithm was adapted from the study conducted by Di Domenicantonio et al, <sup>20</sup> indeed, in that study, ICD9-CM diagnosis code of 555.xx as the primary or secondary diagnosis or the activation of co-payment exemption for CD (code 009.555) was used as a case definition criteria. In our algorithm, we decided to consider only ICD9-CM diagnosis code of 555.xx as primary cause of hospitalization to be sure that the hospital admission was related to CD
2	ICD-9-CM: 555x (primary) <b>OR</b> Exemptions from healthcare service co-payment: 009.555 OR 555 <b>OR</b> ((Exemptions from healthcare service co-payment: 009) AND (Methotrexate: 039153x, 044252x, 044253x, 043416x, 044257x, 044149x, 044224x))	CD algorithm was performed from "Algorithm 1", adding methotrexate dispensing (considering only NDC codes approved for the treatment of CD and other IMIDs in Italy). <sup>41</sup> The combination of both methotrexate dispensing AND the exemption from healthcare service co-payment code, identifying <i>unspecified IBD (009)</i> , could ensure the identification of CD (rather than other IMIDs)

(Continued)

Table I (Continued).

Algorithms	Crohn's Disease	Comment of the Local Expert
3	ICD-9-CM: 555x (primary) <b>OR</b> Exemptions from healthcare service co-payment: 009.555 OR 555 <b>OR</b> ((Exemptions from healthcare service co-payment: 009) AND (Methotrexate: 039153x, 044252x, 044253x, 043416x, 044257x, 044149x, 044224x)) <b>OR</b> Budesonide: 034734x, 036507022, 036507034, 036507059, 036507046, 036507010, 044798x	CD algorithm was performed from "Algorithm 2", adding budesonide dispensing (considering only NDC codes with the only approval for the treatment of CD in Italy). <sup>42</sup> We decided to consider CD-related ICD-9-CM codes OR exemptions from co-payment codes OR non-biological drugs with indication for CD in order to explore all the conditions to identify CD patients from claims databases
4	Exemptions from healthcare service co-payment: 009.555 OR 555 <b>OR</b> ((Exemptions from healthcare service co-payment: 009) AND (Methotrexate: 039153x, 044252x, 044253x, 043416x, 044257x, 044149x, 044224x)) <b>OR</b> (Budesonide: 034734x, 036507022, 036507034, 036507059, 036507046, 036507010, 044798x)	CD algorithm was performed from "Algorithm 3", excluding CD-related ICD-9-CM codes to explore whether these codes have (or not) an impact for the identification of CD
5	ICD-9-CM: 555x (primary) <b>OR</b> Budesonide: 034734x, 036507022, 036507034, 036507059, 036507046, 036507010, 044798x	CD algorithm was performed from "Algorithm 4", replacing exemptions from healthcare service co-payment codes and methotrexate dispensing with CD-related ICD-9-CM codes to explore how the performance of the algorithm changed
Algorithms	Ulcerative colitis	
1	ICD-9-CM: 556x (primary) <b>OR</b> Exemptions from healthcare service co-payment: 009.556 OR 556	UC algorithm was adapted from the study conducted by Di Domenicantonio et al; <sup>20</sup> indeed, in that study, ICD9-CM diagnosis code of 556.xx (excluding 556.0, 556.1, 556.4, 556.8) as the primary or secondary diagnosis or the activation of co-payment exemption for UC (code 009.556) was used as a case definition criteria. In our algorithm, we decided to consider only ICD9-CM diagnosis code of 556.xx as primary cause of hospitalization to be sure that the hospital admission was related to UC
2	ICD-9-CM: 556x (primary) <b>OR</b> Exemptions from healthcare service co-payment: 009.556 OR 556 <b>OR</b> ((Exemptions from healthcare service co-payment: 009) AND (Golimumab 039541x [excl. 039541091]))	UC algorithm was performed from "Algorithm 1", adding Golimumab dispensing (considering only NDC codes approved for the treatment of UC and other IMIDs in Italy <sup>43</sup> ). The combination of both Golimumab dispensing AND the exemption from healthcare service co-payment code, identifying <i>unspecified IBD (009)</i> , could ensure the identification of UC (rather than other IMIDs)
3	ICD-9-CM: 556x (primary) <b>OR</b> Exemptions from healthcare service co-payment: 009.556 OR 556 <b>OR</b> ((Exemptions from healthcare service co-payment: 009) AND (Golimumab 039541x [excl. 039541091])) <b>OR</b> (Mesalazine specific AIC (see <a href="#">Supplement Table 2</a> ) <b>OR</b> (Balsalazide 033858x) <b>OR</b> (Budesonide: 043461x, 036507061, 036507085, 036507073)	UC algorithm was performed from "Algorithm 2", adding mesalazine/balsalazide/budesonide dispensing (considering only NDC codes with the only approval for the treatment of CD in Italy). <sup>42,44,45</sup> We decided to consider UC-related ICD-9-CM codes OR exemptions from co-payment codes OR non-biological drugs with indication for CD in order to explore all the conditions to identify UC patients from claims databases
4	Exemptions from healthcare service co-payment: 009.556 OR 556 <b>OR</b> ((Exemptions from healthcare service co-payment: 009) AND (Golimumab 039541x [excl. 039541091])) <b>OR</b> Mesalazine: specific AIC (see <a href="#">Supplement Table 2</a> ) <b>OR</b> Balsalazide: 033858x <b>OR</b> Budesonide: 043461x, 036507061, 036507085, 036507073	UC algorithm was performed from "Algorithm 3", excluding UC-related ICD-9-CM codes to explore whether these codes have (or not) an impact for the identification of UC
5	ICD-9-CM: 556x (primary) <b>OR</b> Mesalazine: specific AIC (see <a href="#">Supplement Table 2</a> ) <b>OR</b> Balsalazide: 033858x <b>OR</b> (Budesonide: 043461x, 036507061, 036507085, 036507073)	UC algorithm was performed from "Algorithm 4", replacing exemptions from healthcare service co-payment codes and golimumab dispensing with UC-related ICD-9-CM codes to explore how the performance of the algorithm changed

