The use of ICD codes to identify IBD subtypes and phenotypes of the Montreal classification in the Swedish National Patient Register

Sarita Shrestha, Ola Olén, Carl Eriksson, Åsa H. Everhov, Pär Myrelius, Isabella Visuri, The SWIBREG Study Group; Jonas F. Ludvigsson, Ida Schoultz, Scott Montgomery, Michael C. Sachs and Jonas Halfvarson

School of Medical Sciences, Örebro University, Örebro, Sweden; Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden; Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; Department of Pediatric Gastroenterology and Nutrition, Sachs’ Children and Youth Hospital, Stockholm, Sweden; Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; Division of Surgery, Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University; Department of Surgery, County Council of Östergötland Linköping, Linköping, Sweden; Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; Department of Pediatrics, Örebro University Hospital, Örebro University, Örebro, Sweden; Division of Epidemiology and Public Health, School of Medicine, University of Nottingham, Nottingham, UK; Department of Medicine, Celiac Disease Center, Columbia University College of Physicians and Surgeons, New York, NY, USA; Clinical Epidemiology and Biostatistics, School of Medical Sciences, Örebro University, Örebro, Sweden; Department of Epidemiology and Public Health, University College London, London, UK

ABSTRACT

Introduction: Whether data on International Classification of Diseases (ICD)-codes from the Swedish National Patient Register (NPR) correctly correspond to subtypes of inflammatory bowel disease (IBD) and phenotypes of the Montreal classification scheme among patients with prevalent disease is unknown.

Materials and methods: We obtained information on IBD subtypes and phenotypes from the medical records of 1403 patients with known IBD who underwent biological treatment at ten Swedish hospitals and retrieved information on their IBD-associated diagnostic codes from the NPR. We used previously described algorithms to define IBD subtypes and phenotypes. Finally, we compared these register-generated subtypes and phenotypes with the corresponding information from the medical records and calculated positive predictive values (PPV) with 95% confidence intervals.

Results: Among patients with clinically confirmed disease and diagnostic listings of IBD in the NPR (N=1401), the PPV was 97 (96–99)% for Crohn’s disease, 98 (97–100)% for ulcerative colitis, and 8 (4–11)% for IBD-unclassified. The overall accuracy for age at diagnosis was 95% (when defined as A1, A2, or A3). Examining the validity of codes representing disease phenotype, the PPV was 36 (32–40)% for colonic Crohn’s disease (L2), 61 (56–65)% for non-stricturing/non-penetrating Crohn’s disease behaviour (B1) and 83 (78–87)% for perianal disease. Correspondingly, the PPV was 80 (71–89)% for proctitis (E1)/left-sided colitis (E2) in ulcerative colitis.

CONTACT
Prof. Jonas F. Halfvarson, jonas.halfvarson@regionorebrolan.se Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, SE 70182, Sweden
The SWIBREG Study Group: Malin Olsson, Henrik Hjortswang, Jonas Bengtsson, Hans Strid, Marie Andersson, Susanna Jägghult, Michael Eberhardsson, Caroline Nordenvall, Jan Bjork, Ulfrika L. Fagerberg, Martin Rejler, Olof Grip, Pontus Karlung, Mattias Block, Eva Angenete, Per M. Hellsström and Anders Gustavsson

