A Systematic Review of Crohn's Disease Case Definitions in Administrative or Claims Databases

Susan Hutfless, PhD,[*,†,] Ryan A Jasper,[‡] Aman Tilak, MBBS,[§] Tamoghna Ghosh, MBBS,[§]
Saurabh Kedia, MBBS,[¶] Simon Liu, BS,[∥] Nathalie H Urrunaga, MD, MS,[**]
Matthew Josephson, MD,[††] Arshit Narang, MBBS,[§] Steve Miller, MD, MBE,[‡‡]
Po-Hung Chen, MD,[*,] Shelly Joseph, MD,[§§] and Steven R. Brant, MD[*,¶¶]

From the *Division of Gastroenterology and Hepatology, Department of Medicine, Johns Hopkins Medical Institutions, Baltimore, MD, USA
[†]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
[‡]Department of Applied Mathematics and Statistics, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA
[§]Department of Medicine, All India Institute of Medical Sciences, New Delhi, India
[¶]Department of Gastroenterology, All India Institute of Medical Sciences, New Delhi, India
[∥]Medical Genomics Unit, National Human Genome Research Institute, Bethesda, MD, USA
[**]Division of Gastroenterology and Hepatology, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA
[††]Division of Gastroenterology and Hepatology, Albany Medical College, Albany, NY, USA
[‡‡]Division of Pediatric Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[§§]Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, NYU Grossman School of Medicine, New York, NY, USA
[¶¶]Division of Gastroenterology and Hepatology, Department of Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA
Address correspondence to: Susan Hutfless, PhD, 1830 E Monument Street, Suite 427, Baltimore, MD 21287, USA (shutfle1@jhmi.edu).

Background: We sought to review Crohn's disease (CD) case definitions that use diagnosis, procedure, and medication claims.

Methods: We searched PubMed and Embase from inception through January 31, 2022, using terms related to CD, inflammatory bowel disease, administrative claims, or validity. Each article was scrutinized by 2 authors independently screening and abstracting data. Collected data included participant characteristics, case definition characteristics, and case definition validity. When diagnostic accuracy was provided for multiple case definitions, we extracted the case definition selected by the authors. All diagnostic accuracy characteristics were captured.

Results: We identified 30 studies that evaluated a case definition using claims data to identify CD patients. The most common case definition included counts of diagnosis codes (57%) followed by a combination of diagnosis codes and medications (20%). All but 1 study validated the case definition with a medical chart review. In 2 studies, the patient's primary care provider completed a survey to confirm disease status. The positive predictive value of the case definitions ranged from 18% (≥1 code at a single U.S. health plan) to 100% (≥1 code plus a relevant prescription at a U.S. hospital). More complex case definitions (eg, ≥1 code + prescription or ≥2 codes) had lower variability in positive predictive value (≥80%) and specificity (≥85%) than the ≥1 code requirement.

Conclusions: Health services researchers should validate case definitions in their research cohorts. When such validation cannot be performed, we recommend using a more complex case definition. Studies without a validated CD case definition should use sensitivity analyses to confirm the robustness of their results.

Lay Summary

This systematic review of Crohn's disease (CD) case definitions identified that complex case definitions such as ≥1 diagnosis code + ≥1 prescription had desirable diagnostic accuracy properties.

Introduction

Crohn's disease (CD) and ulcerative colitis (UC) together are the 2 major phenotypes of idiopathic chronic inflammatory disease of the intestinal tract that compose inflammatory bowel disease (IBD). The common practice in clinical literature to combine CD and UC under the rubric IBD minimizes the important genetic, epidemiologic, symptomatic, and

treatment differences between the diagnoses. While some clinicians and researchers argue that CD and UC exist on a spectrum, others argue that they are separate diseases.[1–3] Clinical trials and cohort studies must create inclusion criteria that differentiate CD, UC, and IBD based on numerous clinical classification systems that use disease signs, symptoms, endoscopy, pathology, or their combination. Clinicians often

---

Key Points:

What is already known?

- Health services researchers define Crohn's disease (CD) cases in their research using case definitions that combine diagnosis, procedure, and medication records. Informaticians recommend validating case definitions in the cohort under study when possible, but not all data sources allow researchers to identify individual patients to perform such validation.

What is new here?

- This systematic review identified 30 studies that validated case definitions.
- Cases with ≥1 CD diagnosis code and ≥1 CD prescription had ≥80% positive predictive value and ≥85% specificity. For resources that do not have access to prescription information, ≥2 CD diagnosis code may be an alternative.
- When case definitions cannot be validated for new CD cohorts, complex case definitions are recommended.

How can this study help patient care?

- Studies that rely on diagnosis, procedure, and medication records can often examine thousands or hundreds of thousands of CD patients concurrently. These large database projects allow researchers to study patient-important questions about rare events and subgroups of patients that would be impossible to conduct as randomized trials.

---

use the same classification systems, informally or formally, to classify disease in medical and billing records. These medical and billing records are increasingly used to perform comparative effectiveness research. Health services researchers use the codes associated with CD and UC to create cohorts and report associations that impact clinical care and health policy.

Access to large databases of patient histories documented in codes has become increasingly available throughout the world with the adoption of electronic medical records. This has allowed clinical researchers to study larger numbers of patients. However, the diagnostic accuracy of case definitions varies widely depending on the context in which codes are used (eg, tracking for payment vs clinical record keeping, coding for payment purposes or medication access). Informaticians recommend studies use a validated case definition that has been confirmed in that study's population; however, clinician investigators do not always observe this best practice.[4–10] Validation is generally performed by comparing database disease codes with medical record review. Some studies survey the clinicians who provide care for the patients to determine their clinical diagnoses instead of or in addition to medical record review. Diagnostic accuracy can then be calculated using the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), allowing researchers to select the best code-based case definition for their cohort.

Some researchers use aggregated billing records (called claims) purchased from insurers or other companies. Unfortunately, these resources often do not include medical records or physician identifiers, which prevents researchers from validating case definitions in compliance with the RECORD (REporting of studies Conducted using

Observational Routinely-collected health Data) checklist.[11] However, these resources may still be useful for research. If the validated case definitions from existing sources appear to be similar across different database sources and time, validation of new cohorts may not be needed. Similarly, smaller numbers of patients could be included in validation studies of the new cohorts. For example, if most validation studies find that only 1 CD encounter sufficiently identifies true CD cases (ie, have an acceptable range of diagnostic accuracy), then it may be acceptable for additional studies to use 1 CD encounter as the code-based definition for their cohorts.