**Abbreviations:** NDC, National drug code; IMID, immune-mediated inflammatory disease.

exemptions from healthcare service co-payment codes (CD: 009.555; UC: 009.556) and biological and non-biological drugs dispensing with only indication for CD/UC were the criteria included into CD/UC specific coding algorithms. To select the specific medicinal product (National drug codes – NDC codes) with only indication for CD or UC, information from the “therapeutic indications” paragraph of the Summaries of Product Characteristics (SmPCs) for drugs known to be used in CD/UC was extracted.

All the algorithms identifying UC/CD separately were developed by combining in a different way the above-mentioned criteria. In the first algorithm, only CD/UC-related ICD-9-CM codes, as primary cause of hospital admission, and exemptions from co-payment codes were included. For algorithms 2 and 3, biological/non-biological drugs with only indication for CD or UC were added as well. Finally, in algorithms 4 and 5, biological/non-biological drugs and (CD/UC-related ICD-9-CM OR exemptions from co-payment codes) were respectively considered.

The five proposed algorithms were firstly performed on the gold standard cohorts and then tested on the larger cohort of users of biological drugs approved for IBDs (ie irrespective of the presence of any ETP, and thus of any gold standard information).

## Statistical Analysis

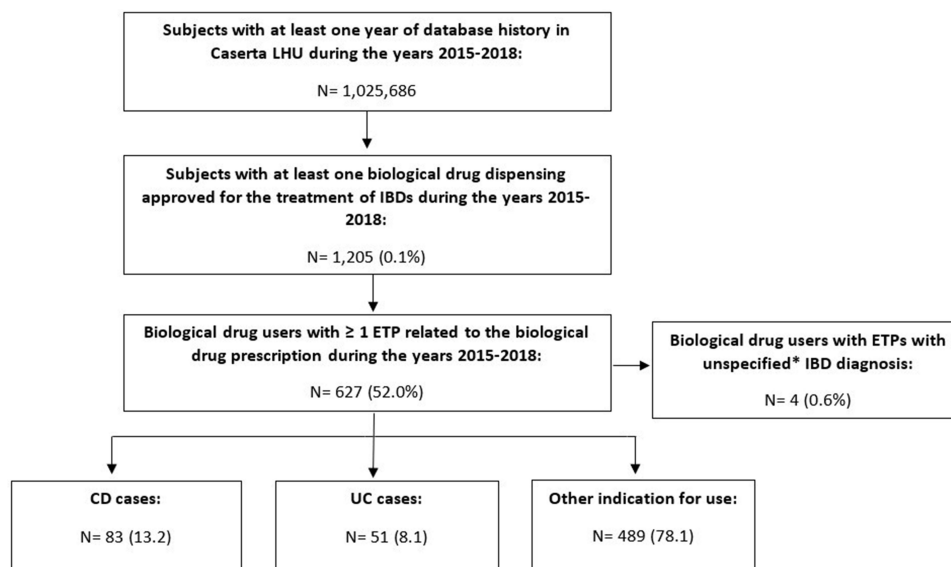
Clinical and demographic characteristics of gold standard cohorts of biological drug users with CD/UC vs other indications for use were reported as mean and standard deviation, as median along with interquartile range (ie first–third quartiles) and as absolute and relative (ie percentage) frequencies for continuous and categorical variables, respectively. To assess whether comparisons were meaningful or “important” in terms of the magnitude of the difference between mean, medians or percentages, the standardized mean difference or “standardized Z-score” from the Mann–Whitney *U*-test (ie Z-score divided by the square root of the total number of compared subjects) was reported as the effect size for continuous variables, respectively, whereas Cramér’s *V* was reported as the effect size for categorical variables. The diagnostic accuracy of each algorithm was assessed by Se, Sp, PPV and NPV, along with their 95% confidence interval (CI). For Se and Sp, the 95% CI was calculated using the “exact” Clopper–Pearson method, whereas for PPV and NPV the 95% CI was calculated following the standard logit transformation. Moreover, both the Youden Index and F-score were computed to evaluate the test’s diagnostic and prognostic accuracy of each algorithm for CD and UC, respectively. In particular, the Youden Index is computed as the sum of Se and Sp minus one<sup>29</sup> and was used to choose “the best algorithm” (ie the one that achieves the highest diagnostic accuracy) while the F-score is the harmonic mean of Se and PPV, using a weight of 0.5 to give more importance to the PPV with respect to Se, and was reported just for sake of completeness.

To show the overlapping number of CD and UC cases detected by each coding algorithm and through ETPs, Venn diagrams were produced. All analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and R version 4.0 (specifically, “caret” and “PropCIs” packages) (The R Foundation for Statistical Computing, Vienna, Austria).

## Results

Among a total population of more than 1 million inhabitants from Caserta LHU in the study period, 1205 (0.1%) subjects received at least one biological drug dispensing approved for the treatment of IBDs and 627 (52.0%) of them received at least one ETP. Of these, 138 (22%) biological drug users received ETPs reporting as indication CD (N=83; 60.1%) and UC (N=51; 36.9%), respectively (Figure 1). Specifically, 4 of them were excluded because they received the same number of ETPs for either CD or UC as indication for use. Therefore, the two gold standard cohorts finally consisted overall of 134 IBD patients (gold standard cases).

