

Original Article

Dan Med J 2023;70(4):A07220458

Identifying patients with inflammatory bowel disease in the Danish National Patient Register

Bobby Lo^{1, 2}, Mirabella Zhao^{1, 2} & Johan Burisch^{1, 2}

1) Gastro Unit, Medical Section, Copenhagen University Hospital – Amager and Hvidovre Hospital, 2) Copenhagen Center for Inflammatory Bowel Disease in Children, Adolescents and Adults, Copenhagen University Hospital – Amager and Hvidovre Hospital, Denmark

Dan Med J 2023;70(4):A07220458

ABSTRACT

INTRODUCTION. The Danish National Patient Register (NPR) is an indispensable source of data for population-based studies of inflammatory bowel disease (IBD). Current case-validation algorithms are at risk of overestimating the occurrence of IBD in Denmark. We aimed to develop a new algorithm for validating IBD patients in the Danish NPR and compared it with the algorithm currently used.

METHODS. We used the Danish NPR to identify all IBD patients between 1973 and 2018. In addition, we compared the traditional two-registration validation method with a newly developed ten-step method. Data were provided by Statistics Denmark.

RESULTS. In total, 69,908 IBD patients (Crohn's disease (CD): 23,500 (33.6%); ulcerative colitis (UC): 38,728 (55.4%); IBD unclassified (IBDU): 7,680 (11.0%)) and 84,872 (UC: 51,304 (60.4%), CD: 20,637 (24.3%), IBDU: 9,931 (11.7%)) were identified using the new and the traditional algorithm, respectively, yielding 21.4% more patients. The sensitivity of each algorithm was 98%; however, the new algorithm demonstrated a superior positive predictive value (PPV) (69% (95% confidence interval (CI): 66-72%) versus 57% (95% CI: 54-59%), $p < 0.05$). The overall incidence rate in 2017 was 44.36 (95% CI: 42.66-46.11) versus 53.41 (95% CI: 51.54-55.33, $p < 0.0001$) for the new and the traditional method, respectively.

CONCLUSION. We developed a new and more refined algorithm for validating IBD patients in the Danish NPR. The algorithm will ensure that new studies based upon one of the world's most comprehensive registers will be of an even higher quality. We recommend that all future studies of IBD in Denmark use the new algorithm.

FUNDING. none.

TRIAL REGISTRATION. not relevant.

The Danish National Patient Register (NPR) is an important resource for epidemiological studies within the field of inflammatory bowel disease (IBD) [1]. The NPR may be used for investigating topics like, e.g., financial burden [2], disease course [3] and morbidity [4]. The validity is crucial to the credibility of any results based upon data from the NPR. However, many diseases may imitate IBD, leading to misregistration [5]. Recognising the risk of overestimating the frequency of IBD by relying solely on the occurrence of a single registration has led to the development of algorithms for validating IBD [6-8].

Even so, few efforts have been made to investigate the validity of case-finding algorithms in the Danish NPR [9, 10]. In a study from 1996, the NPR was indirectly validated by comparing IBD patients identified in a regional hospital system with pathology reports [11]. Since then, several changes in registration methods at hospitals and the NPR have occurred. In 2013, Lophaven et al. [9] investigated the effect of requiring several records of an IBD diagnosis in the NPR to validate cases and found that one record produced overestimations, whereas four underestimated the number of patients. It is hypothesised that using data from medical treatments and pathology records may further improve the case-validation algorithm for IBD patients within the NPR.

Therefore, we aimed to develop a new algorithm for validating IBD patients in the Danish NPR and to compare the new algorithm with to the traditional type of algorithm in the hope of improving accurate case detection for register-based studies.

METHODS

Data

The Danish Civil Registration System provides all residents in Denmark with a unique ten-digit identification number [12]. Using this number, we identified all persons with at least one record of an IBD diagnosis ([Table 1 a07220458_supplementary_material.pdf](#) between 1974 and 2018 in any of the following registers: i) the NPR [13], which contains information about all visits, surgeries and procedures conducted at public hospitals in Denmark; ii) the National Register of Unfinished Investigations [13], which contains all information concerning patients who fail to attend appointments despite being in the middle of a medical investigation; and iii) the National Register of Private Hospitals [13], which contains information about all visits, surgeries and procedures conducted at private hospitals in Denmark.