Department of Surgery, County Council of Östergötland, Linköping, Sweden; Department of Gastroenterology and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden; Department of Surgery, Sahlgrenska University Hospital/Ostra, Gothenburg, Sweden; Department of Internal Medicine, Södra Alvsborgs Hospital, Borås, Sweden; Stockholm Gastro Center, Karolinska Institutet, Stockholm, Sweden; Department of Medicine, Karolinska Institutet, Stockholm, Sweden; Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; Department of Colorectal Cancer, Karolinska University Hospital, Stockholm, Sweden; Unit of Internal Medicine, Institute Medicine Solna, Karolinska Institutet, Stockholm, Sweden; Patient Area Gastroenterology, Dermatovenerology and Rheumatology, Inflammation and Infection Theme Karolinska University Hospital, Stockholm, Sweden; Center for Clinical Research, Västmanland Hospital, Västerås, Sweden and Uppsala University, Uppsala, Sweden; Department of Pediatrics, Västmanland Hospital, Västerås, Sweden; Department of Women’s and Children’s Health, Karolinska Institutet, Stockholm, Sweden; Department of Medicine, Hölandssjukhuset Eksjö, Region Jönköping County Council, Jönköping, Sweden; Jönköping Academy for Improvement of Health and Welfare, Jönköping University, Jönköping, Sweden; Department of Gastroenterology, Skåne University Hospital, Malmö, Sweden; Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; Department of Surgery, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, SSORG - Scandinavian Surgical Outcomes Research Group, Sahlgrenska University Hospital/Ostra, Gothenburg, Sweden; Department of Surgery, SSORG - Scandinavian Surgical Outcomes Research Group, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; Department of Surgery, Region Västra Götaland, Sahlgrenska University Hospital/Ostra, Gothenburg, Sweden; Department of Medical Sciences, Uppsala University, Uppsala, Sweden; Department of Internal Medicine, Central Hospital, Karlstad, Sweden

Supplemental data for this article can be accessed here.

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction

Inflammatory bowel disease (IBD) is a multifactorial, immune-mediated disease entity, characterised by chronic inflammation in the gastrointestinal tract, which comprises the two predominant forms Crohn’s disease and ulcerative colitis. The clinical features of Crohn’s disease and ulcerative colitis may be heterogeneous, both at diagnosis and during the disease course. For the subset of patients with colonic disease, in whom the diagnosis Crohn’s disease or ulcerative colitis cannot be clearly ascertained, the term IBD unclassified (IBD-U) is used [1].

Because of the heterogeneous nature of the diseases, several attempts have been made to classify Crohn’s disease and ulcerative colitis into subgroups, mainly on the basis of clinical, epidemiological and molecular data. The rational is that an accurate classification may provide guidance on (I) differences in the pathophysiology, (II) differences in disease prognosis, including risk of surgery, and (III) most appropriate therapy for each subtype. The currently accepted model for defining subgroups of the diseases is the Montreal classification [2]. Based on this classification system, the diseases are categorised according to age of diagnosis, anatomic distribution of inflammation, and in Crohn’s disease also disease behaviour.

Scandinavian registers such as the Swedish National Patient Register (NPR) represent an important data source for research. The NPR includes individual-level data on hospital discharges diagnoses, coded according to the International Classification of Diseases and Related Health Problems (ICD) [3]. Using the personal identity number, unique to every Swedish resident, the prospectively recorded information can be linked to other registers and analysed at both the individual and the national level [4,5]. The register is recognised for its high validity. In a previous validation study of incident IBD, the positive predictive value (PPV) was 93% for any IBD, 79% for ulcerative colitis and 72% for Crohn’s disease, using ≥2 diagnostic listings of IBD in the NPR to define confirmed IBD [6].

In register-based studies, the proportion of patients with Crohn’s disease, ulcerative colitis and IBD-U depends on the ICD-code definition of IBD subtypes and when, during a patient’s medical history, data are obtained. In prospective studies, where possible associations between exposures and future outcomes are explored, the definition of IBD subtypes should be based on information that has been collected during a patient’s medical history. Whether data on ICD-codes from the NPR correctly correspond to subgroups of patients with prevalent IBD is unknown.

We therefore aimed to assess the accuracy of ICD-codes for identifying IBD subtypes and phenotypes of the Montreal classification scheme among patients with prevalent IBD, by retrieving information on these codes from the NPR and comparing them with the corresponding information in the patient’s medical records.

Materials and methods

Study population

This retrospective multicentre validation study was conducted at ten Swedish hospitals (Regional hospitals, N = 3 and University hospitals, N = 7). Patients recorded in the Swedish IBD Quality Register (SWIBREG), who had been treated with a biological agent between 1999 and 2017 were eligible for inclusion.