Clinical and demographic characteristics of the 623 biological drug users at the date of the first ETP prescription are shown in Table 2. Females with CD and UC were 36 (43.3%) and 23 (45.1%), respectively. The mean age ( $\pm$ SD) was 36 ( $\pm$ 15.4) years in CD and 43.6 ( $\pm$ 15.8) years in UC patients. Both CD and UC cohorts were significantly younger than those with other indications for use (49.5 $\pm$ 14.7 years). Respiratory disease (>62%) and neoplasms (>22%) were the most commonly reported comorbidities in all biological drugs users. CD patients were more likely to be treated with adalimumab (N=56; 67.5%) as index drug, followed by infliximab (N=20; 24.1%); while UC patients were mostly



**Figure 1** Flow-chart of patients identified in the gold standard cohort from Caserta LHM database in the period 2015–2018. \*Unspecified diagnosis=IBD as indication for use, not specified if CD or UC.

**Abbreviations:** LHM, Local Health Unit; IBDs, inflammatory bowel diseases; ETP, electronic therapeutic plan; CD, Crohn’s disease; UC, ulcerative colitis.

treated with infliximab (N=16; 31.4%), followed by adalimumab (N=14; 27.4%). Only one CD patient and no UC patients received ustekinumab, as it was approved for the treatment of CD in September 2018 and for the treatment of UC in October 2019 by the Italian Medicines Agency. It may not be traced in the Caserta database because the mean/

**Table 2** Baseline Characteristics of IBD Biological Drugs Users, Stratified by Indication for Use, as Reported in the First Electronic Therapeutic Plan

	Crohn’s Disease N=83	Ulcerative Colitis N=51	Other Indications for Use N=489	Crohn’s Disease vs Other Indications for Use (Effect Size <sup>a</sup> )	Ulcerative Colitis vs other Indications for Use (Effect Size <sup>a</sup> )
<b>Sex – N (%)</b>					
Male	47 (56.6)	28 (54.9)	231 (47.2)	(0.066)	(0.045)
Female	36 (43.4)	23 (45.1)	258 (52.8)		
<b>Age (years) – mean± SD</b>	36.4 (15.4)	43.6 (15.8)	49.5 (14.7)	(0.887)	(0.394)
<b>Age categories – N (%)</b>					
<45	61 (73.5)	25 (49.0)	158 (32.3)	(0.301)	(0.107)
45–64	19 (22.9)	21 (41.2)	253 (51.7)		
65–79	3 (3.6)	5 (9.8)	72 (14.7)		
≥80	0 (0)	0 (0)	6 (1.3)		
<b>Follow-up period (months) – median (Q1–Q3)</b>	24.6 (11.6–27.1)	22.9 (9.7–26.5)	25.3 (20.2–27.0)	(0.060)	(0.128)
<b>Index drug – N (%)</b>					
Adalimumab	56 (67.5)	14 (27.4)	343 (70.1)	(0.021)	(0.264)
Golimumab	0 (0)	12 (23.5)	85 (17.4)	(0.172)	(0.047)
Infliximab	20 (24.1)	16 (31.4)	32 (6.5)	(0.215)	(0.255)

(Continued)



Table 2 (Continued).

	Crohn's Disease N=83	Ulcerative Colitis N=51	Other Indications for Use N=489	Crohn's Disease vs Other Indications for Use (Effect Size <sup>#</sup> )	Ulcerative Colitis vs other Indications for Use (Effect Size <sup>#</sup> )
Vedolizumab	6 (7.2)	9 (17.6)	0 (0)	(0.250)	(0.403)
Ustekinumab	1 (1.2)	0 (0)	0 (0)	(0.102)	(0)
Other	0 (0)	0 (0)	29 (5.9)	(0.095)	(0.077)
<b>Comorbidities – N (%)<sup>a</sup></b>					
Hypertension	9 (10.8)	9 (17.6)	89 (18.2)	(0.069)	(0.004)
Diabetes mellitus	3 (3.6)	4 (7.8)	64 (13.1)	(0.104)	(0.046)
Respiratory disease	52 (62.6)	37 (72.5)	327 (66.9)	(0.031)	(0.035)
Renal failure	0 (0)	0 (0)	1 (0.2)	(0.017)	(0.014)
Neoplasms	27 (32.5)	16 (31.4)	109 (22.3)	(0.085)	(0.063)
Cerebrovascular disease	2 (2.4)	0 (0)	24 (4.9)	(0.042)	(0.070)
<b>Other autoimmune diseases</b>					
Spondyloarthritis	3 (3.6)	0 (0)	82 (16.8)	(0.130)	(0.137)
Psoriasis	5 (6.0)	5 (9.8)	217 (44.4)	(0.277)	(0.206)
Psoriatic arthritis	2 (2.4)	1 (1.9)	152 (31.1)	(0.228)	(0.189)
Rheumatoid arthritis	2 (2.4)	2 (3.9)	193 (39.5)	(0.275)	(0.216)
<b>Previous use of conventional drugs – N (%)<sup>a</sup></b>					
Aminosalicylic acid and similar agents	80 (96.4)	51 (100.00)	150 (30.7)	(0.472)	(0.419)
Immunosuppressants	50 (60.2)	30 (58.8)	366 (74.8)	(0.116)	(0.106)
Glucocorticoids	64 (77.1)	50 (98.0)	348 (71.2)	(0.047)	(0.179)
Corticosteroids acting locally	56 (67.4)	47 (92.1)	36 (7.4)	(0.576)	(0.688)
<b>Concomitant drugs – N (%)<sup>b</sup></b>					
Antithrombotic agents	11 (13.2)	3 (5.8)	78 (15.9)	(0.026)	(0.083)
Platelet aggregation inhibitors excl. heparin	3 (3.6)	2 (3.9)	57 (11.6)	(0.092)	(0.073)
Antiarrhythmics	4 (4.8)	5 (9.8)	112 (22.9)	(0.158)	(0.093)
Antibacterials for systemic use	65 (78.3)	37 (72.5)	335 (68.5)	(0.075)	(0.026)
Drugs for peptic ulcer and gastro-esophageal reflux disease (GORD)	34 (40.9)	35 (68.6)	294 (60.1)	(0.136)	(0.051)
Antivirals for systemic use	0 (0)	2 (3.9)	16 (3.3)	(0.070)	(0.011)
NSAIDs	5 (6.0)	11 (21.6)	269 (55.0)	(0.345)	(0.196)
Statins	3 (3.6)	3 (5.9)	71 (14.5)	(0.114)	(0.074)