We defined a record as a registration in the NPR when the patient has been in contact with the healthcare system.

Supporting registers

We extracted data to help confirm potential IBD cases from the following registers: i) the National Prescription Registry [14], which contains information about all prescriptions redeemed in Denmark since 1995; and ii) the Danish National Pathology Registry, which contains detailed records of all pathology specimens analysed in Denmark since 1997 [15]).

Development of the algorithm to validate IBD patients

In light of the suggestion by Lophaven et al. [9], two or three records of an IBD diagnosis in the NPR will likely suffice for accurately validating patients. Therefore, we expanded the traditional algorithm by requiring three or more records of an IBD diagnosis, or fewer than three records but in combination with prescriptions for IBD-relevant medications, IBD-related surgeries and/or pathology results suggesting IBD.

Evaluation of the algorithms

We evaluated the new algorithm by comparing it to the most widely adopted approach requiring at least two records of an IBD diagnosis [16]. We evaluated the new algorithm by two methods:

1. Comparing the number of identified patients using both algorithms with a population-based inception cohort from the Copenhagen Country and City from 2003 to 2004 [17]. The main outcome was sensitivity as the Copenhagen Country was more fluidly registered in the registers. The municipality codes used to define the Copenhagen Country and City are provided in [Table 1 a07220458_supplementary_material.pdf](#).

2. Comparing the calculated incidence rates between 1978 and 2017 (the rationale for this is described below) between the two algorithms. In addition, the incidence rates generated by the new algorithm were compared against published incidence rates from previous population-based studies.

Although the registers dated back to 1974, we excluded the years 1974-1977 as the NPR only covered the entirety of Denmark as from 1977. This also excludes the wash-in period. Despite having data until 2018, we chose to report only up to 2017 as registration in 2018 was incomplete because data from hospitals and other centres were still being updated.

Lastly, we conducted supporting evaluating analyses of the registers to estimate the credibility achieved by using these registers. The results were evaluated subjectively based upon expert opinion.

Statistics

Data were processed and analysed using R. We used the Chi-squared test, Fisher's test and Kruskal-Wallis test, as appropriate, with a 95% confidence interval (95% CI). To validate both algorithms against inception cohort, sensitivity, and positive predictive value (PPV) were calculated. We used a polynomial regression analysis with two degrees of freedom to estimate the trajectory of the incidence of IBD patients. All incidence rates were calculated using the method proposed by Ulm (1990) and converted to "per 100,000 person-years".

