Validating inflammatory bowel disease (IBD) in the Swedish National Patient Register and the Swedish Quality Register for IBD (SWIBREG)

Gustav L. Jakobsson, Emil Sternegard, Ola Olen, Pär Myrelid, Rickard Ljung, Hans Strid, Jonas Halfvarson and Jonas F. Ludvigsson

Journal Article

Validating inflammatory bowel disease (IBD) in the Swedish National Patient Register and the Swedish Quality Register for IBD (SWIBREG)

Gustav L. Jakobsson[1]*, Emil Sternegård[1], Ola Olén, Pär Myrelid[5], Rickard Ljung, Hans Strid, Jonas Halfvarson, Jonas F. Ludvigsson[1,2,3, 4]

[1] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, 17177 Sweden

[2] Department of Pediatrics, Örebro University Hospital, Sweden

[3] Division of Epidemiology and Public Health, School of Medicine, University of Nottingham, Clinical Sciences Building 2, City Hospital, Nottingham, UK

[4] Department of Medicine, Columbia University College of Physicians and Surgeons, New York, New York, USA

[5] Division of Surgery, Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, and Department of Surgery, County Council of Östergötland, Linköping, Sweden,

*Corresponding author

Email addresses:

GJ: jakobsson.gu@gmail.com

ES: emil.sternegard@gmail.com

PM: par.myrelid@liu.se

JFL: jonasludvigsson@yahoo.com

Abstract word count: xx

Word count xx.

Key words: population-based, inflammatory bowel disease, validation, national patient register, SWIBREG, – skriv till

Abstract

Background

Both the Swedish National Patient Register (NPR) and the Swedish Quality Register for inflammatory bowel disease (IBD, SWIBREG) are important sources of research data and information. However, the validity of a diagnosis of IBD in these registers is unknown.

Methods

Medical charts of 129 randomly selected patients from the NPR and 165 patients registered both in SWIBREG and the NPR were reviewed in the main analysis. Patients were classified according to standardized criteria for ulcerative colitis (UC), Crohn's disease (CD), or IBD unclassified (IBD-U). Positive predictive values (PPVs) for CD, UC, IBD-U (only SWIBREG) or having any form of IBD were then calculated.

Results

For cases with ≥2 IBD diagnoses, the PPV in the NPR was 93% (87-97) for any IBD, 79% (66-81) for UC and 72% (60-82) for CD. Restricting the analysis to patients with solely registered diagnoses of UC or CD in the NPR, the PPV increased to 90% (77-97) for UC and 81% (67-91) for CD. Combining data from SWIBREG (≥1 record) and the NPR (≥1 record), the PPV was 99% for any IBD (97-100), 96% (89-99) for UC and 90% (82-96) for CD.

Conclusion

The validity of the UC, CD and IBD diagnoses is high in the NPR but even higher when cases were identified through SWIBREG and confirmed in the NPR. These results underline the need for a well-functioning Swedish Quality Register for IBD as a complement to the NPR.

Abbreviations:

CD	Crohn’s disease
IBD	Inflammatory bowel disease
IBD-U	Inflammatory bowel disease unclassified
ICD	International Classification of Diseases
IPR	National Inpatient Register (a part of the NPR)
NPR	National Patient Register
OPR	National Outpatient Register (a part of the NPR)
PIN	Personal identity number
SWIBREG	Swedish Quality Register for IBD
UC	Ulcerative colitis

Background and aims

Inflammatory bowel disease (IBD) is an idiopathic disease that consists of chronic gastrointestinal disorders characterized by gastrointestinal inflammation. IBD usually refers to two disease entities: Crohn's disease (CD) and ulcerative colitis (UC). Clinical differentiation between these two types is often challenging, although some disease-specific histologic, endoscopic and radiologic criteria exist. In approximately 20% of the patients' differentiation between CD and UC is not possible at diagnosis; for these cases, the term IBD unclassified (IBD-U) is recommended.

Trend analyses have shown that the incidence of IBD is increasing worldwide. However, the incidence rate varies considerably, with a recorded incidence of UC and CD ranging from 6.3 and 5.0 per 100 000 person-years in Asia and the Middle East to as high as 19.2 for UC and 20.2 for CD per 100 000 person-years in North America [1]. High incidence rates have also been reported in Northern Europe. In Denmark, the nationwide incidence during the period 2009-2011 for women and men were 23.2/23.4 and 10.3/8.9 per 100 000 person-years for UC and CD, respectively [2]. In 2005-2009, while Sweden (Uppsala region) reported incidence rates of 20 per 100 000 person-years for UC (16.1-23.9) and 9.9 per 100 000 person-years for CD (7.1-12.6) [3, 4]. Correspondingly, a prevalence of 267 per 100 000 person-years for CD (244-291) was observed in another regional Swedish study, but no recent data have been reported for UC [5]. Busch *et al.*, however, estimated the prevalence of IBD to 650 per 100 000 person-years in Sweden based on data retrieved from the National Patient Register (NPR) when using two or more diagnoses of IBD in non-primary care as diagnostic criteria[6].