To improve consistency between different centres with respect to granularity of clinical data in the SWIBREG, data on year of birth, sex, date of IBD diagnosis, and clinical characteristics according to the Montreal Classification [2] were abstracted from the medical records, using a standardized, electronic Case Report Form, integrated with the SWIBREG. Diagnosis of IBD was based on internationally accepted clinical, endoscopic, radiologic, and histologic criteria [1,7]. Date of diagnosis was defined as the date of the first examination consistent with IBD.

Data collection was performed by a medical student at each hospital, between February 2016 and December 2017, as part of an initiative to amend data-quality and coverage of the register. To improve agreement between reviewers, every reviewer attended a one-day consensus initiative on how to abstract information from medical records and how to define the diagnosis of IBD and the Montreal classification categories. The study was approved by the Regional Ethics Committee in Stockholm in 2015 (approval numbers: 2007/785-31/5; 2011/1509-32; 2012/601-32; 2015/0004-31; 2015/615-32; 2015/1030-32; 2016/191-31/2).

The Swedish IBD Quality Register (SWIBREG)

In early 2019, the SWIBREG contained information on 46,400 IBD patients [8]. This register has been described in detail elsewhere [8]. In short, the register includes information on diagnosis, disease duration, medication, endoscopy data, surgery, prescribed and administered drugs, extent/location of inflammation, disease behaviour, patient reported outcomes, and quality of life measures [8].
The National Patient Register

The NPR holds information on hospital admissions since 1964, with national coverage since 1987 [9]. From 1997 and onward surgical day care procedures are captured, and since 2001, outpatient visits are reported to the register. Main and contributory diagnoses are recorded according to the ICD-codes and surgical procedures are coded according to an adapted version of the Nordic Medico-Statistical Committee Classification of Surgical Procedures.

Study design

The information abstracted from the medical records and entered into the SWIBREG served as gold standard for the study. In terms of categories of the Montreal classification, data on the highest ever noted degree of disease behaviour, location and extent were abstracted. Patients were excluded from a specific analysis, if the corresponding information could not be identified in the medical records. Information on IBD associated ICD-codes (procedures or diagnostic codes) was retrieved from the NPR and compared with data obtained from the medical record review. As previously described, we used ICD-codes within 5 years preceding the review of a patient’s medical records to assign IBD subtype [5]. In prevalent patients with mixed Crohn’s disease and ulcerative colitis codes, a previously described algorithm was used (Supplementary Tables 2 and 3) [5]. Correspondingly, phenotypes of the Montreal classification were defined by applying a previous defined algorithm (Table 1) [5]. Age at diagnosis of IBD (defined as A1, ≤16 years; A2, 17–40 years or A3 > 40 years) was assigned based on the first IBD-associated ICD-code [2]. Disease location (defined as ileal (L1)/ileo-colonic (L3)/unknown (LX) or colonic location (L2)), disease behaviour (defined as non-stricturing, non-penetrating (B1) or complicated disease behaviour, i.e., stricturing (B2)/penetrating (B3)), perianal disease modifier (p) and disease extent (defined as proctitis (E1)/left-sided colitis (E2) or extensive colitis (E3)) were assigned from the ICD-codes during the five years preceding the review of a patient’s medical records.

Statistics

PPV, sensitivity, and specificity were calculated with 95% confidence interval (CI). We estimated these measures using Microsoft Excel (release/version 16.27, 2019). Overall accuracy of assigned age group at diagnosis, i.e., A1, A2 or A3, was estimated based on the sum of all true positives divided by the total population.

Results

Altogether, the medical records from 1403 patients with IBD (Crohn’s disease, N = 854; ulcerative colitis, N = 519; IBD-U, N = 30) were reviewed. Demographics and clinical characteristics of the study populations are shown in Supplementary Table 1. The median (inter-quartile) time from the first diagnosis of IBD to review of medical records was 12 (7–20) years.