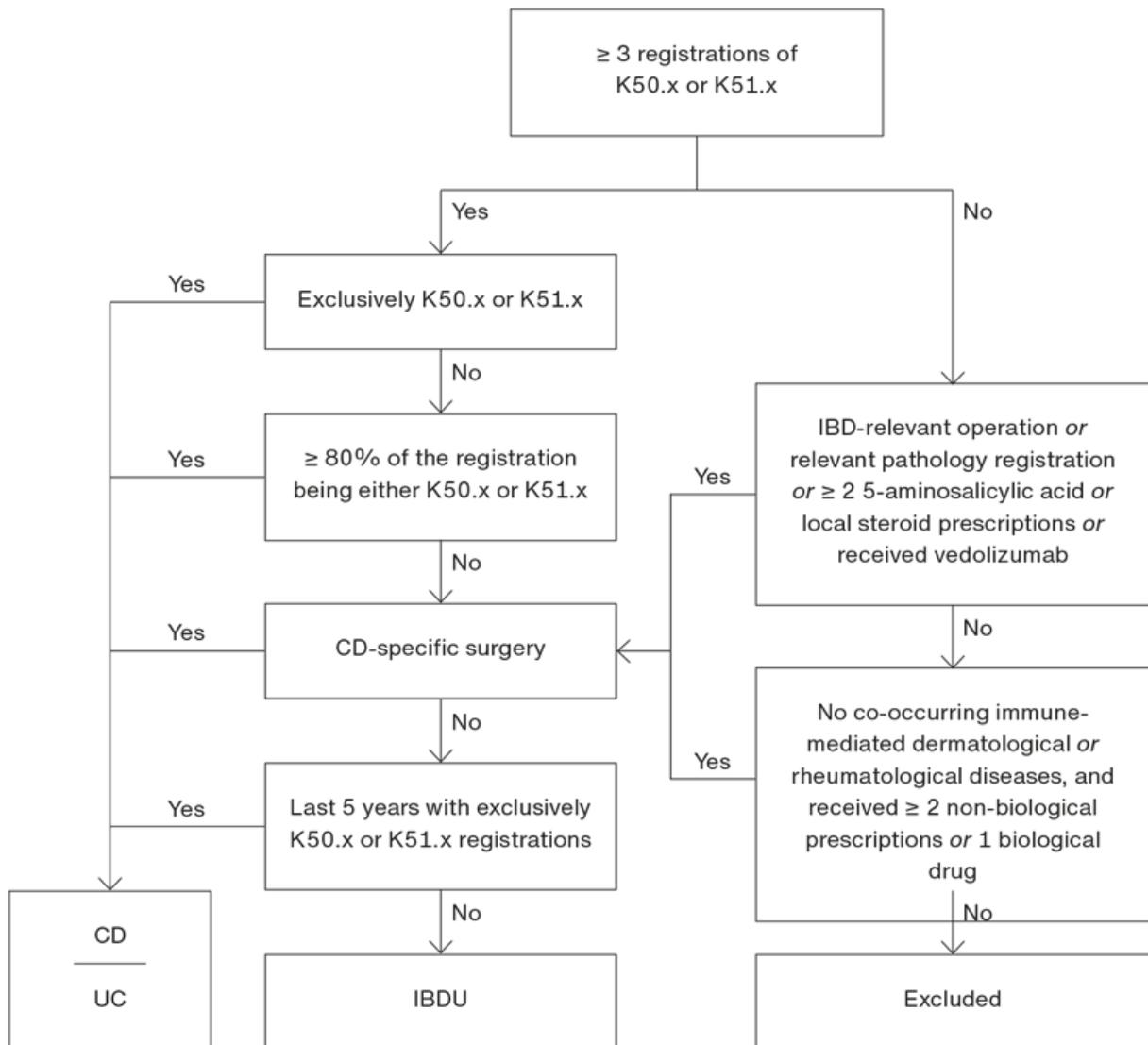
Trial registration: not relevant.

RESULTS

Development of the new algorithm

We developed the following algorithm (**Figure 1**):

FIGURE 1 The newly developed algorithm for validating inflammatory bowel disease (IBD) patients and stratification of Crohn's disease (CD), ulcerative colitis (UC) and IBD-unspecified (IBDU) in the Danish National Patient Register.



1. Identification of all patients with at least one record of an inflammatory bowel disease diagnosis code

All referral records to the hospital were excluded. Records found in the National Registry of Administrative Data (hospital contacts) were removed if the same records were found in the National Registry of Diagnoses (verified diagnoses at the end of each year). This process identified 89,905 potential cases of IBD ([Table 1 a07220458_supplementary material.pdf](#)).

2. Filtering the patients identified

All patients with three or more registrations were included in the cohort. Patients with fewer than three registrations with an IBD code were included if they had a record of an IBD-relevant operation and/or a pathology registration with an unambiguous SNOMED IBD code ([Table 1 a07220458_supplementary material.pdf](#)). In the remaining patients, we searched for redeemed prescriptions for IBD-relevant medications.

First, we included patients with two or more prescriptions for mesalazine or rectal steroids, or vedolizumab, which are treatments used exclusively for IBD ([Table 1 a07220458_supplementary material.pdf](#)) [15]. Among patients not meeting this requirement, we nonetheless included any with a combination of at least two prescriptions or deliveries/treatments at the hospital of thiopurines, mercaptopurine or methotrexate, or one treatment of anti-tumour-necrosis- α inhibitors, anti-interleukin 12/23 antibodies or Janus kinase (JAK) inhibitors. However, we excluded patients with co-occurring immune-mediated dermatological or rheumatological diseases for which these drugs may also be used. This process generated a total of 14,241 IBD cases for classification.

3. Classifying patients as Crohn's disease, ulcerative colitis or inflammatory bowel disease unclassified

Patients with records of Crohn's disease (CD) or ulcerative colitis (UC) diagnosis codes exclusively were categorised as CD or UC, respectively. The remaining patients were categorised as IBD unclassified (IBDU), unless:

- a. 80% or more of their codes were either CD or UC; in such cases, patients were classified as CD or UC, respectively.
- b. Patients with a record of a CD-specific surgery code ([Table 1 a07220458_supplementary material.pdf](#)) were classified as CD.
- c. If patients were recorded with CD or UC codes exclusively during the preceding five years of observation, they were classified as CD or UC, respectively.