We aimed to identify all studies that used codes to identify CD patients. We compared the similarities and differences in the case definitions selected from the validation studies. We then synthesized the evidence to make recommendations on the need for validation studies in new cohorts and reporting strategies for those with limited or no access to records for independent validation.

Methods

We searched the Embase and MEDLINE (via PubMed) databases from inception through January 31, 2022. Covidence software (Covidence, Melbourne, Australia) was used for data management. We combined the concepts of study design ("population-based cohort study" or administrative or ICD or claims or Medicare or Medicaid or military or (validated and cohort) or (validity and cohort)) and study population (Crohn's or "inflammatory bowel disease") to identify potentially relevant articles (Supplementary Appendix Table 1). We then hand-searched the included articles to identify additional validation studies. No year of publication or language restrictions were applied. Title/abstract, and full-text screening was performed using Covidence software. The title/abstract level of review included all studies that may have used codes to identify CD or IBD. The full-text level of review entailed uploading the article's full text article into Covidence for dual independent review. An article was eligible if it reported on CD separately from UC, used codes to define CD, reported the number of CD patients, and provided validation for the CD case definition. Conference abstracts, review articles, and letters to the editor were excluded. Dual independent data extraction was performed using the data comparison function in REDCap (Research Electronic Data Capture).[12] At all phases of review and extraction, 2 authors independently reviewed each article and a third author adjudicated the disagreements. For the detailed synthesis of the chart validation, 1 author synthesized the text and another author reviewed and confirmed the synthesis with the original full-text using serial extraction. Only validation studies that reported at least 1 measure of diagnostic accuracy including sensitivity, specificity, PPV, NPV, accuracy, area under the curve, or likelihood ratio for CD were eligible for synthesis.

We extracted information on characteristics of the study population including the location, years, and population inclusion characteristics. We also collected information about the validation including coding system, validation source (eg, medical records, surveys), and diagnostic accuracy statistics. When the authors reported the diagnostic accuracy for multiple case definitions, we extracted the range of diagnostic accuracy for the definitions considered. In those instances, our tables included the study authors' preferred

Table 1. Crohn's Disease Algorithm Case Definitions, Characteristics of Validation Studies for Each Definition, and Studies Citing the Definition.

Definition	Specific Definition and Diagnostic Accuracy	Characteristics of Validation Study	Reference
≥1 ICD diagnosis code	≥1 ICD-8 563.0 or ICD-9 555.x code during hospitalization No mention of exclusions PPV 87%	**United Kingdom** Barton 1989 Validation years: 1968-1983 Inpatient only No minimum follow-up Pediatric (≤20 y)	37
≥1 ICD diagnosis code	≥1 ICD-8 563.01 during hospitalization If 563.19 or 569.04 then exclude Sensitivity 94%	**Denmark** Fonager 1996 Validation years: 1988-1992 Inpatient only No minimum follow-up No age restriction	58
≥1 ICD diagnosis code	≥1 ICD-9 555.x during hospitalization If 556.x then exclude Sensitivity 82%	**Italy** Di Domenicantonio 2014 Validation years: 2000-2009 Inpatient only No minimum follow-up No age restriction	18
≥1 ICD diagnosis code	≥1 ICD-9 555.x in inpatient or outpatient setting No mention of exclusions PPV 18%-67%	**United States** Herrinton 2007 Validation years: 1999-2001 Inpatient or outpatient No minimum follow-up No age restriction	45
≥1 ICD diagnosis code	≥1 ICD-9 555.x in inpatient or outpatient setting No mention of exclusions Sensitivity 83% PPV 74%-82%	**United States** Herrinton 2008 Validation years: 1996-2002 Inpatient or outpatient ≥12 mo of enrollment No age restriction	55
≥1 ICD diagnosis code	≥1 ICD-9 555.0, 555.1, 555.2 or 555.9 in inpatient or outpatient setting No mention of exclusions for UC. Excluded unspecified intestinal obstruction (560.9). Sensitivity 92% Specificity 99% PPV 88% NPV 99%	**United States** Thirumurthi 2010 Validation years: 2000-2004 Inpatient or outpatient No minimum follow-up Adults (military veterans)	34
≥1 ICD diagnosis code	≥1 ICD-10 K50.x as primary diagnosis during hospitalization No mention of exclusions Sensitivity 98% NPV 99%	**Canada** Stepaniuk 2015 Validation years: 2007-2012 Inpatient only No minimum follow-up Adults (≥65 y)	8
≥1 ICD diagnosis code	≥1 ICD-10 K50.0, K50.1 or K50.8 during hospitalization If K51.x then exclude Sensitivity 30%-95% Specificity 89%-99% PPV 67%-97% NPV 66%-99%	**Canada** Ma 2017 Validation year: 2011 Inpatient only No minimum follow-up Adults (≥18 y)	44
≥1 ICD diagnosis code	≥1 ICD-10-CM code No mention of exclusions Sensitivity: 95% Specificity: 93% PPV: 91% NPV: 97%	**Denmark** Lo 2020 Validation years: 2003-2011 Inpatient and outpatient No minimum follow-up No age restriction	1
≥1 ICD diagnosis code and prescription	≥1 ICD-10-CM K50.x code with 0 K51.x and 0 M35.x and ≥1 medication If K51.x or M35.x then exclude Sensitivity: 100% Specificity: 99% PPV: 97% NPV: 100%	**Japan** Morikubo 2021 Validation years: 2015-2019 Inpatient and Outpatient No minimum follow-up No age restriction	0