Accuracy of IBD subtypes

Overall, 1401 of the 1403 patients with prevalent IBD were assigned to have IBD, based on information from the NPR. The remaining two patients did not have any IBD-diagnosis in the NPR during the 5-year period prior to medical note review and were excluded from this part of the analysis. Altogether, 784 patients were assigned as Crohn’s disease, 405 as ulcerative colitis and 212 as IBD-U, based on ICD-codes. This resulted in a PPV of 97 (96–99)% for Crohn’s disease, 98 (97–100)% for ulcerative colitis and 8 (4–11)% for IBD-U (Table 2).

Accuracy of age at diagnosis

It was not possible to determine the date of IBD diagnosis from the medical records in three patients. Thus, information on age at diagnosis was available for 1400 patients. Patients were classified as A1 (N = 315), A2 (N = 881) or A3 (N = 204) according to information on age at diagnosis in the medical records. Based on the date of the first diagnostic listing of IBD in the NPR, 292 patients were assigned as A1, 887 as A2.

| Table 1. ICD codes assigned for phenotypes of the Montreal Classification*. |
|---------------------------------|------------------|
| Montreal classification         | Diagnostic codes |
| Crohn’s disease location        |                  |
| Ileal (L1)                      | K50.0            |
| Colonic (L2)                    | K50.1            |
| Ileocolonic or location not defined (L3/LX) | K50.8, K50.9 |
| Crohn’s disease behaviour       |                  |
| Non-stricturing, non-penetrating (B1) | None of the ICD-codes for B2 or B3 |
| Strictureing (B2)               | Crohn’s disease AND any of the following codes (K56.5; K56.6; K56.7; K62.4) |
| Penetrating (B3)                | Crohn’s disease AND any of the following diagnostic codes (K63.0, K63.2, K31.6, N82.2, N82.3, N82.4) OR any of the following surgical procedure codes (JFA76, JFA86) |
| Perianal disease modifier (P)   | Crohn’s disease AND any of the following diagnostic codes: (K60.3, K60.4, K60.5, K61.0, K61.1, K61.2, K61.3, K61.4, K62.4) OR any of the following surgical procedure codes: (JHD20, JHD30, JHD33, JHD50, JHD60, JHD63, JHA00, JHA20, JHW96) |
| Ulcerative colitis extent       |                  |
| Proctitis (E1)                  | K51.2            |
| Left-sided colitis (E2)         | K51.3; K51.5     |
| Extensive colitis (E3)          | K51.0            |
| Extent not defined (EX)         | K51.4; K51.8; K51.9 |

*Everhov et al. [5].
Table 2. Positive predictive value, sensitivity and specificity for IBD subtypes.

| Measures          | Crohn’s disease % (95% CI) | Ulcerative colitis % (95% CI) | IBD-U % (95% CI) |
|-------------------|-----------------------------|-------------------------------|------------------|
| Positive predictive value | 97 (96 – 99)               | 98 (97 – 100)                | 8 (4 – 11)       |
| Sensitivity       | 90 (88 – 92)                | 77 (73 – 80)                 | 55 (37 – 73)     |
| Specificity       | 96 (95 – 98)                | 99 (99 – 100)                | 86 (84 – 88)     |

IBD-U: inflammatory bowel disease-unclassified.

Table 3. Positive predictive value, sensitivity and specificity for age at diagnosis (A1, A2 and A3).

| Measures         | A1, <16 years % (95% CI) | A2, 17–40 years % (95% CI) | A3 > 40 years % (95% CI) |
|------------------|--------------------------|-----------------------------|---------------------------|
| Positive predictive value | 95 (93 – 98)            | 96 (95 – 97)                | 90 (86 – 94)              |
| Sensitivity      | 88 (85 – 92)             | 97 (95 – 98)                | 98 (95 – 100)             |
| Specificity      | 99 (98 – 99)             | 93 (91 – 95)                | 98 (97 – 99)              |

Table 4. Positive predictive value, sensitivity and specificity for disease behaviour, including perianal disease, and disease location among patients with Crohn’s disease.