Comparison of algorithms

In total, 69,908 IBD patients were validated using the new algorithm. The median time from first record to validation was 11 days (interquartile range (IQR): 183).

Among those IBD patients, 23,500 (33.6%) were categorised as CD, 38,728 (55.4%) as UC and 7,680 (11.0%) as IBDU. In contrast, using the algorithm requiring a minimum of two records of an IBD diagnosis yielded a total of 84,872 IBD patients, i.e. 21.5% more patients. Among those, 23,637 (27.9%) were categorised as CD, whereas 51,304 (60.4%) were classified as UC. The remaining patients (9,931 (11.7%)) were categorised as IBDU.

Among the 513 patients from the inception cohort, 504 (98%) of the patients were identified by the new algorithm and 505 (98%) by the traditional algorithm. In the same period between 2003 and 2004, 729 persons were registered as IBD using the new algorithm, where 889 using the traditional algorithm were registered; yielding a PPV of 69% (95% CI: 66-72%) and 57% (95% CI: 54-59%, $p < 0.05$) for the new and the traditional algorithm, respectively.

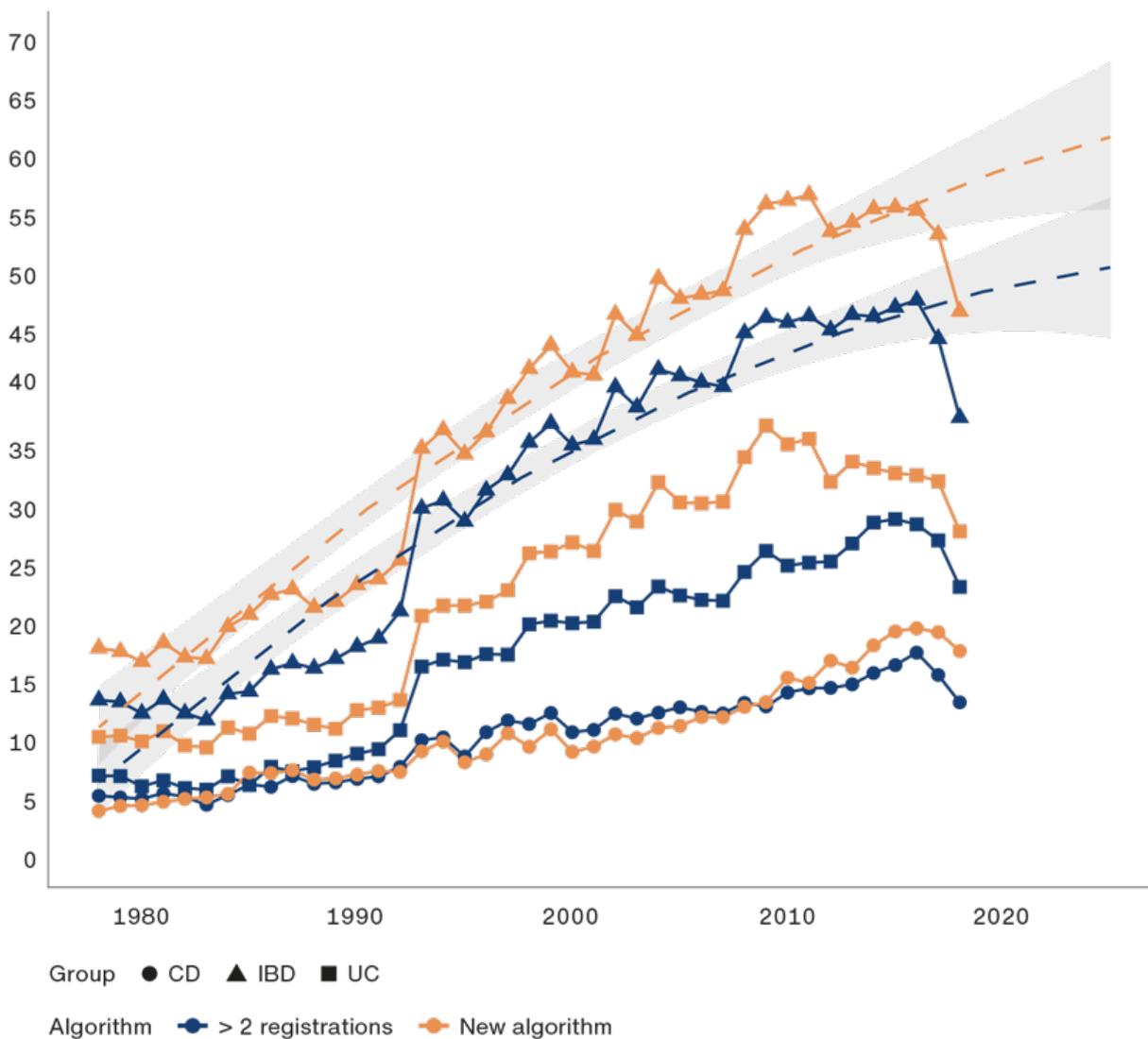
The annual incidence rate for IBD, CD and UC and according to each algorithm are shown in ([Table 2 a07220458_supplementary material.pdf](#)) . In summary, for the 1978-2017 period, the algorithm requiring a minimum of two records generated a significantly higher (21.4%) incidence rate than the newly developed algorithm (37.914 (95% CI: 37.653-38.177) versus 31.242 (95% CI: 31.005-31.480) per 100,000 person years, $p < 0.0001$). When comparing the incidence rates at the end of follow-up, the rate was 20.4% higher according to the traditional algorithm (2017: 53.410 (95% CI: 51.542-55.329) versus 44.359 (95% CI: 42.658-46.110) per 100,000 person years, $p < 0.0001$). At the end of the observation period, the incidence rate generated by the newly developed algorithm was more in line with the rates of previous studies from North America and Northern Europe ([Figure 1 a07220458_supplementary material.pdf](#)), with the exception of rates from the Faroe Islands and Canada [18].