Table 1. Continued

Definition	Specific Definition and Diagnostic Accuracy	Characteristics of Validation Study	Reference
≥1 ICD diagnosis code	≥1 OXMIS/Read code 5630CR, 5630C, 5630ER, 0092LR in primary care records If 5631 then exclude Sensitivity 94%	**United Kingdom** Lewis 2002 Validation year: 1988-1997 Primary care only No minimum follow-up No age restrictions	46
≥1 ICD diagnosis code and ≥1 prescription	≥1 ICD-9 555.x in electronic health record AND ≥1 mention of 5-aminosalicylate, antibiotic, corticosteroid, immunomodulator, anti-TNF or natalizumab Excluded if had code for rheumatoid arthritis (714.x) or multiple sclerosis (340, 341.9, 323.9). Determined CD vs UC based on proportion. PPV 100%	**United States** Restrepo 2016 Validation years: Not reported Electronic health record (presumably inpatient or outpatient) No minimum follow-up No age restriction	6
≥1 ICD diagnosis code and ≥1 prescription	≥1 code and recorded in the SWIBREG Quality Register and received a biologic No mention of exclusions Sensitivity: 90% Specificity: 96% PPV: 97%	**Sweden** Shrestha 2020 Validation years: 1999-2017 Included in Register No minimum follow-up No age restriction	3
≥1 ICD diagnosis code and ≥1 prescription	≥1 ICD-10 K50.x code and ≥1 prescription for 5-aminosalicylate, immunomodulator, anti-TNF No mention of exclusions Sensitivity: 98% Specificity: 98% PPV: 98% NPV: 98%	**South Korea** Lee 2019 Validation year: 2011 Inpatient or outpatient No minimum follow-up No age restriction	6
≥1 ICD diagnosis code and ≥1 prescription and met criteria for diagnosis based on NLP algorithm	≥1 ICD-9 555.x in inpatient or outpatient setting AND ≥1 electronic prescription for 5-aminosalicylate, corticosteroids, immunomodulator or anti-TNF AND NLP definition met No mention of exclusions Sensitivity 69%-72% Specificity 97% PPV 98% AUC 0.95	**United States** Ananthakrishnan 2013 Liao 2015 Validation years: Not reported Inpatient or outpatient No minimum follow-up No age restriction	75 & 98
≥1 ICD diagnosis code and included in registry	≥1 ICD-10 K50.x in inpatient or outpatient setting AND registered in rare disease database as Crohn's disease (RID V130) No mention of exclusions Sensitivity 94.5%-98.3% Specificity 93.5%	**South Korea** Soh 2019 validated 2010-2013 Park 2018 validated 2010-2013 Kim 2015 validated 2006-2012 Inpatient or outpatient No minimum follow-up No age restriction	10 & 19 & 58
≥1 ICD diagnosis code and included in registry	≥1 ICD-7 572.00, 572.09; ICD-8 563.00; ICD-9 555.x; or ICD-10 K50.x code in inpatient or outpatient specialty visit AND registered at least once in a national IBD registry Mixed CD and UC codes considered IBD-U PPV 90%	**Sweden** Jakobsson 2017 Validation years: 1987-2015 Inpatient or outpatient specialty visit No minimum follow-up No age restriction	28
≥2 diagnosis codes	≥2 ICD-9 555.x codes with ≥1 outpatient If 556.x then exclude PPV 84%	**United States** Hou 2014 Validation years: 1999-2009 Inpatient or outpatient No minimum follow-up Adults (military veterans)	35
≥2 diagnosis codes	≥2 ICD-9-CM 555.x codes with ≥1 colonoscopy on same date as a 555.x code If 556.x then exclude Sensitivity: 49% Specificity: 88% PPV: 81%	**United States** Singla 2020 Validation years: 1996-2012 Inpatient or outpatient 3-y minimum follow-up Adults (active duty military)	1
≥3 diagnosis codes	≥3 ICD-9-CM 555.x codes from a gastroenterologist or surgeon If 556.x or 558.x then exclude Sensitivity: 71% Specificity: 86% PPV: 84%	**United States** Limketkai 2019 Validation years: 1998-2011 Inpatient or outpatient 3-y minimum follow-up Adults (active duty military)	2

Table 1. Continued

Definition	Specific Definition and Diagnostic Accuracy	Characteristics of Validation Study	Reference
≥2 ICD diagnosis codes	≥2 ICD-9 555.x in inpatient or outpatient setting No mention of exclusions Sensitivity 82% PPV 88%	**United States** Liu 2009 Validation years: 1996-2002 Inpatient or outpatient ≥12 mo of enrollment No age restriction	21
≥3 diagnosis codes	≥3 ICD-8 563.x or ICD-10 K50.x in inpatient or outpatient setting No mention of exclusions PPV: 75%	**Denmark** Rye 2021 Validation years: 1998-2015 Inpatient or outpatient No minimum follow-up Age ≥50 y	0
Number of ICD encounters within a time period	≥2 hospitalizations or ≥4 physician claims or ≥2 ambulatory surgery visits for ICD-9 555.x or ICD-10 K50.x within 2 y Used number of CD visits among IBD visits to distinguish CD from UC Sensitivity: 94% Specificity: 99%	**Canada** Rezaie 2012 Validation years: 1997-2007 Inpatient or outpatient No minimum follow-up No age restriction	24
Number of ICD encounters within a time period	≥9 ICD-9 555.x or 556.x or ICD-10 K50.x or K51.x. 5/9 of the most recent visits must be for CD (ICD-9 555.x or ICD-10 K50.x). All 9 codes must occur within a 4-y period. Accuracy 95.6%	**Canada** Benchimol 2014 Validation years: 2001-2006 Inpatient or outpatient No minimum follow-up Adults (≥18 y)	58
Number of ICD encounters within a time period	**≥7 IBD physician visits:** ≥5 of last 7 coded for CD (555.x, K50.x) **<7 IBD physician visits:** All visits for ICD-9 555.x or ICD-10 K50.x All visits within a 3-y period. Sensitivity: 95% Specificity: 86% PPV: 92% NPV: 91%	**Canada** Benchimol 2009 Validation years: 2001-2005 Outpatient No minimum follow-up Pediatric (<18 y)	109
Number of ICD encounters within a time period	**For residents of the province for ≥2 years:** ≥5 outpatient encounters or hospitalizations for ICD-9 555.x or 556.x **For residents of the province for <2 years:** ≥3 outpatient encounters or hospitalizations for ICD-9 555.x or 556.x If a combination of CD and UC codes observed, required at least 9 codes. If majority of 9 most recent codes are 555.x then CD. Sensitivity 89% Specificity 90%	**Canada** Bernstein 1999 Validation years: 1984-1995 Inpatient or outpatient No minimum follow-up No age restriction	147

Abbreviations: AUC, area under the curve; CD, Crohn's disease; IBD, inflammatory bowel disease; IBD-U, inflammatory bowel disease unclassified; ICD-9, International Classification of Diseases–Ninth Revision; ICD-9-CM, International Classification of Diseases–Ninth Revision–Clinical Modification; ICD-10, International Classification of Diseases–Tenth Revision; NLP, natural language processing; NPV, negative predictive value; PPV, positive predictive value; TNF, tumor necrosis factor alpha; UC, ulcerative colitis.

case definition, when available. Alternatively, when there was not a clear preferred case definition, we reported multiple definitions.