The NPR was established in 1964 in some regions of Sweden in order to collect information on inpatient care; nationwide coverage was achieved in 1987. In 2001, data on specialized hospital-based outpatient care were added and today the coverage of the NPR is virtually 100% [7]. Data registered in the NPR include main diagnosis, secondary diagnoses, external cause of injury and poisoning, procedure codes and personal identity number (PIN) to allow for linkage with other Swedish registers[8]. The diagnoses registered in the NPR are coded according to the Swedish version of the International Statistical Classification of Diseases (ICD) and Related Health Problems. Different revisions have been used during different periods: ICD-7 in 1964-1967, ICD-8 in 1968-1986, ICD-9 in 1987-1996 and ICD-10 from 1997 onwards. For inpatient diagnoses in general, the positive predictive value (PPV), i.e. the proportion of registered diagnoses that is actually correct, is most often between 85 and 95%[7]. While Ekbom *et al.* examined the validity of 'possible IBD' in a regional NPR dataset from 1965-1983, the authors did not examine individual IBD diagnoses (UC and CD).

The Swedish Quality Register for IBD (SWIBREG), established in 2005, contains clinical data that are either missing in the NPR or lacking in detail. Such variables include disease duration, disease extent, surgery (described in detail), disease characteristics, endoscopy data, prescribed and administered drugs as well as their side-effects, disease activity and laboratory test results. SWIBREG has a national coverage rate of about 50%, with some 33,000 registered patients as of December 2015 [9, 10]. The IBD diagnoses in SWIBREG have never been validated.

The present study therefore aimed to validate the diagnoses of UC, CD and IBD in the NPR and UC, CD, IBD and IBD-U in SWIBREG.

Materials and Methods

Study population

All patients with at least one diagnosis of either CD or UC registered in the NPR after 1987 were identified through their ICD coding (see supplementary Table S1 for used ICD codes). Of these patients, 200 were randomly selected for this study, including patients from all types of hospital (from regional to university hospitals). Another 200 patients were randomly selected from SWIBREG. For each patient, information about the date of the first IBD diagnosis and hospital in which the diagnosis was made was obtained from the National Board of Health and Welfare. We contacted relevant hospitals and requested medical records that included physician notes, surgery notes, discharge notes, laboratory results, radiology/biopsy/endoscopy referrals and other written referrals for at least 2 years after the first registered diagnosis in the NPR (which could be before 1987). Hospitals failing to provide the medical records within 2 months were sent a reminder and contacted by phone. Figure 1 displays an overview of the selection process of patients eligible for the study.