| Measures                              | Colonic (L2)% (95% CI) | Ileal/ileocolonic/unknown (L1/L3/LX)% (95% CI) | Non-stricturing/non-penetrating (B1)% (95% CI) | Stricturing/penetrating (B2/B3)% (95% CI) | Perianal disease% (95% CI) |
|---------------------------------------|------------------------|-----------------------------------------------|-----------------------------------------------|----------------------------------------|---------------------------|
| Positive Predictive Value             | 36 (32 – 40)          | 81 (76 – 85)                                 | 61 (56 – 65)                                  | 76 (72 – 81)                           | 83 (78 – 87)               |
| Sensitivity                           | 86 (82 – 91)          | 48 (44 – 53)                                 | 75 (71 – 80)                                  | 62 (57 – 67)                           | 81 (76 – 85)               |
| Specificity                           | 48 (44 – 52)          | 78 (73 – 83)                                 | 62 (57 – 67)                                  | 75 (71 – 80)                           | 89 (87 – 92)               |

and 221 as A3. The overall accuracy for age at diagnosis was 95%, when age was defined as A1, A2 or A3. Correspondingly, the PPV was 95 (93–98)% for A1, 96 (95–97)% for A2 and 90 (86–94)% for A3 (Table 3).

Accuracy of phenotypes of Crohn’s disease

Information on disease location could be abstracted from the medical records for 756 patients with Crohn’s disease. The PPV was 36 (32–40)% for colonic disease (L2) and 81 (76–85)% for ileal involvement/unknown location (L1/L3/LX) (Table 4). Correspondingly, information on disease behaviour was available in the medical records for 753 patients. The PPV was 61 (56–65)% for non-stricturing, non-penetrating disease (B1) and 76 (72–81)% for complicated disease behaviour (B2 or B3) (Table 4). The medical record review identified 303 Crohn’s disease cases with perianal disease. Based on the NPR data, 295 of these patients were assigned as perianal disease, corresponding to a PPV of 83 (78–87)% (Table 4).

Accuracy of phenotypes of ulcerative colitis

Information on extent of inflammation was missing or inconsistent in the medical records for 11 of the 405 patients with ulcerative colitis in the NPR. The remaining 394 patients were included in the analysis. This produced a PPV of 80 (71–89)% for proctitis/left-sided colitis (E1/E2) and 82 (78–87)% for extensive colitis (E3) (Table 5). Based on the NPR data, information on extent of disease could not be retrieved for 13 patients.

Discussion

This study aimed to validate ICD-code based algorithms for identification of IBD subtypes as well as phenotypes of the Montreal classification scheme among prevalent IBD cases in the NPR. Our data demonstrate a high accuracy of Crohn’s disease and ulcerative colitis cases in the NPR among patients with prevalent IBD, but a low PPV for IBD-U, as some Crohn’s disease and ulcerative colitis cases were erroneously classified as IBD-U. Likewise, a high PPV was observed for age at diagnosis (defined as A1, A2 or A3), with an overall accuracy of 95%. Among patients with IBD, a high level of accuracy was seen for ICD-codes representing complicated Crohn’s disease behaviour, perianal disease and disease extent (defined as E1/2 or E3), whereas the PPV for colonic Crohn’s disease was 36% only.

Attempts to validate IBD diagnoses in Swedish health registries have previously been reported. Ekbom et al. reported a PPV of 74% for IBD among patients within the Uppsala region with ≥1 IBD-associated ICD-code based on pathology records (both inpatients and outpatients) and the inpatient register between 1965 and 1983 [10]. A more recent validation study evaluated 129 incident patients with at least two diagnoses of Crohn’s disease or ulcerative colitis in the NPR. PPVs of 93% for any IBD, 72% for Crohn’s disease and 79% for ulcerative colitis were observed [6]. The predictive accuracy improved when the analyses were restricted to patients who had ≥1 diagnosis of IBD in the SWIBREG and ≥1 IBD-associated ICD-code in the NPR. Among these patients PPVs of 99 (97–100)% for any IBD, 90 (82–96)% for Crohn’s disease and 96 (89–99)% for ulcerative colitis were observed [6].