Results

Of the 7804 unique studies identified by our search, we identified 29 studies that evaluated a case definition using claims data to define CD patients (Supplementary Appendix Table 2). The most common case definition included counts of diagnosis codes (57%), followed by a combination of diagnosis codes and medications (20%) (Table 1 and

Supplementary Appendix Table 3). All but 1 study validated the case definition with medical chart review. In 2 studies, the patient's primary care provider completed a survey to confirm disease status.[13,14] The International Classification of Diseases (ICD) was the most common case definition, with newer studies using ICD–Tenth Revision (ICD-10) and older studies using ICD–Ninth Revision (ICD-9). Several studies covered extended time periods that used multiple coding systems. They did not report meaningful coding system–specific differences in diagnostic accuracy.

Only 8 countries have validated case definitions for CD (Table 1 and Supplementary Appendix Table 3). These

countries are Canada, Denmark, Italy, Japan, South Korea, Sweden, United Kingdom, and the United States. The criteria for validation varied across the studies we included from these countries depending on that individual country's classification systems. Review of endoscopy, pathology, and radiology reports as well as prescription records and clinic notes were common. Most studies did not report on the credentials of the reviewers performing medical record validation of CD diagnosis. Chart reviewer credentials varied from medical students, to trained chart reviewers (at an insurance company), to gastroenterologists specializing in IBD. Only 2 studies were dedicated to validating a case definition in pediatrics (Supplementary Appendix Table 3).[15,16] One study's study population was restricted to individuals 65 years of age and older,[17] while another was restricted to 50 years of age and older.[18]

PPV was the most frequently reported measure of diagnostic accuracy (n = 15), followed by sensitivity (n = 12) and specificity (n = 8) (Figures 1-3). PPV distribution varied widely, from 18% in a United States health plan[19] to 100% in a validation of patients enrolled in a genetics study.[20] The studies that required ≥1 diagnosis code plus a prescription had the highest PPVs (97%-100%).[20–25] Two studies applied a natural language processing (NLP) algorithm to medical record text in addition to requiring ≥1 diagnosis code plus a prescription; the NLP algorithm did not perform better than the simpler diagnosis code plus prescription algorithm.[23,24] Specificity was also highest with ≥1 diagnosis code plus a prescription (96%-99%).[21–23] Sensitivity remained high in definitions with the ≥1 diagnosis code plus a prescription[21,22] but dropped in definitions that required the addition of the NLP algorithm or colonoscopy.[23,24] Case definitions that required ≥1 diagnosis code varied greatly in diagnostic accuracy. In contrast, the definitions using ≥1 diagnosis code plus prescription or ≥2 diagnosis codes tended to perform consistently across settings, with ≥80% PPV and ≥85% specificity. Sensitivity for more complex definitions were lower (eg, only 49% for ≥2 diagnosis codes plus a colonoscopy).[26] Among the 9 (31% of the total) studies that reported sensitivity, specificity, and PPV, those with ≥1 diagnosis code plus a prescription maximized sensitivity and specificity and had high PPV (Figure 4).

The pediatric definitions were similar to the all-ages definitions in studies coming from the same country. A validation restricted to patients ≥65 years of age found 98% sensitivity for ≥1 ICD-10 diagnosis code in the principal diagnosis position of a hospitalization; no other diagnostic accuracy measures were reported.[17] Another validation study in older patients (ie, ≥50 years of age) found a PPV of only 75% with 3 or more codes in inpatient or outpatient settings.[18]

Sensitivity of Crohn's Disease Case Definitions Utilizing Coding Systems
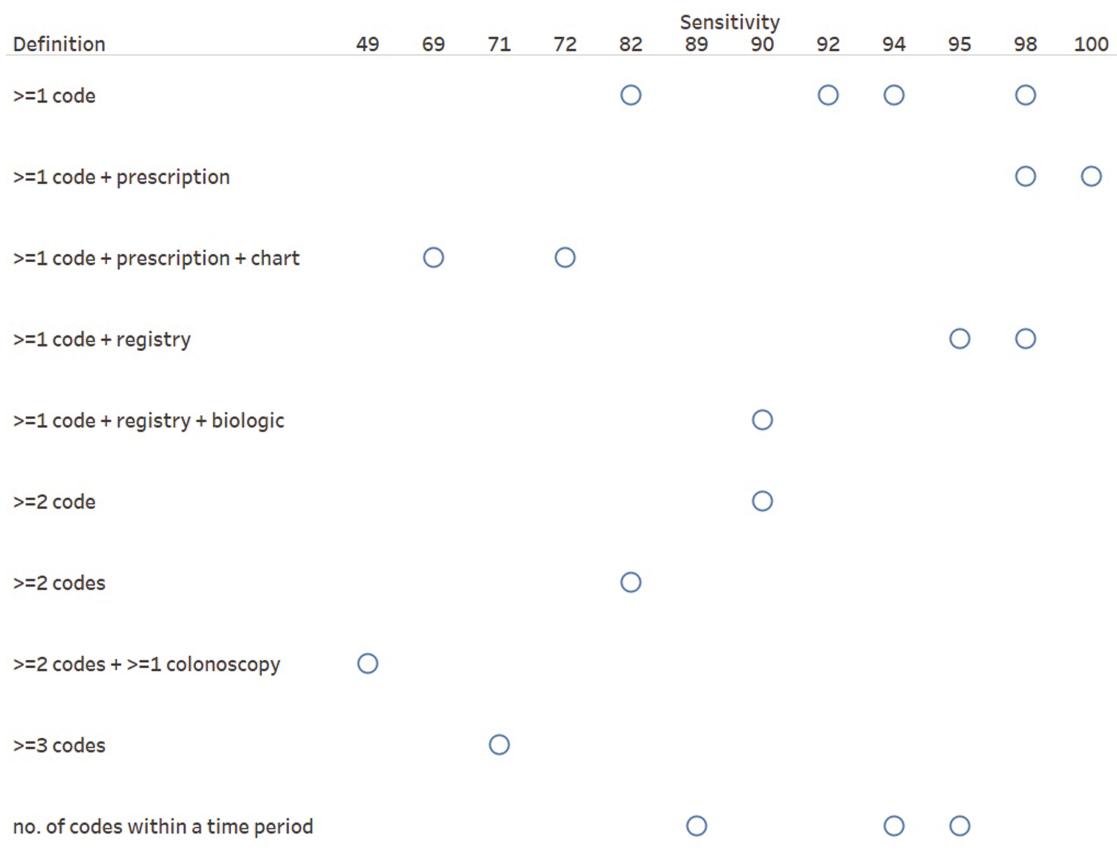
Figure 1. Positive predictive value of the case definitions. *Each circle represents an individual study. If a study reported the diagnostic accuracy of multiple case definitions, the definition with the highest positive predictive value was selected. If diagnostic accuracy was reported separately for multiple groups without an overall summary, each estimate is included in the figure (this applied to only 1 study: Herrinton et al).[19]