When looking at the overall changes in the incidence rates between 1978 and 2017, the algorithm utilising two or

more registrations tended to overestimate the future IBD incidence rate (equation: $y = -0.012279x^2 + 50.230x - 51.301$), whereas the newly developed algorithm (equation: $y = -0.014267x^2 + 58.047x - 58.993$) produced more realistic projections (estimate difference: -6.3 (-8.16 – -4.49), $p < 0.001$); where x is the year, and y is the estimated incidence rate. The changes in the incidence rates for CD, UC and IBD according to both algorithms are shown in **Figure 2**.

FIGURE 2 Incidence rates (95% confidence intervals) generated by the newly developed inflammatory bowel disease (IBD) validation algorithm and the traditional algorithm using two or more registrations, stratified by IBD, Crohn's disease (CD) and ulcerative colitis (UC).

Incidence rate per 100,000 person years



Comparing the patients validated by each algorithm, the traditional algorithm using two or more registrations found significantly more UC patients (60.4% versus 55.4%, $p < 0.001$); furthermore, these patients tended to be older, even if the median age was the same across the two algorithms (see **Table 1**).

TABLE 1 Demographics of patients with inflammatory bowel disease validated by the two algorithms in the Danish National Patient Register.

Algorithm	New algorithm (N _n = 69,908)	Traditional algorithm ^a (N _t = 84,872)	p value
<i>Diagnosis, n (%)</i>			< 0.001
Crohn's disease	23,500 (33.6)	23,637 (27.9)	
Ulcerative colitis	38,723 (55.4)	51,304 (60.4)	
IBD unclassified	7,680 (11.0)	9,931 (11.7)	
Females, n (%)	37,580 (53.8)	45,715 (53.9)	0.678
Age at diagnosis, median (IQR), yrs	40.00 (26.00-58.00)	40.00 (26.00-58.00)	0.022

IBD = inflammatory bowel disease; IQR = interquartile range.

a) Requiring ≥ 2 diagnosis codes for IBD.

Lastly, when evaluating which patients were identified exclusively by one or the other of the two algorithms (Table 2); patients identified exclusively by the new algorithm were mainly female CD patients with a median age of 62 years (IQR: 44-75).

TABLE 2 Demographics of patients with inflammatory bowel disease validated exclusively by one or the other of the two algorithms in the Danish National Patient Register.