In the current study, we validated subtypes of IBD among prevalent patients with known IBD, undergoing treatment with biologics at ten Swedish hospitals. Different IBD...
diagnoses may be recorded during a patient’s medical history as the disease evolves and may shift from one IBD subtype to another subtype. In a prospective follow-up of the Norwegian IBSEN cohort, a change in IBD diagnosis was observed among 9% of the patients who were initially diagnosed with Crohn’s disease or ulcerative colitis [11]. A shift in IBD diagnosis during a patient’s medical history is even more common among patients who are initially diagnosed as IBD-U [12]. Previous studies have reported a shift to either Crohn’s disease or ulcerative colitis among 23–84% of these patients [11,13–18]. We previously reported a shift in IBD diagnosis at some point among 18% of patients with IBD in the NPR [12]. To validate prevalent cases, we included ICD-codes during the 5 years preceding the review of medical records. This algorithm is in line with previous recommendations for the definition of prevalent disease in register-based studies [12]. Using these definitions, we observed PPVs of 97% for prevalent Crohn’s disease and 98% for prevalent ulcerative colitis. The observed level of accuracy for Crohn’s disease and ulcerative colitis is higher than among most previous reports [6,9,19–23]. However, this difference is partly explained by differences in study design, as we included patients with known IBD who were treated with biologics and not a random subset of patients with IBD-associated ICD-codes in a national or regional register.

IBD is a heterogeneous disease entity, and the categorization of patients into Crohn’s disease, ulcerative colitis and IBD-U is in many cases too blunt, since there are pronounced differences with respect to clinical presentations and manifestations even within these subgroups. Because of the clinical heterogeneity, efforts have been made to classify the disease entity further. The Montreal classification system represents the most well-established and currently recognized clinical classification system. Algorithms to define phenotypes of the Montreal classification scheme based on ICD-code data, have previously been used [5,12]. Our data demonstrate that the predictive capacity of these algorithms differs between different phenotypes of the Montreal classification. Age at diagnosis, defined as A1, A2 or A3, was accurately assigned with an overall accuracy of 95%. This finding supports the categorization of patients, based on age at diagnosis, in any study using data from the Swedish NPR. Consistently, high PPVs were observed for perianal Crohn’s disease (83%), extensive colitis (E3, 82%) and ulcerative colitis limited to the rectum or left-sided colon (E1/E2, 80%). Similar to categorization of patients based on age at diagnosis, these data encourage stratification of patients into these subgroups when examining prevalent cases with Crohn’s disease or ulcerative colitis. However, a lower accuracy was observed for some categories of Crohn’s disease location and behaviour. The poor validity for these phenotypes and the difference between the ICD-codes and the Montreal classification, in terms of how disease location is defined, indicate that any results generated from these subgroups of patients need to be interpreted with caution. In fact, our results discourage the use of ICD-codes only for defining disease location and behaviour and indicate that additional sources of data, such as quality registers, should be used to define these phenotypes.

**Strengths and limitations**

A major strength of this study is the large study population (N = 1403) consisting of patients from both regional hospitals and specialized university centres. The acquisition of the patients’ entire medical records allowed us to collect information on diagnosis and disease phenotype based on data recorded later in the follow-up period. To our knowledge, this is the first study to validate algorithms for IBD subtypes and phenotypes according to the Montreal classification by using information from the NPR.

The facts that all Swedish residents can be traced by a unique personal identification number, healthcare is government-financed in Sweden, and that all healthcare providers are obliged to report to the NPR also strengthen the study.

One potential limitation of the study was that the medical records were scrutinised by medical students, who are less experienced than physicians. However, several measures were taken to improve the quality of the review process: (a) all reviewers attended a one-day IBD course, (b) data were collected using a standardised case report form and (c) every reviewer was assigned a supervisor who was specialized in IBD. Although ICD based diagnostic codes were highly indicative in defining IBD subtypes and most phenotypes, the use of diagnostic codes can be subject to errors. This, since codes may not be reported and coding errors can occur when ICD-codes are registered in the NPR. Notably, two of the included patients did not have any diagnosis of IBD in the register during the five years prior to the medical record’s review. The medical record review served as gold standard. However, it cannot be ruled out that some information was overlooked during the review as three patients lacked data necessary to calculate the age at diagnosis. The fact that we identified patients treated with biologics at ten hospitals only and that the study population, therefore, did not represent a random subset of Swedish IBD patients may have overestimated the overall accuracy. This, since patients treated with biologics are monitored more frequently and characterised in more detail, compared to patients with milder disease. In addition, our validation was restricted to previously used algorithms and did not include any refinements of the existing algorithms. The external validity is limited further by the fact that the results may not be applicable to other health care systems in which ICD-codes might be used differently.