**Notes:** <sup>a</sup>Evaluated any time prior to index date (ID), defined as the first date of the biologic dispensing. <sup>b</sup>Evaluated within one year prior to ID. <sup>#</sup>Effect sizes are reported as Cramér's V for categorical variables and standardized mean difference or "standardized Z-score" from the Mann-Whitney U-test for continuous variables as appropriate. SD, standard deviation; Q1–Q3: first–third quartiles; Antithrombotic agents: vitamin K antagonists, heparin group, direct thrombin inhibitors, direct factor Xa inhibitors, other antithrombotic agents; Aminosalicylic acid and similar agents: sulfasalazine, mesalazine, olsalazine, balsalazide; Corticosteroids acting locally: budesonide; Glucocorticoids: prednisone, hydrocortisone, methylprednisolone; Immunosuppressants: azathioprine, mercaptopurine, methotrexate, cyclosporine.

median times lag between the Italian Medicines Agency and Campania Drug Formulary Committee approval could reach some months.

Regarding previous use of conventional drugs, both CD and UC subjects showed higher use of aminosalicylic acid or similar agents (CD=96%; UC=100%), as compared to subjects with other indications for use (31%; Cramér's  $V>0.3$ ), and corticosteroids acting locally (CD=67%; UC=92%; other indications for use=7%; Cramér's  $V>0.3$ ).

As regarding concomitant drugs, both CD and UC subjects received platelet aggregation inhibitors excl. heparin, antiarrhythmics, NSAIDs and statins less commonly than those with other indications for use.

After testing the five coding algorithms among the 623 biological drug users, the best performing algorithm for CD (Youden Index=0.960) was defined as follows: specific CD diagnosis code (ie ICD-9-CM: 555.xx) as a primary cause of hospital admission OR specific exemption from healthcare service co-payment code (009.555/555) OR [non-specific exemption code (009) AND methotrexate dispensing only approved for CD] OR budesonide dispensing only approved for CD (Se: 98.8%; Sp: 97.2%; PPV: 84.5%; NPV: 99.8%) ("Algorithm 3"), as reported in [Table 3](#). Due to very high Se and PPV, its prognostic ability was also good (F-score=0.87). The number of true positive, false positive, true negative and false negative cases for each coding algorithm developed for identification of the 83 CD patients is reported in [Supplement Table 3](#).

As for UC, the best performing algorithm (Youden Index=0.857) was defined as follows: specific UC diagnosis code (ie ICD-9-CM: 556.xx) as a primary cause of hospital admission OR specific exemption from healthcare service co-payment code (009.556/556) OR [non-specific exemption code (009) AND golimumab dispensing] OR dispensing of non-biological drugs with only UC indication (ie mesalazine/balsalazide/budesonide) (Se: 94.1%; Sp: 91.6%; PPV: 50.0% NPV: 99.4%) ("Algorithm 3"), as reported in [Table 4](#). On the other hand, due to the low PPV, its prognostic ability was judged as fair (F-score=0.55). The number of true positive, false positive, true negative and false negative cases for each coding algorithm developed for identification of the 51 UC patients is reported in [Supplement Table 4](#). The PPVs related to the other algorithms developed for UC identification were generally low, ranging from 48.5% to 64.4%.

In general, 51 overlapping CD cases (ie, 61.4% of confirmed CD, [Figure 2A](#)) and 22 overlapping UC cases (ie, 43.1% of confirmed UC, [Figure 2B](#)) were detected by coding algorithms.

Finally, both "Algorithms 3" identifying CD and UC were tested on the whole cohort of 1205 users of biological drugs approved for IBDs (irrespective of the presence of any ETP). The algorithms identified 132/1205 (10.9%) and 138/1205 (11.4%) biological drug users treated for CD and UC, respectively ([Table 3](#) and [Table 4](#)); however, the highest number of false positive subjects identified as having CD/UC was reached (15 and 48 false positives for CD and UC, respectively), compared to the other algorithms ([Supplement Tables 3](#) and [4](#)).

**Table 3** Diagnostic Accuracy of the Coding Algorithms Developed to Identify Patients with Crohn's Disease in the Cohort of Biological Drug Users from Caserta LHU in the Years 2015–2018

Algorithms	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)	F-score	Youden Index	N of Identified CD Cases/Users of Biological Drugs Approved for IBDs (%)
1	91.6 (83.4–96.5)	98.1 (96.6–99.1)	88.4 (80.4–93.4)	98.7 (97.4–99.4)	0.89	0.897	116/1205 (9.6)
2	91.6 (83.4–96.5)	97.8 (96.1–98.8)	86.4 (78.3–91.7)	98.7 (97.4–99.3)	0.87	0.894	119/1205 (9.9)
3	98.8 (93.5–100.0)	97.2 (95.5–98.4)	84.5 (76.8–90.1)	99.8 (98.7–100.0)	0.87	0.960	132/1205 (10.9)
4	71.1 (60.1–80.5)	97.9 (96.4–98.9)	84.3 (74.6–90.7)	95.7 (94.0–96.9)	0.81	0.690	96/1205 (7.9)
5	96.4 (89.8–99.2)	97.8 (96.1–98.8)	86.9 (79.2–92.1)	99.4 (98.3–99.8)	0.88	0.942	127/1205 (10.5)

**Abbreviations:** CD, Crohn's disease; CI, confidence interval; ETP, electronic therapeutic plan; IBDs, inflammatory bowel diseases.