Algorithm	New algorithm (N _n = 1,667)	Traditional algorithm ^a (N _t = 16,672)	p value
<i>Diagnosis, n (%)</i>			< 0.001
Crohn's disease	840 (50.4)	6,416 (38.5)	
Ulcerative colitis	214 (12.8)	9,933 (59.6)	
IBD unclassified	613 (36.8)	323 (1.9)	
<i>Decade, n (%)</i>			< 0.001
1977-1979	55 (3.3)	768 (4.6)	
1980-1989	245 (14.7)	3,036 (18.2)	
1990-1999	252 (15.1)	3,192 (19.1)	
2000-2009	443 (26.6)	4,408 (26.4)	
2010-2017	672 (40.3)	5,268 (31.6)	
Females, n (%)	968 (58.1)	9,126 (54.7)	0.010
Age at diagnosis, median (IQR), yrs	62.00 (44.00-75.00)	44.00 (26.00-64.00)	< 0.001

IBD = inflammatory bowel disease; IQR = interquartile range.

a) Requiring ≥ 2 diagnosis codes for IBD.

DISCUSSION

This study developed a new composite algorithm for validating IBD patients in the Danish NPR. The results generated by the algorithm indicate that the traditional method used in previous studies tends to overestimate the incidence of IBD by more than 20% with a much lower PPV. We suggest that the newly developed algorithm

be used in future research into IBD that utilises the national Danish registers to avoid overestimation.

The use of a combination of different criteria to define administrative cases in population-based registers is not new [7, 19, 20]. The overall conclusion is that by increasing the number of criteria, sensitivity may be improved without the considerable reduction of specificity, to a certain point, thereby reducing the number of false positives. The same was demonstrated in our study, increasing the PPV from 57% to 70%. A PPV of 70% is not perfect. The moderate PPV was due to regional registrational limitations as the inception cohort by Vind et al. [17] was conducted on very specific municipalities and between Jan 2003 and Dec 2004. These registrations were not as strictly recorded in the available data from the national registers due to, e.g., registrational delays. The PPV is presumably higher when evaluated at a national level and covering a longer time period.

This study is not without its limitations, the most conspicuous of which is the lack of validation of a national cohort of verified IBD patients; we were only able to validate the data using a regional cohort. However, we believe that the method used in this study is comprehensive enough to evaluate both algorithms. Firstly, there is no known reason why the Danish incidence rate should vary significantly from that of neighbouring countries in the Northern hemisphere. Secondly, this study made use of the National Prescription Registry and the National Pathology Registry, which were established in 1995 and 1997, respectively. As our cohort dates back to 1973, a gap of 22 and 24 years exists, respectively, where the additional criteria for filtering IBD patients could not be applied. However, when looking at the patients validated by only one or the other of the algorithms (Table 2), most of the patients validated exclusively by the traditional algorithm were mainly from 2000 and later, suggesting that any underestimation by the new algorithm may not be due to this factor. Lastly, the new algorithm requires use of several registers besides the Danish NPR. However, when requesting data from Statistics Denmark, access to all the required databases may be achieved by filling in a single form.

The main strength of this study is our access to a vast quantity of data, enabling us to set a wide range of criteria for the newly developed algorithm. Secondly, we used a clinical inception cohort to estimate the sensitivity and PPV. Lastly, we were able to evaluate uncertain cases of IBD using surgical, histopathological and other medical information, thereby reducing the risk of false positives.

CONCLUSION

We developed a new and more refined algorithm for validating IBD patients in the Danish NPR. The algorithm should ensure that new studies relying on data from one of the world's most comprehensive national registers will be of a high quality. We recommend that all future studies of IBD in Denmark use the new algorithm owing to its superior PPV to reduce the risk of false positives.

Correspondence Bobby Lo. E-mail: bobby.lo@regionh.dk

Accepted 16 February 2023

Conflicts of interest Potential conflicts of interest have been declared. Disclosure forms provided by the authors are available with the article at ugeskriftet.dk/dmj

Cite this as Dan Med J 2023;70(4):A07220458

Link for Supplementary Material https://content.ugeskriftet.dk/sites/default/files/2023-03/a07220458_-_supplementary.pdf

REFERENCES

1. Ng SC, Shi HY, Hamidi N et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2017;390(10114):2769-78. doi: 10.1016/S0140-6736(17)32448-0.