Specificity of Crohn's Disease Case Definitions Utilizing Coding Systems

Definition	86	88	90	Specificity 94	96	97	98	99
>=1 code								O
>=1 code + prescription							O	O
>=1 code + prescription + chart						O		
>=1 code + registry				O				
>=1 code + registry + biologic					O			
>=2 code					O			
>=2 codes + >=1 colonoscopy			O					
>=3 codes		O						
no. of codes within a time period	O		O					O
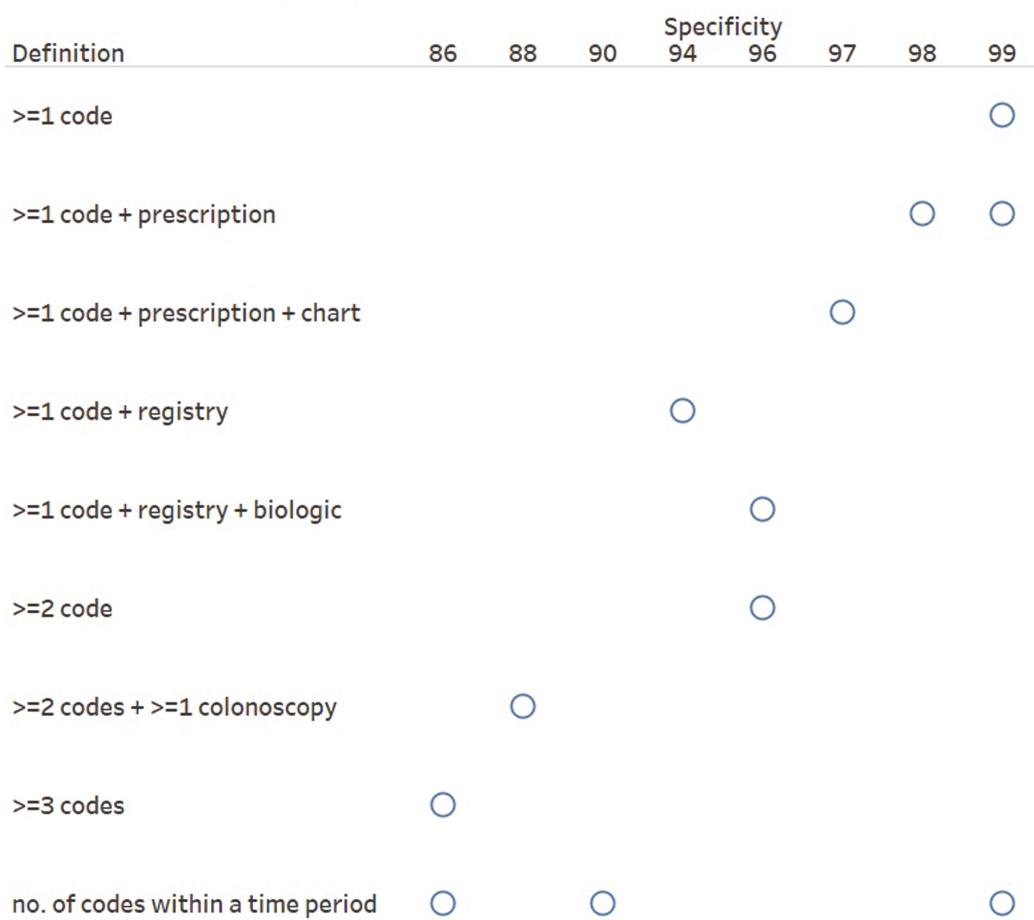
**Figure 2.** Sensitivity of the case definitions.

The diagnostic accuracy of specific ICD codes as markers of disease location or complications was low.[14,27–29] The PPV for ileocolonic disease identified by 2 instances of K50.1 was 64%.[29] However, in another study when the population was restricted to those who had a surgical resection, the diagnostic accuracy for at least 1 instance of K50.1 for ileocolonic disease was high (sensitivity: 95%; specificity: 99%; PPV: 95%; NPV: 99%).[27] The PPV for intestinal fistula was 18%.[28] An algorithm to identify hospitalizations yielded a sensitivity of 52% with a PPV of 48%.[14]

Discussion

Requiring at least 1 encounter with a diagnosis code for CD and a CD-related prescription appeared to produce the best PPV and specificity across the validation studies. Two or more diagnosis codes also tended to have consistent PPV and specificity. However, while some studies found that a case definition with at least 1 CD diagnosis code and no prescription requirement had acceptable diagnostic accuracy, overall the PPV with only 1 diagnostic code was far more variable, including 2 studies that had unacceptable PPVs (18% and 67%) for reliable research. When a case definition cannot be validated in a cohort, a more complex case definition (eg, ≥1

ICD-10 diagnosis plus a prescription or ≥2 ICD-10 diagnosis codes) may maximize specificity. A single encounter definition may be acceptable if health services researchers perform a validation study to determine the algorithm sufficient for their cohort. Case definitions that require colonoscopy or incorporate NLP algorithms of medical records in addition to codes do not increase diagnostic accuracy and instead may decrease it. Different case definitions do not appear necessary for pediatric patients, though they may be helpful for older patients.