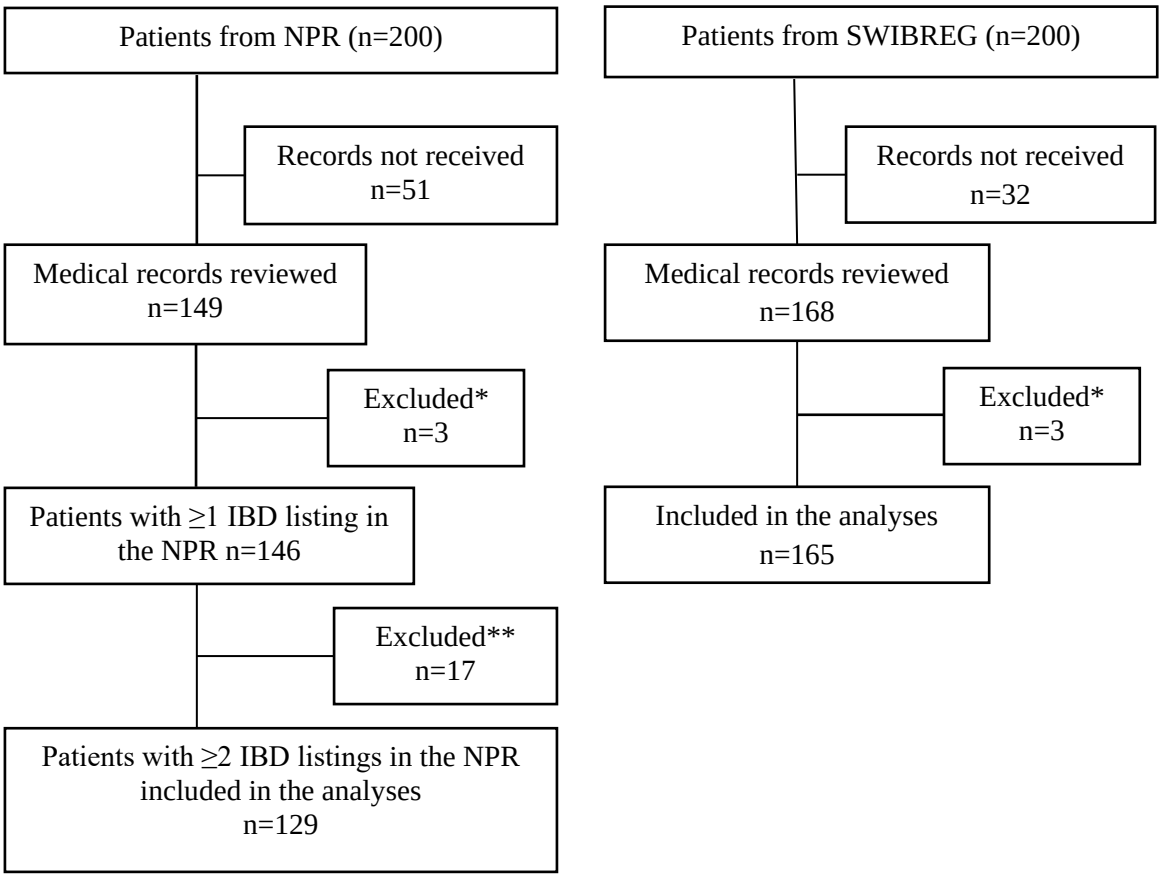

Figure 1. Flow chart of the selection process of patients eligible for the study.
Abbreviations: NPR=National Patient Register; NBHW=National Board of Health and Welfare; IBD=inflammatory bowel disease.
*Excluded because of incomplete medical records. ** Excluded because of only 1 recorded diagnosis of any IBD in the NPR.

For the validation of NPR, we excluded 17 individuals with only one registered IBD: unspecified colitis or proctitis, n=5 (not explicitly assessed as UC or CD by the treating physician); suspected IBD, which was later dismissed, n=2; diverticulosis, n=2; suspected terminal ileitis on computerized tomography (CT), n=2; polyposis, n=1; self-reported IBD in the 1940s with no later symptoms, n=1; anal fissure with no other IBD symptoms, n=1; and resection of terminal ileum believed to be caused by CD, n=1. In addition, two patients were probably misclassified as having CD (ICD code K50.9) but should instead have been registered with C50.9 (breast cancer) and R50.9 (‘unspecified fever’).

Data Elements

Data from the medical charts were abstracted using a standardized form. We retrieved data from pathology, endoscopy and radiology reports, as well as on symptoms and medication at the time of the first registered diagnosis in the NPR. If no information were explicitly recorded for a specific symptom (e.g., abdominal pain) or examination (e.g., endoscopy), the patient was classified as negative for this specific symptom/examination (i.e. 'no abdominal pain' and 'no endoscopy performed'). This method of classifying if a case had undergone an examination or suffered from symptom is likely to underestimate the prevalence of symptoms and procedures but this was considered when setting up the diagnostic criteria by using a probable criteria. From the Swedish National Prescribed Drug Register, additional information about patient dispensation of drugs was retrieved from 2006 and onwards [11]. The anatomical extent/location of inflammation was assessed using data from endoscopies, barium enemas, CT and magnetic resonance imaging (MRI). All this information was then classified according to the Montreal classification [12].

Case definition

Cases were classified as *definite*, *probable* or *negative* for UC, CD or IBD-U. Patients fulfilling the Copenhagen criteria were classified as having a *definite* diagnosis [13, 14]; patients failing to fulfill the Copenhagen criteria but judged by the treating physician as having IBD with at least two registered IBD diagnoses in the NPR were given a *probable* diagnosis in line with the treating physician's assessment (UC, CD or IBD-U). If none of these criteria were fulfilled, the patient was classified as *negative* for IBD. If the medical records were insufficient, i.e. contained too little data to allow assessment, the patient was excluded from the study. The full criteria are presented in Supplementary Table S2.

Ethical approval

This project was approved by the Stockholm Ethics Review Board (2014/1288-31/4) [Ludvigsson, 2015, PMID 26648756].

Statistics

Data management was performed using Microsoft Excel and SPSS software 22.0. PPVs were presented with 95% CIs calculated using the Clopper-Pearson method [15].

Results

In total, 149/200 requested medical records were retrieved to validate the NPR. Three of these charts were excluded because of incompleteness (Figure 1). The main analyses of the NPR included 129 cases (88.4%) with ≥2 recorded IBD diagnoses in the NPR. For the SWIBREG validation, medical records from 168 cases were retrieved, three of which were excluded for incompleteness. The SWIBREG validation was thus based on 165 medical records (Figure 1). Patients in SWIBREG tended to be diagnosed more recently than the average patient in the NPR, which is because registration in SWIBREG started in 2005. This fact could explain the higher retrieval rate for patient charts related to SWIBREG (83.5%) than for the patient charts identified through the NPR (74.5%).

Table 1 presents patient characteristics in relation to diagnosis (UC, CD, IBD-U) (n=129). From the NPR, 120 patients with IBD had either a definite or probable IBD diagnosis (UC n=57, 44%; CD n=53, 41%; IBD-U n=10, 8%). Of the nine patients negative for IBD, six had initially been diagnosed with IBD, but the diagnosis was later dismissed (ischemic colitis n=1, functional gastrointestinal disease n=2, unclassified disease n=3). The remaining three non-IBD cases consisted of one case of miscoding (the patient had heart failure, which has the

ICD-10 code I50), one with *infectious* proctitis and one with infectious enteritis based on positive stool culture.

Of the cases validated from SWIBREG, 81 (49%) had UC, 79 (48%) CD and 4 (2%) IBD-U. One patient did not fulfill the diagnostic criteria for IBD. This patient had a negative histopathology report and an ileocolonoscopy without any macroscopic signs of inflammation.