**Conclusion**

The Swedish NPR is a reliable data source to define IBD subtypes, age at diagnosis, perianal Crohn’s disease and disease extent among patients with ulcerative colitis, but less accurate for defining Crohn’s disease location and behaviour and also IBD-U. Our findings suggest that ICD-codes can reliably be used to classify IBD subtypes and some phenotypes of the Montreal classification, but additional sources of data are needed to classify Crohn’s disease location and behaviour.
Acknowledgement

We would like to acknowledge the medical students who reviewed all medical records; Anna Fröborg, Emelie Mårdberg, Helena K Olohs, Ida Holm, Johan Weiber, Johanna Engström, Lars Becker, Natalia Mouratidou and Per Olsson.

Disclosure statement

JH has served as speaker and/or advisory board member for AbbVie, Celgene, Celltrion, Ferring, Hospira, Janssen, MEDA, Medivir, MSD, Olink Proteomics, Pfizer, Prometheus Laboratories, Sandoz/Novartis, Shire, Takeda, Thermo Fisher Scientific, Tillotts Pharma, Vifor Pharma, and UCB. He also has received grant support from Janssen, MSD, and Takeda. JFL coordinates a study on behalf of the Swedish IBD quality register (SWIBREG). This study has received funding from Janssens corporation. CE has received grant support/lecture fee/advisory board from Takeda, Janssen, Ferring, and also report a grant from Pfizer in the context of a national safety monitoring program. None of those boards from Janssen, Ferring, and Takeda, and Pfizer regarding topics not related to the present study. ÅHE has worked on projects at Karolinska Institutet partly financed by investigator-initiated grants from Janssen and Ferring, and also report a grant from Pfizer in the context of a national safety monitoring program. None of those studies have any relation to the present study. Karolinska Institutet has received fees for lectures by OO and participation by OO on advisory boards from Janssen, Ferring, Takeda, and Pfizer regarding topics not related to the present study. ÅHE has worked on projects at Karolinska Institutet and SWIBREG partly financed by grants from Ferring and Janssens. IS has received lecture fees from Nutrica and Meda. SM has received financial support for research from AstraZeneca, Roche and Novartis, advisory board: IOVIA, lecture fee(s): Teva. For the remaining authors, there were no conflicts of interests.

Funding

This project (Newbreed) has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754285.

ORCID

Carl Eriksson http://orcid.org/0000-0002-1046-383X
Pär Myrelid http://orcid.org/0000-0001-7518-9213