Incorporating NLP into the research process is of increasing interest as more tools become available to perform NLP. However, the 2 related studies that used NLP[23,24] did not see improvements in diagnostic accuracy in their comparison of the case definitions with and without NLP. This may reflect the language that was identified from the NLP tool did not add more information than the diagnosis codes. For example, searching for the phrase "Crohn's" may not yield more than the ICD-10 code for Crohn's. Alternatively, as healthcare providers capture different information in notes vs structure fields, the notes searched as part of NLP may not yield useful information on the diagnosis of disease captured by the diagnosis code for the same encounter. When medical records transitioned from paper to electronic records, researchers noticed that the quality and quantity of diagnosis
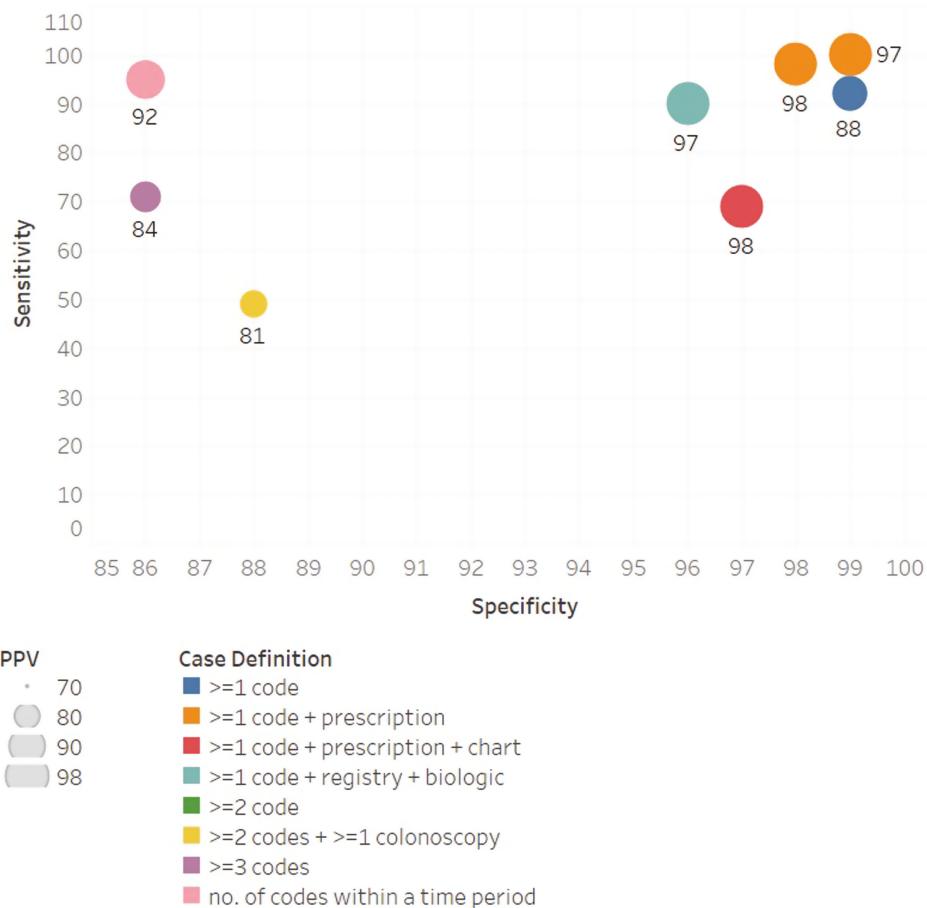
Figure 4. Sensitivity, specificity, and positive predictive value (PPV) among the studies reporting all 3 measures.

and procedure coding was improved in the electronic over the paper record.[30] Similarly, requiring colonoscopy did not lead to an improved case definition despite the need for a colonoscopy to clinically diagnosis CD. This seemingly paradoxical finding may be partially due to databases that tested such a definition having incomplete birth-to-death follow-ups with individuals; they may instead show only portions of patients' lives, which may be years after an initial CD diagnosis. Requiring a colonoscopy may increase diagnostic accuracy for incidence CD; however, none of the included studies aimed to create a definition exclusively for incident patients.

Several studies highly cited as sources for case definitions did not meet our eligibility criteria. These include a study that did not perform any independent confirmation of cases,[31] as well as several studies that validated an IBD definition but did not distinguish a CD-specific case definition.[32–36] Studies calculating population prevalence using multiple nonvalidated case definitions tended to prefer more complex case definitions (eg, ≥1 ICD-10 diagnosis plus a prescription or ≥2 ICD-10 diagnosis codes).[31,35,37–40] These studies that compared rate differences by coding definitions did not meet the inclusion criteria because no validation against an independent source (eg, chart review) was performed. Another study that did not meet the inclusion criteria asked clinicians to code encounter scenarios.[41] Some scenarios were coded inappropriately more often than they were coded appropriately. In particular, CD of the large intestine with an abscess was coded inappropriately 58% of the time.[27] This is consistent with our finding that

codes can generally distinguish CD, but the disease location and complications have worse diagnostic accuracy. Although the accuracy of patient self-report of healthcare utilization in IBD patients is very high,[42] another Canadian study using patient self-report found that IBD patients had the lowest kappa (5%) among the chronic conditions studied.[43] A national, population-based United States survey that asked about IBD status in 2015 estimated that 3 million respondents (1.3% of the U.S. population) have IBD.[44] No follow-up questions were asked to distinguish which patients had CD vs UC or possibly reported irritable bowel syndrome as IBD. Also, no validation was performed, so it is unknown if self-report of IBD status is sufficient to validate a proposed case definition.

This systematic review chose to validate case definitions specifically for CD instead of IBD generally. Ordinarily, health services researchers now report on CD and UC separately because they are thought to be different clinical entities with overlapping, but unique, treatments. Evaluating CD specifically allows us to understand how case definitions for CD patients may differ from those for IBD patients. Studies that compared different CD definitions without validation were also excluded. Although these studies are useful to estimate the bounds of CD prevalence and incidence using different case definitions, they are not as rigorous as a true diagnostic accuracy study wherein trained investigators manually confirm case definitions with individual patient records. However, when a validated case definition cannot be identified in a cohort, sensitivity analyses using multiple case definitions are essential to understanding