For patients in the NPR who met the criteria for UC, diarrhea (79%) and bloody stool (70%) were the most frequent symptoms (Table 1). These symptoms were also most frequent in UC in SWIBREG: diarrhea 83% and bloody stool 74% (Table 1). Among patients in the NPR who met the criteria for CD, abdominal pain and diarrhea were the most prevalent symptoms; the same pattern was evident for patients with CD in SWIBREG (Table 1). Fistulas had been reported in 13.2 and 12.7% of the cases classified as CD in the NPR and SWIBREG cohort, respectively. Table 1 also presents information on received endoscopy referrals as well as the percentage of individuals with radiological signs of stenosis. If the medical chart contained no information on endoscopy (MRI/CT) or any histopathology report, the patient was classified as not having undergone the procedures.

Table 1. Diagnosis based on set criteria and symptoms and findings

	National Patient Register (N=120)										SWIBREG (N=164)									
Diagnosis based on criteria*	**Any IBD****		**UC**		**CD**		**IBD-U**		**Not IBD**		**Any IBD****		**UC**		**CD**		**IBD-U**		**Not IBD**	
Number	120	100%	57	44%	53	41%	10	8%	9	7%	164	100%	81	49%	79	48%	4	2%	1	1%
Examinations & findings																				
Colonoscopy	79	66%	41	72%	33	62%	5	50%	3	33%	109	66%	51	63%	54	68%	4	100%	1	100%
Reached ileum	30	38%	12	29%	15	45%	3	60%	1	33%	47	43%	16	31%	30	56%	1	25%	1	100%
Gastroscopy	19	16%	6	11%	13	25%	0	0%	0	0%	30	18%	8	10%	22	28%	0	0%	0	0%
MRI/CT of small intestines	18	15%	5	9%	12	23%	1	10%	2	22%	27	16%	5	6%	21	27%	1	25%	0	0%
Sign of stenosis	*9*	*50%*	*0*	*0%*	*9*	*75%*	*0*	*0%*	*0*	*0%*	*14*	*52%*	*0*	*0%*	*14*	*67%*	*0*	*0%*	*0*	*0%*
No sign of stenosis	*9*	*50%*	*5*	*100%*	*3*	*25%*	*1*	*100%*	*2*	*100%*	*13*	*48%*	*5*	*100%*	*7*	*33%*	*1*	*100%*	*0*	*0%*
Small bowel follow-through	31	26%	7	12%	20	38%	4	40%	1	11%	50	30%	15	19%	34	43%	1	25%	0	0%
Sign of stenosis	*12*	*39%*	*0*	*0%*	*12*	*60%*	*0*	*0%*	*0*	*0%*	*20*	*40%*	*0*	*0%*	*20*	*59%*	*0*	*0%*	*0*	*0%*
No sign of stenosis	*19*	*61%*	*7*	*100%*	*8*	*40%*	*4*	*100%*	*1*	*100%*	*30*	*60%*	*15*	*100%*	*14*	*41%*	*1*	*100%*	*0*	*0%*
Symptoms																				
Hematochezia	60	50%	40	70%	13	25%	7	70%	1	11%	91	55%	60	74%	28	35%	3	75%	0	0%
Diarrhea	88	73%	45	79%	36	68%	7	70%	4	44%	121	74%	67	83%	51	65%	3	75%	0	0%
Weight loss	34	28%	16	28%	18	34%	0	0%	1	11%	47	29%	19	23%	27	34%	1	25%	0	0%
Abdominal pain	60	50%	19	33%	37	70%	4	40%	4	44%	86	52%	29	36%	56	71%	1	25%	1	100%
Fatigue	10	8%	4	7%	6	11%	0	0%	0	0%	17	10%	6	7%	10	13%	1	25%	0	0%

Abbreviations: CD= Crohn's disease, CT= Computed tomography, IBD = inflammatory bowel disease, IBD-U = inflammatory bowel disease unclassified MRI=Magnetic resonance imaging, UC=Ulcerative colitis.

*See Supplementary Table S2 for criteria. ** UC, CD or IBD-U

The extent of inflammation was similar in UC patients from the NPR and SWIBREG. In total, 11% of the NPR cases fulfilling UC criteria had ulcerative proctitis (E1), 30% left-sided colitis (E2) and 44% extensive colitis (E3). Patient chart data did not allow us to classify the extent of inflammation in the remaining 16% (Table 2). Corresponding figures for the SWIBREG cohort were 10%, 26%, 48%, respectively (impossible to define extent: 16% also here). Similarly, of the CD patients in the NPR, 40% had ileal disease (L1), 23% colonic disease (L2) and 15% ileocolonic disease (L3) (no data on location in 23%). Coexisting upper gastrointestinal (GI) disease (L4) was present in 0%, 25% and 13% of the cases with ileal, colonic and ileocolonic disease, respectively (Table 2). The corresponding figures for SWIBREG cases were 37%, 26% and 24%, respectively (no data on location in 23%). Coexisting upper GI disease (L4) was present in 7%, 10% and 0% of the cases, respectively.