2. Lo B, Vind I, Vester-Andersen MK et al. Direct and indirect costs of inflammatory bowel disease: ten years of follow-up in a Danish population-based inception cohort. *J Crohn's Colitis*. 2020;14(1):53-63. doi: 10.1093/ecco-jcc/jjz096.
3. Rungoe C, Langholz E, Andersson M et al. Changes in medical treatment and surgery rates in inflammatory bowel disease: a nationwide cohort study 1979-2011. *Gut*. 2014;63(10):1607-16. doi: 10.1136/gutjnl-2013-305607.
4. Waljee AK, Higgins PDR, Jensen CB et al. Anti-tumour necrosis factor- α therapy and recurrent or new primary cancers in patients with inflammatory bowel disease, rheumatoid arthritis, or psoriasis and previous cancer in Denmark: a nationwide, population-based cohort study. *Lancet Gastroenterol Hepatol*. 2020;5(3):276-84. doi: 10.1016/S2468-1253(19)30362-0.
5. Feakins R, Torres J, Borralho-Nunes P et al. ECCO topical review on clinicopathological spectrum and differential diagnosis of inflammatory bowel disease. *J Crohns Colitis*. 2022;16(3):343-68. doi: 10.1093/ECCO-JCC/JJAB141.
6. Jakobsson GL, Sternegård 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):216-21. doi: 10.1080/00365521.2016.1246605.
7. Lee CK, Ha HJ, Oh SJ, Kim JW, Lee JK, Kim HS, et al. Nationwide validation study of diagnostic algorithms for inflammatory bowel disease in Korean National Health Insurance Service database. *J Gastroenterol Hepatol* 2020;35:760-8. <https://doi.org/10.1111/JGH.14855>.
8. Jones GR, Lyons M, Plevris N et al. IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. *Gut*. 2019;68(11):1953-60. doi: 10.1136/gutjnl-2019-318936.
9. Lophaven SN, Lyng E, Burisch J. The incidence of inflammatory bowel disease in Denmark 1980-2013: a nationwide cohort study. *Aliment Pharmacol Ther*. 2017;45(7):961-72. doi: 10.1111/APT.13971.
10. Lo B, Vind I, Vester-Andersen MK, Burisch J. Validation of ulcerative colitis and Crohn's disease and their phenotypes in the Danish National Patient Registry using a population-based cohort. *Scand J Gastroenterol*. 2020;55(10):1171-5. doi: 10.1080/00365521.2020.1807598.
11. Fonager K, Sørensen HT, Rasmussen SN et al. Assessment of the diagnoses of Crohn's disease and ulcerative colitis in a Danish hospital information system. *Scand J Gastroenterol*. 1996;31(2):154-9. doi: 10.3109/00365529609031980.
12. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health*. 2011;39(7 suppl):22-5. doi: 10.1177/1403494810387965.
13. Schmidt M, Schmidt SAJ, Sandegaard JL et al. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449-90. doi: 10.2147/CLEP.S91125.
14. Kildemoes HW, Sørensen HT, Hallas J. The Danish National Prescription Registry. *Scand J Public Heal*. 2011;39(7 suppl):38-41. doi: 10.1177/1403494810394717.
15. Sands BE, Peyrin-Biroulet L, Loftus EV et al. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. *N Engl J Med*. 2019;381(13):1215-26. doi: 10.1056/NEJMoa1905725.
16. Burisch J, Zhang H, Choong CKC et al. Validation of claims-based indicators used to identify flare-ups in inflammatory bowel disease. *Therap Adv Gastroenterol*. 2021;14:17562848211004841. doi: 10.1177/17562848211004841.
17. Vind I, Riis L, Jess T et al. Increasing incidences of inflammatory bowel disease and decreasing surgery rates in Copenhagen City and County, 2003-2005: a population-based study from the Danish Crohn colitis database. *Am J Gastroenterol*. 2006;101(6):1274-82. doi: 10.1111/j.1572-0241.2006.00552.x.
18. Ng SC, Shi HY, Hamidi N et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2017;390(10114):2769-78. doi: 10.1016/S0140-6736(17)32448-0.
19. Rezaie A, Quan H, Fedorak RN et al. Development and validation of an administrative case definition for inflammatory bowel diseases. *Can J Gastroenterol*. 2012;26(10):711-7.
20. Jakobsson GL, Sternegård E, Olén 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):216-21. doi: 10.1080/00365521.2016.1246605.