**Table 4** Diagnostic Accuracy of the Coding Algorithms Developed to Identify Patients with Ulcerative Colitis in the Cohort of the Biological Drug Users from Caserta LHU in the Years 2015–2018

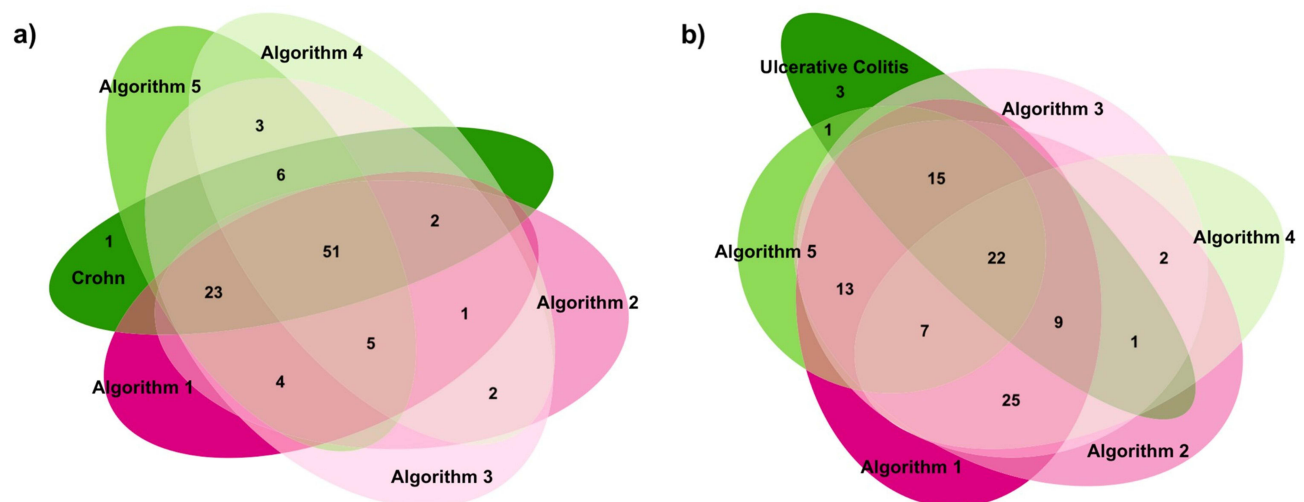
Algorithms	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)	F-Score	Youden Index	N of Identified UC Cases/Users of Biological Drugs Approved for IBDs (%)
1	90.2 (78.6–96.7)	92.1 (89.6–94.2)	50.5 (43.2–57.8)	99.1 (97.9–99.6)	0.55	0.823	133/1205 (11.0)
2	92.2 (81.1–97.8)	91.8 (89.2–93.9)	50.0 (42.9–57.1)	99.2 (98.1–99.7)	0.55	0.840	135/1205 (11.2)
3	94.1 (83.8–98.8)	91.6 (89.1–93.7)	50.0 (43.1–56.9)	99.4 (98.3–99.8)	0.55	0.857	138/1205 (11.4)
4	64.7 (50.1–77.6)	93.9 (91.6–95.7)	48.5 (39.2–58.0)	96.8 (95.4–97.8)	0.51	0.586	89/1205 (7.4)
5	74.5 (60.4–85.7)	96.3 (94.4–97.7)	64.4 (53.6–73.9)	97.7 (96.4–98.5)	0.66	0.708	98/1205 (8.1)

**Abbreviations:** CI, confidence interval; ETP, electronic therapeutic plan; IBDs, inflammatory bowel diseases; UC, ulcerative colitis.

## Discussion

To our knowledge, this is the first Italian population-based study assessing and comparing the accuracy of different coding algorithms for the identification of UC and CD, separately, as indication for use in a cohort of users of biological drugs approved for the treatment of IBDs, using claims databases. Testing of any coding algorithms for the disease identification using claims databases are two fundamental steps that must be followed before the validation process. Several validation studies aimed at detecting a reliable diagnosis of IBD within different claims databases were conducted,<sup>30–35</sup> but preexisting differences in the structure of national data sources have to be considered when specific algorithms are developed.

Two population-based studies using claims databases validated algorithms identifying IBD patients.<sup>33,34</sup> However, both studies were aimed at identifying IBD patients from the general population (not specifically from a cohort of biological drug users, as indication for use); moreover, Rezaie et al did not distinguish specifically patients with Crohn's disease from those with ulcerative colitis.<sup>33</sup> Other population-based studies were aimed at identifying CD/UC cases from the general population (not specifically from a cohort of biological drug users), but not aimed at testing and validating



**Figure 2** Frequency distribution (Venn diagrams) of the number of biological drug users identified by the coding algorithms developed for Crohn's disease (a) and ulcerative colitis (b) identification.

Notes: The dark green sets showed the true positive cases for Crohn's disease (a) and ulcerative colitis (b), identified through electronic therapeutic plans. The other sets showed the Crohn's disease (a) and ulcerative colitis (b) positive cases identified by the 5 different tested algorithms. Full details about algorithms definition are reported in Table 1.

algorithms identifying CD/UC.<sup>30–32,35</sup> In particular, Lewis et al identified IBD patients from the UK General Practice Research Database, that is a reliable database for research purposes, but this is not a claims database.<sup>32</sup>

Moreover, a previous Italian study using the Lazio region claims databases to test an algorithm identifying CD/UC patients showed a good sensitivity; however, the authors did not focus on the identification of IBD as indication for use of biological drugs specifically.<sup>20</sup>

Therefore, in general, all the published studies were not aimed at identifying CD or UC diagnosis, separately, specifically from a cohort of biological drug users, contrarily to our study focusing specifically on the CD/UC identification as indication for use of biological drugs approved for IBD, using claims databases.