Table 2. Extent of inflammation based on the Montreal classification

	NPR (N=129)		SWIBREG (N=165)	
Ulcerative colitis				
E1 - Proctitis	6	11%	8	10%
E2 – Left-sided colitis	17	30%	21	26%
E3 - Extensive colitis	25	44%	39	48%
Unknown	9	16%	13	16%
Total:	57	100%	81	100%
Crohn's disease				
L1 -Ileal disease	21	40%	27	34%
L1 + L4	0	0%	2	3%
L2 – Colonic disease	9	17%	19	23%
L2 + L4	3	6%	2	3%
L3 – Ileocolonic disease	7	13%	19	24%
L3+L4	1	2%	0	0%
L4 - Upper GI disease	0	0%	0	0%
Unknown	12	23%	10	14%
Total:	53	100%	79	100%

GI, Gastrointestinal. NPR, National patient register.

Overall, 120/129 (93%) of the patients with ≥2 IBD diagnoses in the NPR and 164/165 (99%) of those with a SWIBREG record of IBD and ≥1 IBD diagnosis in the NPR met our *a priori* criteria for any IBD. This resulted in a PPV of 79% for UC (95% CI: 66-88), 72% for CD (95% CI: 60-82) and 93% for any-IBD (95% CI: 87-97) (Table 3). For SWIBREG, the corresponding figures were 96% (95% CI: 89-99) for UC, 90% for CD (95% CI: 82-96) and 99% for any IBD (95% CI: 97-100) (Table 3). In- or outpatient status at first registered diagnosis in the NPR did not substantially change the PPV for the respective diagnosis (Table 3). Limiting the analysis to cases from NPR with a solely registered diagnosis of UC (n=41) or CD (n=48) resulted in increased PPVs of 90% (77-97) for UC and 81% (67-91) for CD. For an overview of the PPVs, see Table 3.

Table 3. Positive predictive values (PPVs) by register

	UC (95% CI)	CD (95% CI)	IBD-U (95% CI)	Any IBD (95% CI)
PPV NPR:	**79%** (66- 88)	**72%** (60- 82)	**NA**	**93%** (87- 97)
*PPV IPC**	**80%** (66- 89)	**71%** (57- 82)	**NA**	**95%** (89- 97)
*PPV OPC**	**71%** (29- 96)	**80%** (44- 97)	**NA**	**82%** (57- 96)
PPV NPR:**	**90%** (77- 97)	**81%** (67- 91)	**NA**	**92%** (84- 97)
PPV SWIBREG:	**96%** (89- 99)	**90%** (82- 96)	**33%** (1- 91)	**99%** (97- 100)

Abbreviations: CD= Crohn's disease, IBD = inflammatory bowel disease, IBD-U = inflammatory bowel disease unclassified, IPC= Inpatient care, OPC=Outpatient care, UC=Ulcerative colitis.
*First diagnosis registered at out- or inpatient care. **Patients exclusively diagnosed with ulcerative colitis (UC, n=41) or Crohn's disease (CD, n=48) (i.e. no other inflammatory bowel diagnosis) in the National Patient Register.

For cases registered as UC or CD in the NPR, only 8 and 6%, respectively, were classified as *negative for IBD*; instead, most misclassifications were due to individuals first being diagnosed with UC or CD but later declared to have another type of IBD (see Table 4 for data on re-classifications). Corresponding figures for SWIBREG were 0%, 1% and 0% for UC, CD and IBD-U, respectively.

Table 4. Diagnosis recorded in the National Patient Register/SWIBREG with corresponding diagnoses according to criteria.

National Patient Register (n=129)										
		Diagnosis fulfilled according to criteria*:								
		UC		**CD**		**IBD-U**		**Not IBD**		**Total:**
Diagnosis in the NPR:	**UC**	48	79%	4	7%	4	7%	5	8%	**61**
	CD	9	13%	49	72%	6	9%	4	6%	**68**
Stratification whether first NPR diagnosis was recorded at in-(n=112) or outpatient (n=17) care.										
Inpatient care:	**UC**	43	80%	4	7%	4	7%	3	6%	**54**
	CD	9	16%	41	71%	5	9%	3	5%	**58**
Outpatient care**:	**UC**	5	71%	0	0%	0	0%	2	29%	**7**
	CD	0	0%	8	80%	1	10%	1	10%	**10**
Analysis of patients exclusively diagnosed with UC (N=41) or CD (N=48) in the NPR										
Diagnosis in the NPR:	**UC**	37	90%	0	0%	1	2%	3	7%	**41**
	CD	1	2%	39	81%	4	8%	4	8%	**48**
SWIBREG (n=165)										
		Diagnosis according to criteria*:								
		UC		**CD**		**IBD-U**		**Not IBD**		**Total:**
Diagnosis in SWIBREG:	**UC**	75	96%	3	4%	0	0%	0	0%	**78**
	CD	4	5%	76	90%	3	4%	1	1%	**84**
	IBD-U	2	67%	0	0%	1	33%	0	0%	**3**

*See Supplementary Table S2 for criteria. **Outpatient visits are registered only from 2001 and onwards

37

38

39 **Sub-analysis**

40 When including all cases (N=146) randomly selected from the NPR (≥1 registered diagnosis), 75%

41 (95% CI: 64-85) met the UC criteria and 68% (95% CI: 56-78) the CD criteria (see Supplementary

42 Table S3).

43

44 Restricting the analyses of SWIBREG to cases with congruent diagnoses in SWIBREG and the first

45 diagnosis in NPR (N=143) showed that 99% (95% CI: 92-100) and 95% (95% CI: 87-98) met the

46 criteria for UC and CD, respectively (see Supplementary Table S4).

