Inflammatory bowel disease and health registry data
Estimates of incidence, prevalence and regional treatment variation

Sandre Svatun Lirhus

Thesis submitted for the Degree of Philosophiae Doctor (PhD)

Department of Health Management and Health Economics
Institute of Health and Society
Faculty of Medicine
University of Oslo
Norway
January 2022
Summary

The aim of this thesis is to increase the understanding of how we can use health registry data in inflammatory bowel disease (IBD) research and provide knowledge on the epidemiology and treatment of IBD in Norway. IBD includes two major disorders, Crohn’s disease (CD) and ulcerative colitis (UC).

For IBD patients, there are several accepted treatment options available, including both different medical therapies and surgery. Paper I studies the regional variation of surgery and advanced medical therapy among IBD patients in Norway’s four health regions. The study found that the cumulative incidence of advanced medical therapy and surgery varied across the health regions.

There is conflicting evidence on the development of IBD incidence over time and geographical areas. Since IBD is a chronic disease with low mortality, the prevalence will likely increase over time due to an aging population. Paper II provides estimates for IBD incidence and prevalence in Norway and assess the impact of different case definitions on these estimates. The base case incidence was 14.6 per 100,000 person-years for CD and 25.7 per 100,000 person-years for UC in 2017. The base case prevalence was 0.27% for CD and 0.50% for UC in 2017. The results suggest that Norway has among the highest incidence and prevalence of IBD in the world.

Paper III introduces a novel method to estimate the incidence of CD using hospital data. Based on the results from Paper II & III we provide recommendations for how an incident patient should be defined when utilizing registry data, and different ways to determine incidence and prevalence dependent on the data available. The findings from this thesis enable future register-based research on IBD to be based on more informed methodological choices and increased awareness of the strength and limitations of two of Norway’s largest health registries; the Norwegian Patient Registry and the Norwegian Prescription Database.
Acknowledgments

First, I would like to thank my main supervisor Hans Olav Melberg. You have an exceptional positive and motivating presence. This PhD would not have been possible without you. A big thanks and appreciation to my two co-supervisors Marte Lie Høivik and Bjørn Moum. I could not ask for a better supervisor trio.

I would also like to thank Fredrik Arneberg who initiated the project. Thanks to Borghild, Hallstein and Merete who saw potential in the project and increased the financing. This allowed me to go from part time researcher to a PhD student. Thanks to the University of Oslo for hosting me.

Thanks to all my friends and colleagues at the department of health economics and health management. A special thanks to Mathyn for sharing an office with me and making my days at the university more enjoyable. You have been of more importance than you can imagine.

Thanks to all my friends and family who care about my wellbeing, and not so much about the PhD. This has been highly appreciated.
Norsk sammendrag

Inflammatorisk tarmsykdom (IBD) er en kronisk betennelse i mage-tarmkanalen som hovedsakelig inkluderer Crohn's sykdom (CD) og ulcerøs kolitt (UC). UC forekommer i tykktarmen, mens CD kan ramme hele mage-tarmkanalen. Det er usikkert hvorfor IBD oppstår, men det er definert som en immunmediert sykdom der kroppens immunforsvar feilaktig angriper egne celler. En IBD-diagnose er basert på kliniske, endoskopiske, radiologiske og histologiske funn, og er vanligvis satt i spesialisthelsetjenesten. Det er estimert at ca. 40 000 lever med IBD i Norge.


I Norge blir all informasjon om sykehusbesøk og resepter for norske innbyggere registrert i sentrale databaser. Disse databasene har et stort forskningspotensiale da de dekker millioner av mennesker over mange år. Det er derfor norske registerdata ofte omtales som en gullgruve. Forskning på sykdommer basert på registerdata har mange potensielle fallgruver, og denne oppgaven ser nærmere på hvordan noen av disse begrensningene kan håndteres.


- IBD kan behandles både med flere ulike medikamenter og kirurgi. Artikkel I undersøkte variasjon i bruk av ulike behandlingsalternativer blant IBD-pasienter i Norges fire helseregioner. I Artikkel I fant vi at det var forskjeller i bruk av
avansert medisinsk behandling og kirurgi på tvers av helseregionene.

- Eksisterende litteratur viser motstridende resultater om hvordan IBD-insidens har utviklet seg over tid og geografiske områder. Siden IBD er en kronisk sykdom med lav dødelighet vil prevalansen sannsynligvis øke over tid på grunn av en aldrende befolkning. En av utfordringene med kroniske sykdommer i registerdata er å skille prevalente og insidente pasienter. Artikkel II estimerte insidens og prevalens av IBD i Norge, samt effekten av å variere definisjonen av hvem som ble klassifisert som en IBD pasient. I 2017 var insidensen av CD 14,6 per 100 000 personår og av UC 25,7 per 100 000 personår, mens prevalensen var henholdsvis 0,27% og 0,50%. Resultatene antyder at Norge er et av landene i verden med høyest insidens og prevalens av IBD.

- Artikkel III introduserte en ny metode for å estimere forekomsten av CD ved bruk av data kun fra Norsk pasientregister. Basert på resultatene fra Artikkel II & III gir vi anbefalinger for hvordan insidens bør beregnes ved bruk av registerdata, samt ulike måter å bestemme insidens og prevalens avhengig av tilgjengelige data.