All of the five algorithms tested were able, to different extent, to identify true CD/UC patients from claims databases. Based on this, our study results clearly show that “Algorithms 3” achieved the best performance in diagnostic accuracy to identify biological drug users with CD or UC. Both “Algorithms 3” considered CD/UC-related ICD-9-CM codes as primary cause of hospital admissions, exemption from healthcare service co-payment code and biological/non-biological drugs dispensing with only indication for CD/UC.

As regards CD, “Algorithms 3” showed the highest Se but lower Sp, compared to the other algorithms tested for CD identification. On the other hand, excluding the use of biological/non-biological drugs dispensing with only indication for CD as a proxy of disease (ie, “Algorithm 1”), higher Sp and PPV, compared to the other algorithms, were achieved.

As regards UC, the highest Sp and PPV values were achieved with “Algorithm 5”, including UC-related ICD-9-CM codes as primary cause of hospital admissions and non-biological drugs dispensing with only indication for UC.

Interestingly, the exclusion of ICD-9-CM codes may decrease the Se of algorithms tested. Indeed, both “Algorithms 4”, which did not include information regarding CD/UC-related ICD-9-CM codes, showed the lowest Se.

In general, adding biological/non-biological drugs dispensing, as a proxy of IBD identification, together with ICD-9-CM codes as cause of hospital admissions and exemption from healthcare service co-payment code (just for Italian claims databases) increased the performance of algorithms identifying CD/UC, as reported in previous studies.<sup>17,20</sup>

In general, our study suggested the importance of defining standardized algorithms in order to identify the exact indication for use of biological drugs approved for both CD and UC.

The application of the two “Algorithms 3” (identifying CD and UC, respectively) on the users of biological drugs approved for IBDs from Caserta LHU database yielded almost 11% of detected cases, as expected.<sup>36</sup>

Our study highlights the importance of selecting an appropriate algorithm to identify a chronic disease from different data sources. Using only IBD-related medications as a proxy of IBD or only CD/UC-related ICD-9-CM codes from hospital admissions (without any other claims as a proxy of disease) may identify a high number of patients recorded as having IBD.<sup>37,38</sup> However, considering specific ICD-9-CM codes together with information on IBD-related medication and exemption codes had the advantage to increase the accuracy of the algorithms without decreasing the sensitivity.<sup>17</sup>

This study has some limitations to take into consideration. First, due to the nature of the claims database and the potential inaccurate coding of the CD/UC diagnosis during routine clinical practice, a potential misclassification of the diagnosis could be observed.<sup>39</sup> However, the study patients are subject treated with biological drugs recommended to treat severe patients to induce remission in the short term and maintain remission in the long term, according to the guidelines.<sup>10,11</sup> A second limitation is the unavailability of an IBD register that could better validate our estimates. However, our study provides a reasonably accurate evaluation of the sensitivity of claims databases in identifying users of biological drugs treated for CD/UC. Third, the present study highlights the value of different claims data to identify IBDs, and specifically CD and UC separately. The inconsistencies in clinical practice regarding differential diagnosis of CD and UC and related treatment may lead to irregular coding of ICD-9-CM resulting in patients being registered with multiple diagnoses and having ETPs both for CD than UC.<sup>40</sup> Fourth, since biological drugs are mainly indicated in those patients who have not responded to conventional therapy or who are intolerant or have contraindications for such therapies, the extraction of the gold standard cohort from the ETP database may not be fully representative of CD/UC. Nevertheless, this has unlikely affected the comparison of the accuracy of different coding algorithms for CD/UC identification. Finally, validation of identified algorithms in an external population was not possible; this may represent a limitation because the estimates of validity indices rely on the subjective judgment of the local expert. Nevertheless, in

absence of previously validated algorithms identifying CD/UC, separately, as indication for use, in a cohort of users of biological drugs approved for the treatment of IBDs, this approach could be helpful.

## Conclusions

The identification of Crohn's disease and ulcerative colitis, as indication for use of biological drugs approved for the treatment of inflammatory bowel disease, using real-world data is challenging. In a population-based claims database from southern Italy, new coding algorithms including diagnostic and exemption codes plus specific drug dispensing yielded highly accurate identification of CD and UC as distinct indication for use of biological drugs. This study showed that robust validity testing may yield the identification of accurate coding algorithms. The tested algorithms should be validated in an external cohort of biological drug users, in order to use them in future studies to identify CD/UC patients treated with biological drugs.

## Ethics Approvals

This study was conducted in the context of the multiregional active pharmacovigilance VALORE project, funded by the Italian Medicines Agency. This retrospective study protocol was notified to the Ethical Committee of the Academic Hospital of Verona, which was the project's coordinating center at the time of protocol submission, according to the current national law. The manuscript does not contain clinical studies or patient data. Formal consent is not required for this type of study.

## Funding

This study was funded by the Italian Medicines Agency in the context of the multiregional pharmacovigilance project (AIFA 2012–2014: Post-marketing evaluation of the benefit–risk profile of originator biological drugs and biosimilars in the dermatological, rheumatological, gastroenterological and onco-hematological areas through the establishment of a single multiregional network for the integrated analysis of data from health databases, active surveillance and clinical registers; VALORE project).