47

Discussion

This study aimed to validate IBD, and specifically, UC and CD in the NPR and SWIBREG. Our main finding is the high validity of the UC and CD diagnoses in the NPR, especially for any IBD, as most misclassified cases of UC and CD were made up of the other IBD diagnosis or IBD-U. For ≥2 IBD diagnoses in the NPR, the PPV for any IBD was 93%. The validity was even higher for patients identified through SWIBREG and confirmed by ≥1 NPR diagnosis: 99% for IBD, 96% for UC and 90% for CD.

The results of our validation study should be compared with those of Ekbom *et al*. in 1991, who studied IBD registered diagnoses in the NPR from Uppsala County, Sweden in 1965-1983. That study reported a PPV of 74% for IBD. However, for several reasons, we expected the PPV for IBD in our study to be higher. First, we included cases with at least two NPR diagnoses of IBD, whereas Ekbom *et al.* included patients with at least one IBD diagnosis from the NPR, but also included patients with *possible* IBD (e.g., patients with a diverticulitis diagnosis in the NPR, as such patients may represent misclassified IBD). Considering that our study used more stringent inclusion criteria, a higher PPV would be expected. Restricting the analysis to cases with a UC diagnosis (and never a CD diagnosis) or a CD diagnosis (and never a UC diagnosis) in the NPR, the PPVs increased to 90% for UC and 81% for CD. It seems that out- or inpatient status at first diagnosis does not affect PPV more than marginally (Table 3). However, because of the small number of cases that had their first diagnoses recorded as outpatients (n=17), no firm conclusions can be drawn as to the importance of out- or inpatient status at first IBD diagnosis.

The PPV for IBD in our study is in line with the 85-95% reported for other diagnoses in the NPR, and with the highest PPVs generally seen for severe diseases [7]. Earlier validation has

reported PPVs close to 90% for other inflammatory diseases (e.g., rheumatoid arthritis, 90% and ankylosing spondylitis, 89%)[16, 17]).

Individual IBD diagnoses (CD and UC) in SWIBREG have not previously been validated. SWIBREG diagnoses showed slightly higher PPVs than those of the NPR. This finding was expected because the SWIBREG record could be seen as a confirmation of the NPR diagnosis and because SWIBREG diagnoses are generally assigned by specialists or residents in gastroenterology or internal medicine with a special interest in IBD. Furthermore, the diagnoses from SWIBREG were more recent (given that this quality registry started in 2005) with better access to patient chart data. When a more stringent criterion was applied to identify cases from SWIBREG (only those with a congruent first diagnosis in NPR and diagnosis in SWIBREG), specificity increased as predicted for both UC (PPV=99%, 95% CI: 92-100) and CD (PPV=95%, 95% CI; 87-98) (Supplementary Table S4).

Our study shows a higher PPV for diagnoses of UC or CD registered in SWIBREG and confirmed by ≥1 NPR record than for cases with ≥2 records of IBD in the NPR. Several explanations may account for this finding. The diagnoses in the NPR are recorded at the first visit with suspected IBD: sometimes the initial diagnosis is incorrect (e.g., initially classified as UC but later re-classified as CD), which lowers the calculated PPV. This argument is supported by the substantially higher PPV for any IBD than for UC or CD separately (Table 3). In contrast, subjects from SWIBREG are not included in this register (SWIBREG) until the treating physician is certain about the diagnosis (IBD or not) and which type of IBD is present (UC, CD or IBD-U). Moreover, as mentioned above cases are exclusively diagnosed by a specialist or a resident in internal medicine or gastroenterology. Taken together, these factors should lead to a higher PPV for diagnoses in SWIBREG than in the NPR.

This study also examined symptoms and signs in IBD. In patients with UC diarrhea (79/83%: NPR/SWIBREG cohort) and hematochezia (70/74%) were the most common symptoms. Abdominal pain (70/71%) and diarrhea (68/65%) were the most common symptoms in CD. These data are consistent with international data that have shown that hematochezia and diarrhea are the most common symptoms in UC[18]. For CD, diarrhea is the most common symptom and abdominal pain is present in about 70% of the cases before diagnosis [19].

Finally, disease extent and location of disease were examined. In UC, the inflammation was mostly an extensive colitis (44/48%: NPR/SWIBREG cohort) followed by left-sided colitis (30/26%). In CD, inflammation mostly involved the terminal ileum (53/58%). International data have shown that the inflammation for UC at time of diagnosis is located most often in the rectum or sigmoid colon (30-50%), followed by left-sided-colitis (20-30%) and pancolitis (20%) [20]. The more extensive inflammation for UC in our material than in the literature could be explained by progression of inflammation since our data is from the time of first registration in the NPR, which in some cases could be after the time of initial diagnosis.