Positive Predictive Values of Crohn's Disease Case Definitions Utilizing Coding Systems
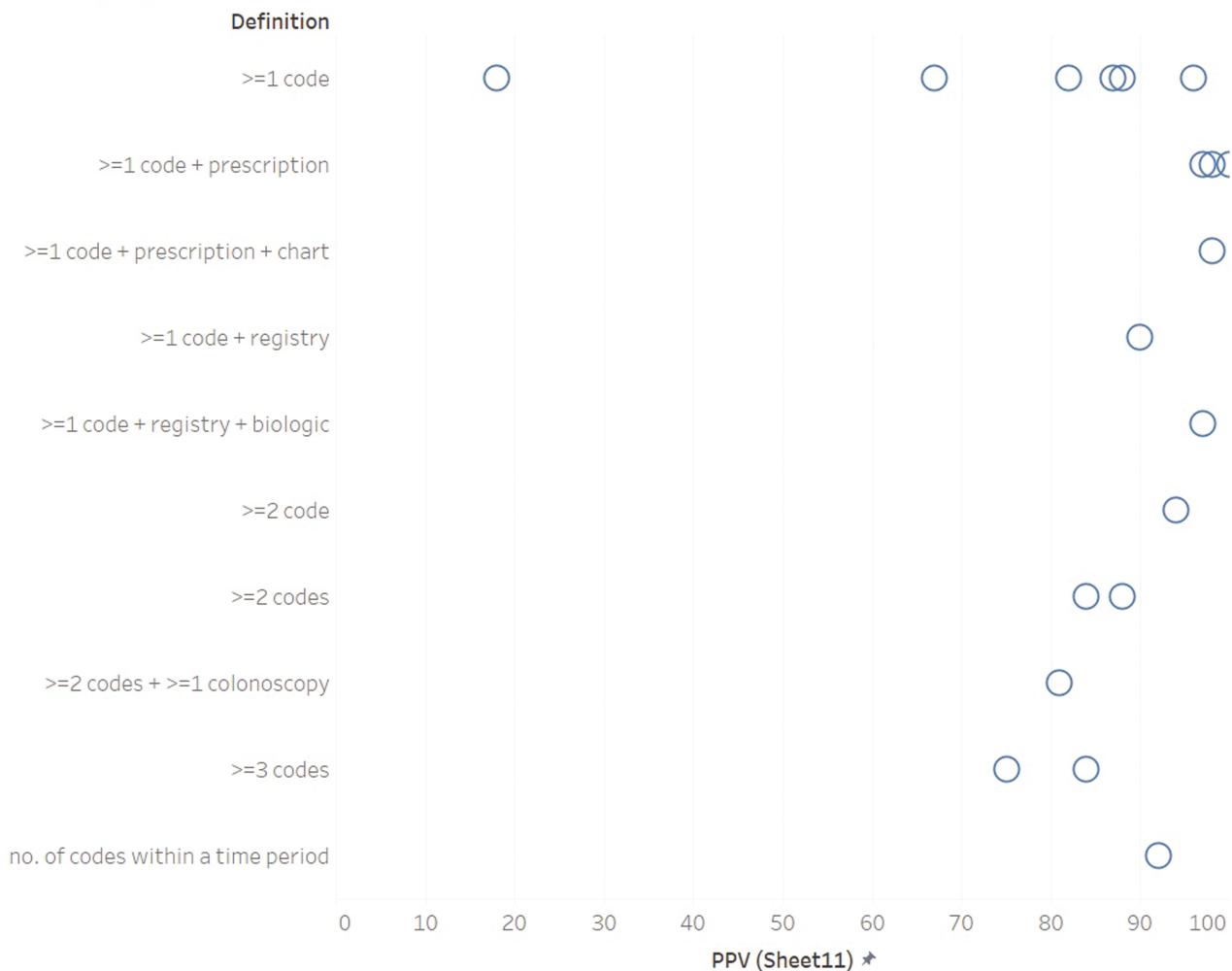
Figure 3. Specificity of the case definitions.

CD rates and outcomes. PPV was the most frequently reported measure of diagnostic accuracy. PPV is susceptible to the underlying population prevalence, and PPV will increase with increasing population prevalence. Another limitation of the approach was the reliance on only 2 databases to identify the literature base.[45] The reference lists of included articles were examined and the search strategy was modified to capture any articles missing by the search strategy or the article was searched for by identifier if its terms were too vague to incorporate into the search. Four older articles were identified independent of the search terms (Supplementary Appendix Table 1). One limitation of the underlying data is that the validation studies were primarily from North America, Europe, and Asia; no validation studies were from South America or Africa. China and India are countries with the largest populations and, thus, likely the largest number of CD patients, yet neither has a validation study that would facilitate the study of large numbers of CD patients from electronic data resources. Another limitation is that only a subset of all studies reported on sensitivity, specificity, and PPV; future validation studies should try to report on all relevant measures.

Conclusions

We recommend that health services researchers validate case definitions in their research cohorts. When such validation cannot be performed, we recommend using a more complex case definition (eg, ≥1 code + prescription or ≥2 codes). Studies without a validated CD case definition should use sensitivity analyses to confirm the robustness of their results.

Acknowledgments

Author Contributions

S.R.B. and S.H. were involved in design. All authors were involved in systematic review screening and data extraction. S.H. and R.A.J. were involved in analysis. S.H. was involved in drafting. All authors were involved revision and approved of manuscript for submission.

Funding

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*. 2016;387(10014):156-167.

2. DeWeerdt S. Genetics: clues in the code. *Nature*. 2016;540(7634):S104-S105.

3. Vermeire S, Van Assche G, Rutgeerts P. Classification of inflammatory bowel disease: the old and the new. *Curr Opin Gastroenterol*. 2012;28(4):321-326.

4. Kaplan GG. Pitfalls and perils of using administrative databases to evaluate the incidence of inflammatory bowel disease overtime. *Inflamm Bowel Dis*. 2014;20(10):1777-1779.

5. McBrien KA, Souri S, Symonds NE, et al. Identification of validated case definitions for medical conditions used in primary care electronic medical record databases: a systematic review. *J Am Med Inform Assoc*. 2018;25(11):1567-1578.

6. Spratt SE, Pereira K, Granger BB, et al. Assessing electronic health record phenotypes against gold-standard diagnostic criteria for diabetes mellitus. *J Am Med Inform Assoc*. 2017;24(e1):e121-e128.

7. Ehrenstein V, Petersen I, Smeeth L, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. *Clin Epidemiol*. 2016;8:49-51.

8. Hripcsak G, Duke JD, Shah NH, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574-578.

9. Rosenbloom ST, Carroll RJ, Warner JL, Matheny ME, Denny JC. Representing knowledge consistently across health systems. *Yearb Med Inform*. 2017;26(1):139-147.

10. U.S. Food and Drug Administration. Guidance for industry and FDA staff: best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data sets. 2013. Accessed February 22, 2022. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/best-practices-conducting-and-reporting-pharmacoepidemiologic-safety-studies-using-electronic.

11. RECORD reporting guidelines. Accessed February 22, 2022. http://record-statement.org

12. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.