## Disclosure

Y.I. is the CEO of the academic spin-off “INSPIRE srl” of the University of Messina, which has received funding for conducting observational studies from contract research organizations (RTI Health Solutions, Pharmo Institute N.V.) and from pharmaceutical Companies (Chiesi Italia, Kyowa Kirin s.r.l., Daiichi Sankyo Italia S.p.A.). G.T. has served on advisory boards/seminars funded by SANOFI, Eli Lilly, AstraZeneca, Abbvie, Servier, Mylan, Nova Nordisk, Gilead and Amgen; he was the scientific director of a II level Master on pharmacovigilance, pharmacoepidemiology and real-world evidence, which has received non-conditional contributions from various pharmaceutical companies; he coordinates a pharmacoepidemiology team at the University of Messina, which has received funding for conducting observational studies from various pharmaceutical companies (Boehringer Ingelheim, Daichii Sankyo, PTC Pharmaceuticals); and he is also chief of the academic spin-off “INSPIRE srl”, which has received funding for conducting observational studies from contract research organizations (RTI Health Solutions, Pharmo Institute N.V.) from pharmaceutical Companies (Chiesi Italia, Kyowa Kirin s.r.l., Daiichi Sankyo Italia S.p.A.). The authors report no other conflicts of interest in this work.

## References

1. Kappelman MD, Rifas-Shiman SL, Kleinman K, et al. The prevalence and geographic distribution of Crohn's disease and ulcerative colitis in the United States. *Clin Gastroenterol Hepatol.* 2007;5(12):1424–1429.
2. Romano C, Famiani A, Gallizzi R, Comito D, Ferrau' V, Rossi P. Indeterminate colitis: a distinctive clinical pattern of inflammatory bowel disease in children. *Pediatrics.* 2008;122(6):e1278–81. doi:10.1542/peds.2008-2306
3. Alatab S, Sepanlou SG, Ikuta K, GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol.* 2020;5(1):17–30. doi:10.1016/S2468-1253(19)30333-4
4. Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology.* 2012;142:46–54. doi:10.1053/j.gastro.2011.10.001

5. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel disease. *Gastroenterology*. 2011;140:1785–1794. doi:10.1053/j.gastro.2011.01.055
6. Burisch J, Pedersen N. East-West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. *Gut*. 2014;63:588–597. doi:10.1136/gutjnl-2013-304636
7. Lakatos L, Kiss LS, David G, et al. Incidence, disease phenotype at diagnosis, and early disease course in inflammatory bowel diseases in Western Hungary, 2002–2006. *Inflamm Bowel Dis*. 2011;17:2558–2565. doi:10.1002/ibd.21607
8. 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):2769–2778. doi:10.1016/S0140-6736(17)32448-0
9. Galeone C, Pelucchi C, Barbera G, Citterio C, La Vecchia C, Franchi A. Crohn's disease in Italy: a critical review of the literature using different data sources. *Dig Liver Dis*. 2017;49:459–466. doi:10.1016/j.dld.2016.12.033
10. Torres J, Bonovas S, Doherty G, et al. ECCO Guidelines on Therapeutics in Crohn's Disease: medical Treatment. *J Crohns Colitis*. 2020;14(1):4–22. doi:10.1093/ecco-jcc/jjz180
11. Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD, Clinical Guideline ACG. Ulcerative Colitis in Adults. *Am J Gastroenterol*. 2019;114(3):384–413. doi:10.14309/ajg.000000000000152
12. Trifirò G, Gini R, Barone-Adesi F, et al. The Role of European Healthcare Databases for Post-Marketing Drug Effectiveness, Safety and Value Evaluation: where Does Italy Stand? *Drug Saf*. 2019;42(3):347–363. doi:10.1007/s40264-018-0732-5
13. Ingrasciotta Y, Giorgianni F, Bolcato J, et al. How much are biosimilars used in clinical practice? A retrospective Italian population-based study of erythropoiesis-stimulating agents in the years 2009–2013. *BioDrugs*. 2015;29(4):275–284. doi:10.1007/s40259-015-0132-7
14. Marcianò I, Ingrasciotta Y, Giorgianni F, et al. How did the introduction of biosimilar filgrastim influence the prescribing pattern of granulocyte colony stimulating factors? Results from a multicentre, population-based study, from five Italian centres in the years 2009–2014. *BioDrugs*. 2016;30(4):295–306. doi:10.1007/s40259-016-0175-4
15. Trifirò G, Isgrò V, Ingrasciotta Y, et al. VALORE Project Collaborators. Large-Scale Postmarketing Surveillance of Biological Drugs for Immune-Mediated Inflammatory Diseases Through an Italian Distributed Multi-Database Healthcare Network: the VALORE Project. *BioDrugs*. 2021;35(6):749–764. doi:10.1007/s40259-021-00498-3
16. Trifirò G, Isgrò V, Ingrasciotta Y, et al.; VALORE Project Collaborators. Large-Scale Postmarketing Surveillance of Biological Drugs for Immune-Mediated Inflammatory Diseases Through an Italian Distributed Multi-Database Healthcare Network: the VALORE Project. *BioDrugs*. 2021;35(6):749–764.
17. Benchimol EI, Guttman A, Mack DR, et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. *J Clin Epidemiol*. 2014;67(8):887–896. doi:10.1016/j.jclinepi.2014.02.019
18. Thirumurthi S, Chowdhury R, Richardson P, Abraham NS. Validation of ICD-9-CM diagnostic codes for inflammatory bowel disease among veterans. *Dig Dis Sci*. 2010;55(9):2592–2598. doi:10.1007/s10620-009-1074-z
19. Lee H, Ford JA, Jin Y, et al. Validation of claims-based algorithms for psoriatic arthritis. *Pharmacoepidemiol Drug Saf*. 2020;29(4):404–408. doi:10.1002/pds.4950
20. Di Domenicantonio R, Cappai G, Arcà M, et al. Occurrence of inflammatory bowel disease in central Italy: a study based on health information systems. *Dig Liver Dis*. 2014;46(9):777–782. doi:10.1016/j.dld.2014.04.014
21. 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. doi:10.19191/EP19.4.S2.P008.089
22. Ye Y, Manne S, Bennett D. Identifying Patients With Inflammatory Bowel Diseases in an Administrative Health Claims Database: do Algorithms Generate Similar Findings? *Inquiry*. 2019;56:46958019887816. doi:10.1177/0046958019887816
23. Ingrasciotta Y, Sultana J, Giorgianni F, et al. The burden of nephrotoxic drug prescriptions in patients with chronic kidney disease: a retrospective population-based study in Southern Italy. *PLoS One*. 2014;9(2):e89072. doi:10.1371/journal.pone.0089072
24. Ingrasciotta Y, Giorgianni F, Bolcato J, et al. How Much Are Biosimilars Used in Clinical Practice? A Retrospective Italian Population-Based Study of Erythropoiesis-Stimulating Agents in the Years 2009–2013. *BioDrugs*. 2015;29(4):275–284.
25. Ingrasciotta Y, Lacava V, Marcianò I, et al. In search of potential predictors of erythropoiesis-stimulating agents (ESAs) hyporesponsiveness: a population-based study. *BMC Nephrol*. 2019;20(1):359. doi:10.1186/s12882-019-1554-0
26. Sultana J, Hurtado I, Bejarano-Quisoboni D, et al. Antipsychotic utilization patterns among patients with schizophrenic disorder: a cross-national analysis in four countries. *Eur J Clin Pharmacol*. 2019;75(7):1005–1015. doi:10.1007/s00228-019-02654-9
27. Crisafulli S, Fontana A, L'Abbate L, et al. Development and testing of diagnostic algorithms to identify patients with acromegaly in Southern Italian claims databases. *Sci Rep*. 2022;12(1):15843. doi:10.1038/s41598-022-20295-4
28. Lee CK, Ha HJ, Oh SJ, et al. Nationwide validation study of diagnostic algorithms for inflammatory bowel disease in Korean National Health Insurance Service database. *J Gastroenterol Hepatol*. 2020;35(5):760–768. doi:10.1111/jgh.14855
29. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
30. Bernstein CN. "Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population based study. *Am J Epidemiol*. 1999;149:916–924. doi:10.1093/oxfordjournals.aje.a009735
31. Fonager K, Sørensen RT, Rasmussen SN, et al. Assessment of the diagnoses of Crohn's disease and ulcerative colitis in a Danish hospital information system. *Scand J Gastroenterol*. 1996;31:pp. 154–159. doi:10.3109/00365529609031980
32. Lewis JD, Brensinger C, Bilker WB, Strom BL. Validity and completeness of the General Practice Research Database for studies of inflammatory bowel disease. *Pharmacoepidemiol Drug Saf*. 2002;11:pp. 211–218. doi:10.1002/pds.698
33. Rezaie A, Quan H, Fedorak RN, et al. Development and validation of an administrative case definition for inflammatory bowel diseases. *Can J Gastroenterol*. 2012;26(10):711–717. doi:10.1155/2012/278495
34. Benchimol EI, Guttman A, Mack DR. Validation of International Algorithms to Identify Adults with Inflammatory Bowel Disease in Health Administrative Data from Ontario, Canada. *J Clin Epidemiol*. 2014;vol. 67(8):887–896.
35. Coward S, Clement F, Benchimol EI, et al. Past and Future Burden of Inflammatory Bowel Diseases Based on Modeling of Population-Based Data. *Gastroenterology*. 2019;156(5):1345–1353.e4. doi:10.1053/j.gastro.2019.01.002
36. Degli Esposti L, Sangiorgi D, Perrone V, et al. Adherence and resource use among patients treated with biologic drugs: findings from BEETLE study. *Clinico Econ Outcomes Res*. 2014;18(6):401–407. doi:10.2147/CEOR.S66338

37. 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. doi:10.1186/1471-2458-11-688
38. Kim JW, Lee CK, Rhee SY, Oh CH, Shim JJ, Kim HJ. Trends in health-care costs and utilization for inflammatory bowel disease from 2010 to 2014 in Korea: a nationwide population-based study. *J Gastroenterol Hepatol*. 2018;33(4):847–854. doi:10.1111/jgh.14027
39. Ladha KS, Eikermann M. Codifying healthcare--big data and the issue of misclassification. *BMC Anesthesiol*. 2015;15:179. doi:10.1186/s12871-015-0165-y
40. Gunaseelan V, Kenney B, Lee JS, Hu HM. Databases for surgical health services research: clinformatics Data Mart. *Surgery*. 2019;165(4):669–671. doi:10.1016/j.surg.2018.02.002
41. Italian Medicines Agency. Methotrexate SPC. Available from: <https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/farmaco?farmaco=033647>. Accessed October 28, 2022.
42. Italian Medicines Agency. Budesonide SPC. Available from: <https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/farmaco?farmaco=034734#>. Accessed October 28, 2022.
43. Italian Medicines Agency. Golimumab SPC. Available from: <https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/farmaco?farmaco=039541>. Accessed October 28, 2022.
44. Italian Medicines Agency. Balsalazide SPC. Available from: <https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/farmaco?farmaco=033858>. Accessed October 28, 2022.
45. Italian Medicines Agency. Mesalazine SPC. Available from: [https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/cerca-per-principio-attivo?princ\\_att=MESALAZINA%20\(5-ASA\)](https://farmaci.agenziafarmaco.gov.it/bancadatifarmaci/cerca-per-principio-attivo?princ_att=MESALAZINA%20(5-ASA)). Accessed October 28, 2022.

## Clinical Epidemiology

Dovepress

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

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