Our study has several strengths. First, the random selection from the NPR and SWIBREG and the nationwide approach support the contention that the results are likely representative of expected results nationally. Second, we requested charts from a broad timespan (at least 2 years), a strategy that increases the possibility of including relevant diagnostic examinations and medication.

Some limitations should be acknowledged. First, we were not always able to retrieve complete medical records (e.g., we identified histopathology referrals in about 75% of the

patients and colonoscopy referrals in 65%, although most, if not all, the patients are likely to have undergone these procedures). To address the possible lack of data we used a *probable* criterion for IBD, which was based on the assessment of the treating physicians (who likely had read the referrals) after their investigation. Second, we did not retrieve all of the requested charts (51/200 for the NPR and 32/200 for SWIBREG were missing). We cannot exclude that missing charts have biased our results somewhat. We tried to minimize the risk of bias by re-requesting charts first by letter and then by telephone to those hospitals and clinics that had not responded. Third, we did not include patients with other forms of IBD, such as collagenous and lymphocytic colitis, but these constitute only a small percentage of the total number of IBD cases.

Conclusions

In conclusion, the validity of a diagnosis of UC, CD and IBD is high in the NPR but even higher when cases were identified through SWIBREG and confirmed in the NPR. These results strengthen the need for a well-functioning Swedish Quality Register for IBD as a complement to the NPR.

Competing interests

None.

Authors' contributions

Study concept: JFL.

Acquisition of data: GLJ, ES, JFL.

Statistical analyses: GLJ.

First draft of the manuscript: GLJ, JFL.

Study design, critical revision of the manuscript for important intellectual content: GLJ, ES,

JFL, OO, PM, RL, HS, JH.

Study supervision: JFL.

Acknowledgements

This project was supported by a grant from SWIBREG.

References:

1. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, Benchimol EI, Panaccione R, Ghosh S, Barkema HW *et al*: **Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review**. *Gastroenterology* 2012, **142**(1):46-54 e42; quiz e30.
2. Norgard BM, Nielsen J, Fonager K, Kjeldsen J, Jacobsen BA, Qvist N: **The incidence of ulcerative colitis (1995-2011) and Crohn's disease (1995-2012) - based on nationwide Danish registry data**. *J Crohns Colitis* 2014, **8**(10):1274-1280.
3. Sjoberg D, Holmstrom T, Larsson M, Nielsen AL, Holmquist L, Ekbom A, Ronnblom A: **Incidence and natural history of ulcerative colitis in the Uppsala Region of Sweden 2005-2009 - results from the IBD cohort of the Uppsala Region (ICURE)**. *J Crohns Colitis* 2013, **7**(9):e351-357.
4. Sjoberg D, Holmstrom T, Larsson M, Nielsen AL, Holmquist L, Ekbom A, Ronnblom A: **Incidence and clinical course of Crohn's disease during the first year - results from the IBD Cohort of the Uppsala Region (ICURE) of Sweden 2005-2009**. *J Crohns Colitis* 2014, **8**(3):215-222.
5. Zhulina Y, Udumyan R, Henriksson I, Tysk C, Montgomery S, Halfvarson J: **Temporal trends in non-stricturing and non-penetrating behaviour at diagnosis of Crohn's disease in Orebro, Sweden: a population-based retrospective study**. *J Crohns Colitis* 2014, **8**(12):1653-1660.
6. Busch K, Ludvigsson JF, Ekstrom-Smedby K, Ekbom A, Askling J, Neovius M: **Nationwide prevalence of inflammatory bowel disease in Sweden: a population-based register study**. *Aliment Pharmacol Ther* 2014, **39**(1):57-68.
7. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, Heurgren M, Olausson PO: **External review and validation of the Swedish national inpatient register**. *BMC Public Health* 2011, **11**:450.
8. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A: **The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research**. *Eur J Epidemiol* 2009, **24**(11):659-667.
9. **Swedish Inflammatory Bowel Disease Registry. www.swibreg.se**. In.
10. Ludvigsson JF, Myrelid P: **[Swibreg--a new version of national IBD registry]**. *Lakartidningen* 2009, **106**(45):3014-3015.
11. Wettermark B, Hammar N, Fored CM, Leimanis A, Otterblad Olausson P, Bergman U, Persson I, Sundstrom A, Westerholm B, Rosen M: **The new Swedish Prescribed Drug Register--opportunities for pharmacoepidemiological research and**

experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007, **16**(7):726-735.
12. Satsangi J, Silverberg MS, Vermeire S, Colombel JF: **The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications**. *Gut* 2006, **55**(6):749-753.
13. Munkholm P: **Crohn's disease--occurrence, course and prognosis. An epidemiologic cohort-study**. *Dan Med Bull* 1997, **44**(3):287-302.
14. Langholz E: **Ulcerative colitis. An epidemiological study based on a regional inception cohort, with special reference to disease course and prognosis**. *Dan Med Bull* 1999, **46**(5):400-415.
15. Fagerland MW, Lydersen S, Laake P: **Recommended tests and confidence intervals for paired binomial proportions**. *Stat Med* 2014, **33**(16):2850-2875.
16. Waldenlind K, Eriksson JK, Grewin B, Askling J: **Validation of the rheumatoid arthritis diagnosis in the Swedish National Patient Register: a cohort study from Stockholm County**. *BMC Musculoskelet Disord* 2014, **15**:432.
17. Lindstrom U, Exarchou S, Sigurdardottir V, Sundstrom B, Askling J, Eriksson JK, Forsblad-d'Elia H, Turesson C, Kristensen LE, Jacobsson L: **Validity of ankylosing spondylitis and undifferentiated spondyloarthritis diagnoses in the Swedish National Patient Register**. *Scand J Rheumatol* 2015:1-8.
18. Dignass A, Eliakim R, Magro F, Maaser C, Chowers Y, Geboes K, Mantzaris G, Reinisch W, Colombel JF, Vermeire S *et al*: **Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 1: definitions and diagnosis**. *J Crohns Colitis* 2012, **6**(10):965-990.
19. Van Assche G, Dignass A, Panes J, Beaugerie L, Karagiannis J, Allez M, Ochsenkuhn T, Orchard T, Rogler G, Louis E *et al*: **The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: Definitions and diagnosis**. *J Crohns Colitis* 2010, **4**(1):7-27.
20. Ordas I, Eckmann L, Talamini M, Baumgart DC, Sandborn WJ: **Ulcerative colitis**. *Lancet* 2012, **380**(9853):1606-1619.

Supplementary Table S1. ICD codes used to identify patients with inflammatory bowel disease (IBD) in the Swedish National Patient Register.

	ICD-7	ICD-8	ICD-9*	ICD-10*
Crohn's disease	572.00	563.00	555	K50
	572.09			
Ulcerative colitis	572.20	563.10	556	K51
	572.21	569.02		
Non-specific IBD#	572,30	563,98-99		

*Only ICD-9 and ICD-10 codes were used to identify patients with inflammatory bowel disease (IBD) in the Patient Registry.

#Only used to define first date of IBD diagnosis in patients who already had a relevant ICD-9 or -10 code.

Supplementary Table S2. Diagnostic criteria for definite and probable diagnoses of Crohn's disease, ulcerative colitis and inflammatory bowel disease unclassified. NPR, national patient register; CD, Crohn's disease; UC, ulcerative colitis.

Crohn's disease	
Definite (at least two of the criteria present)	History of abdominal pain, weight loss *and/or* diarrhea for more than three months
	Characteristic endoscopic findings of ulceration (aphtous lesions, snail track ulceration) *or* cobble stoning *or* radiological features of stricture or cobble stoning
	Histopathology consistent with Crohn's disease (epithelioid granuloma of Langerhans type *or* transmural discontinuous focal or patchy inflammation)
	Fistula and/or abscess in relation to affected bowel segments
Probable	At least two registered diagnoses of Crohn's disease in the NPR *or* at least one registered diagnosis of Crohn's disease in SWIBREG; and
	Documentation of the treating physician judging the patient to have Crohn's disease
Ulcerative colitis	
Definite (all of the criteria present)	History of diarrhea *and/or* rectal bleeding and pus for more than one week or repeated episodes; and
	Characteristic endoscopic findings of continuous ulceration, vulnerability or granulated mucosa; and
	Histopathology consistent with ulcerative colitis (neutrophils within epithelial structures, cryptitis, crypt distortion, crypt abscesses)
Probable	At least two registered diagnoses of ulcerative colitis in the NPR *or* at least one registered diagnosis of ulcerative colitis in SWIBREG; and
	Documentation of the treating physician judging the patient to have ulcerative colitis
Inflammatory bowel disease unclassified	
Definite	Not all criteria for either CD or UC are fulfilled; and
	The patient has received IBD treatment

Probable	Documentation of the treating physician judging the patient to have IBD, but failing to differentiate between Crohn's disease and ulcerative colitis

233
234
235

Supplementary Table S3. Additional analysis of NPR

All patients selected from NPR ** (N=146)								
		Diagnosis according to criteria:						
		UC	**CD**	**IBD-U**	**Not IBD**	**PPV (95% CI)**		
Diagnosis NPR:	**UC**	52	4	5	8	75%	(64-	85)
	CD	9	52	6	10	68%	(56-	78)
	Total:	61	56	11	18	88%*	(81-	93)

*PPV for IBD-U. ** Including those with only one registered diagnosis in NPR

236

Supplementary Table S4. Additional analysis of SWIBREG

Cases with congruent diagnosis in SWIBREG and the NPR (N=143)								
		Diagnosis according to criteria:						
		UC	**CD**	**IBD-U**	**Not IBD**	**PPV (95% CI)**		
Diagnosis in	**UC (N=70)**	69	1	0	0	99%	(92-	100)
NPR/SWIBREG:	**CD (N=73)**	1	69	2	1	95%	(87-	98)

237