13. Bernstein CN, Blanchard JF, Rawsthorne P, Wajda A. Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study. *Am J Epidemiol*. 1999;149(10):916-924.

14. 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(3):211-218.

15. Barton JR, Gillon S, Ferguson A. Incidence of inflammatory bowel disease in Scottish children between 1968 and 1983; marginal fall in ulcerative colitis, three-fold rise in Crohn's disease. *Gut*. 1989;30(5):618-622.

16. Benchimol EI, Guttmann A, Griffiths AM, et al. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. *Gut*. 2009;58(11):1490-1497.

17. Stepaniuk P, Bernstein CN, Nugent Z, Singh H. Characterization of inflammatory bowel disease in elderly hospitalized patients in a large central Canadian health region. *Can J Gastroenterol Hepatol*. 2015;29(5):274-278.

18. Rye C, Rubin KH, Moller FT, Julsgaard M, Jess T, Andersen V. Positive predictive value of diagnostic codes for inflammatory bowel disease in the Danish national patient registry among individuals 50+ years, using patient records as reference standard. *Clin Epidemiol*. 2021;13:335-344.

19. Herrinton LJ, Liu L, Lafata JE, et al. Estimation of the period prevalence of inflammatory bowel disease among nine health plans using computerized diagnoses and outpatient pharmacy dispensings. *Inflamm Bowel Dis*. 2007;13(4):451-461.

20. Restrepo NA, Butkiewicz M, McGrath JA, Crawford DC. Shared genetic etiology of autoimmune diseases in patients from a biorepository linked to de-identified electronic health records. *Front Genet*. 2016;7:185.

21. Morikubo H, Kobayashi T, Fukuda T, Nagahama T, Hisamatsu T, Hibi T. Development of algorithms for identifying patients with Crohn's disease in the Japanese health insurance claims database. *PLoS One*. 2021;16(10):e0258537.

22. 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.

23. Ananthakrishnan AN, Cai T, Savova G, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis*. 2013;19(7):1411-1420.

24. Liao KP, Cai T, Savova GK, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ*. 2015;350:h1885.

25. Shrestha S, Olén O, Eriksson C, et al. The use of ICD codes to identify IBD subtypes and phenotypes of the Montreal classification in the Swedish national patient register. *Scand J Gastroenterol*. 2020;55(4):430-435.

26. Singla M, Hutfless S, Al Kazzi E, Rodriguez B, Betteridge J, Brant S. Clinical codes combined with procedure codes increase diagnostic accuracy of Crohn's disease in a US military health record. *BMJ Open Gastroenterol*. 2020;7(1):e000378. doi:10.1136/bmjgast-2020-000378

27. Ma C, Moran GW, Benchimol EI, et al. Surgical rates for Crohn's disease are decreasing: a population-based time trend analysis and validation study. *Am J Gastroenterol*. 2017;112(12):1840-1848.

28. 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.

29. Liu L, Allison JE, Herrinton LJ. Validity of computerized diagnoses, procedures, and drugs for inflammatory bowel disease in a northern California managed care organization. *Pharmacoepidemiol Drug Saf*. 2009;18(11):1086-1093.

30. Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based with electronic patient records: lessons learned during a study on diagnosis and procedure codes. *J Am Med Inform Assoc*. 2003;10(5):470-477.

31. 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.

32. Sørensen JØ, Nielsen OH, Andersson M, et al. Inflammatory bowel disease with primary sclerosing cholangitis: a Danish population-based cohort study 1977-2011. *Liver Int*. 2018;38(3):532-541.

33. Lee M, Krishnamurthy J, Susi A, et al. Association of autism spectrum disorders and inflammatory bowel disease. *J Autism Dev Disord*. 2018;48(5):1523-1529.

34. Friedman MY, Leventer-Roberts M, Rosenblum J, et al. Development and validation of novel algorithms to identify patients with inflammatory bowel diseases in Israel: an epi-IIRN group study. *Clin Epidemiol*. 2018;10:671-681.

35. Kwak MS, Cha JM, Lee HH, et al. Emerging trends of inflammatory bowel disease in South Korea: a nationwide population-based study. *J Gastroenterol Hepatol*. 2019;34(6):1018-1026.

36. Ekbom A, Helmick C, Zack M, Adami HO. The epidemiology of inflammatory bowel disease: a large, population-based study in Sweden. *Gastroenterology*. 1991;100(2):350-358.

37. Arseneau KO, Stukenborg GJ, Connors AF Jr, Cominelli F. The incidence of lymphoid and myeloid malignancies among hospitalized Crohn's disease patients. *Inflamm Bowel Dis*. 2001;7(2):106-112.

38. Kurti Z, Vegh Z, Golovics PA, et al. Nationwide prevalence and drug treatment practices of inflammatory bowel diseases in Hungary: a population-based study based on the National Health Insurance Fund Database. *Dig Liver Dis*. 2016;48(11):1302-1307.

39. Camacho-Soto A, Gross A, Searles Nielsen S, Dey N, Racette BA. Inflammatory bowel disease and risk of Parkinson's disease in Medicare beneficiaries. *Parkinsonism Relat Disord*. 2018;50:23-28.

40. Holko P, Kawalec P, Stawowczyk E. Prevalence and drug treatment practices of inflammatory bowel diseases in Poland in the years 2012-2014: an analysis of nationwide databases. *Eur J Gastroenterol Hepatol*. 2018;30(4):456-464.

41. Horsky J, Drucker EA, Ramelson HZ. Accuracy and completeness of clinical coding using ICD-10 for ambulatory visits. *AMIA Annu Symp Proc*. 2018;2017:912-920.

42. Longobardi T, Walker JR, Graff LA, Bernstein CN. Health service utilization in IBD: comparison of self-report and administrative data. *BMC Health Serv Res*. 2011;11:137-137.

43. Griffith LE, Gruneir A, Fisher KA, et al. The hidden complexity of measuring number of chronic conditions using administrative and self-report data: a short report. *J Comorb*. 2020;10:2235042X20931287.doi:10.1177/2235042X20931287.

44. Dahlhamer JM, Zammitti EP, Ward BW, Wheaton AG, Croft JB. Prevalence of inflammatory bowel disease among adults aged ≥18 years — United States, 2015. *MMWR Morb Mortal Wkly Rep*. 2016;65:1166-1169.

45. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev*. 2017;6(1):245. doi:10.1186/s13643-017-0644-y