References

[1] Gomollón F, Dignass A, Annese V, et al. Third European evidence-based consensus on the diagnosis and management of Crohn’s disease 2016: part 1: diagnosis and medical management. J Crohns Colitis. 2017;11(1):3–25.
[2] Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol. 2005; 19(Suppl A):5A–36A.
[3] Ludvigsson JF, Otterblad-Olausson P, Petersson BU, et al. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol. 2009;24(11): 659–667.
[4] Olén O, Askling J, Sachs MC, et al. Childhood onset inflammatory bowel disease and risk of cancer: a Swedish nationwide cohort study 1964–2014. BMJ. 2017;358:j3951.
[5] Everhov AH, Halfvarson J, Myrelid P, et al. Incidence and treatment of patients diagnosed with inflammatory bowel diseases at 60 years or older in Sweden. Gastroenterology. 2018;154(3):518–528 e15.
[6] Jakobsson GL, Sternegard E, Olen O, et al. Validating inflammatory bowel disease (IBD) in the Swedish National Patient Register and the Swedish Quality Register for IBD (SWIBREG). Scand J Gastroenterol. 2017;52(2):216–221.
[7] Magro F, Gionchetti P, Elakim R, et al. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. Part 1: definitions, diagnosis, extra-intestinal manifestations, pregnancy, cancer surveillance, surgery, and ileo-anal pouch disorders. J Crohns Colitis. 2017;11(6):649–670.
[8] Ludvigsson JF, Andersson M, Bengtsson J, et al. Swedish Inflammatory Bowel Disease Register (SWIBREG) – a nationwide quality register. Scand J Gastroenterol. 2019;54(9):1089–1101.
[9] Ludvigsson JF, Andersson E, Ekboom A, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11(1):450.
[10] Ekboom A, Helmick C, Zack M, et al. The epidemiology of inflammatory bowel disease: a large, population-based study in Sweden. Gastroenterology. 1991;100(2):350–358.
[11] Henriksen M, Jahnsen J, Lygren I, et al. Change of diagnosis during the first five years after onset of inflammatory bowel disease: results of a prospective follow-up study (the IBSEN Study). Scand J Gastroenterol. 2006;41(9):1037–1043.
[12] Everhov AH, Sachs MC, Malmborg P, et al. Changes in inflammatory bowel disease subtype during follow-up and over time in 44,302 patients. Scand J Gastroenterol. 2019;54(1):55–63.
[13] Carvalho RS, Abadom V, Dilworth HP, et al. Indeterminate colitis: a significant subgroup of pediatric IBD. Inflamm Bowel Dis. 2006; 12(4):258–262.
[14] Malaty HM, Mehta S, Abraham B, et al. The natural course of inflammatory bowel disease-indeterminate from childhood to adulthood: within a 25 year period. Clin Exp Gastroenterol. 2013; 6:115–121.
[15] Meucci G, Bortoli A, Riccioli FA, et al. Frequency and clinical evolution of indeterminate colitis: a retrospective multi-centre study in northern Italy, GSMII (Gruppo di Studio per le Malattie Infiammatorie Intestinali). Eur J Gastroenterol Hepatol. 1999;11(8):909–913.
[16] Pozler O, Maly J, Bonova O, et al. Incidence of Crohn disease in the Czech Republic in the years 1990 to 2001 and assessment of pediatric population with inflammatory bowel disease. J Pediatr Gastroenterol Nutr. 2006;42(2):186–189.
[17] Rinawi F, Assa A, Elakim R, et al. The natural history of pediatric-onset IBD-unclassified and prediction of Crohn’s disease reclassification: a 27-year study. Scand J Gastroenterol. 2017;52(5):558–563.
[18] Winter DA, Karolewska-Bochenek K, Lazowska-Przeorek I, et al. Pediatric IBD-unclassified is less common than previously reported; results of an 8-year audit of the EUROKIDS registry. Inflamm Bowel Dis. 2015;21(9):2145–2153.
[19] Benchimol EI, Guttmann 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.
[20] Friedman MY, Leventer-Roberts M, Rosenblum J, et al. Development and validation of novel algorithms to identify patients with inflammatory bowel disease in health administrative data from Ontario, Canada. J Clin Epidemiol. 2014;67(8):887–896.
[21] Herrinton LJ, Liu L, Lafata JE, et al. Estimation of the period prevalence of inflammatory bowel disease reclassification in the first five years after onset of inflammatory bowel disease: results of a prospective follow-up study (the IBSEN Study). Scand J Gastroenterol. 2017;52(2):1043–1050.
[22] Hou JK, Tan M, Stidham RW, et al. Accuracy of diagnostic codes for inflammatory bowel disease reclassification: results of an 8-year audit of the EUROKIDS registry. Inflamm Bowel Dis. 2015;21(9):2145–2153.
[23] Herrinton LJ, Liu L, Lafata JE, et al. Estimation of the period prevalence of inflammatory bowel disease among veterans. Dig Dis Sci. 2010;55(9):2592–2598.