Avhandlingen er den første i Norge som bruker data om IBD fra to store, nasjonale helseregistre, og baner vei for fremtidig registerbasert forskning på IBD. Funnene i avhandlingen vil gjøre det enklere å gjøre gode metodologiske valg med økt bevissthet om styrkene og begrensningene til Norsk pasientregister og Reseptregisteret, både for videre forskning på IBD, men også andre kroniske sykdommer.
List of papers

Paper I – Regional differences in anti-TNF-α therapy and surgery in the treatment of inflammatory bowel disease patients: A Norwegian nationwide cohort study
https://doi.org/10.1080/00365521.2018.1495258

Paper II – Incidence and prevalence of inflammatory bowel disease in Norway and the impact of different case definitions: A nationwide registry study
https://doi.org/10.2147/CLEP.S303797

Paper III - Wash-out and look-forward bias when estimating incidence based on patient registry data: A new method and a case study of Crohn’s Disease in Norway
Melberg HO, Lirhus SS, Moum B, Bauer M, Burisch J, Høivik ML,
Submitted
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATC</td>
<td>Anatomic Therapeutic Chemical classification system</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn's disease</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>IBDU</td>
<td>Inflammatory bowel disease unclassified</td>
</tr>
<tr>
<td>IBSEN</td>
<td>Inflammatory Bowel Disease in South Eastern Norway</td>
</tr>
<tr>
<td>IC</td>
<td>Indeterminate colitis</td>
</tr>
<tr>
<td>ICD-10</td>
<td>International Classification of Diseases, 10th edition</td>
</tr>
<tr>
<td>ICPC</td>
<td>International Classification of Primary Care</td>
</tr>
<tr>
<td>NCMP</td>
<td>Norwegian Classification of Medical Procedures</td>
</tr>
<tr>
<td>NCRP</td>
<td>Norwegian Classification of Radiological Procedures</td>
</tr>
<tr>
<td>NCSP</td>
<td>Norwegian Classification of Surgical Procedures</td>
</tr>
<tr>
<td>NorPD</td>
<td>Norwegian prescription database</td>
</tr>
<tr>
<td>NPR</td>
<td>Norwegian Patient Registry</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative predictive value</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive predictive value</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised control trial</td>
</tr>
<tr>
<td>SSB</td>
<td>Statistics Norway</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour necrosis factor</td>
</tr>
<tr>
<td>UC</td>
<td>Ulcerative colitis</td>
</tr>
</tbody>
</table>
# Table of Contents

1 Introduction and aim of thesis ................................................................. 13

2 Background ............................................................................................... 16
  2.1 IBD ......................................................................................................... 16
  2.2 Medical treatment ................................................................................... 16
  2.3 Surgery .................................................................................................... 18
  2.4 Surgery & biologics .................................................................................. 19
  2.5 Treatment variation ................................................................................. 20

3 Data ........................................................................................................... 21
  3.1 The Norwegian Patient Registry (NPR) ................................................... 21
  3.2 The Norwegian Prescription Database (NorPD) ....................................... 22
  3.3 Statistics Norway (SSB) ........................................................................... 23
  3.4 Data quality ............................................................................................. 23

4 Methods ..................................................................................................... 25
  4.1 Statistical methods .................................................................................. 25
  4.2 Ethical considerations ............................................................................. 26

5 Summary of results .................................................................................... 27
  5.1 Paper I .................................................................................................... 27
  5.2 Paper II ................................................................................................... 29
  5.3 Paper III ................................................................................................. 32

6 Discussion of main findings ...................................................................... 33
  6.1 Paper I .................................................................................................... 33
  6.2 Paper II and III ....................................................................................... 34
    6.2.1 Incidence .......................................................................................... 34
    6.2.2 Prevalence ....................................................................................... 36

7 Discussion of methods ............................................................................... 37
  7.1 Observational versus interventional studies ........................................... 37
  7.2 Change of IBD subtype and IBD unclassified ......................................... 39
  7.3 UC vs CD ............................................................................................... 40
  7.4 Definition of an incident IBD patient ..................................................... 41
  7.5 Regional differences .............................................................................. 42
  7.6 Validation ............................................................................................... 43
1 Introduction and aim of thesis

During the last decades an increasing amount of health data has been collected in large registries all over the world\(^1\). Often the main reason for the data collection is administrative purposes in order to calculate payments. However, such data may also have important secondary uses and it has been argued that better utilisation of administrative data has “significant potential to facilitate research, improve quality of care for individuals and populations, and reduce healthcare costs”\(^2\). In this thesis, we used such data with the aim to increase our knowledge of the epidemiology and treatment of inflammatory bowel disease (IBD) in Norway.

IBD is characterized by chronic inflammation of the gastrointestinal tract and it is a group of diseases that affects about 40 000 individuals in Norway\(^3\). There are two main subtypes of IBD: Crohn’s disease (CD) and ulcerative colitis (UC). The etiology and pathogenesis of IBD is still largely unknown, but it is defined as an immunomodulated disease, where the immune system attacks the gastrointestinal tract\(^4\). IBD is a heterogeneous group of diseases as it include different sub- and phenotypes, varies in severity and includes multiple disease pathways. An IBD diagnosis is based on clinical, endoscopic, radiological and histologic findings and is normally set in the secondary health care services\(^5,6\).

In Norway, all information on hospital visits and prescriptions are registered in central databases. This information holds great potential as it covers millions of people for many years. This is why Norwegian registry data is often referred to as a gold mine\(^7\). But it is also a gold mine that is difficult to enter, which is why it has also been referred to as an inaccessible treasure\(^8\). In addition, once you enter, there are many pitfalls and this thesis will also explore how some of these limitations can be handled. Registry data can be used to analyse many aspects of a disease, and in this thesis we have focused on the following:

**Paper I:** For IBD patients, there are generally more than one accepted treatment option available, such as elective surgery or medical therapy for patients with moderate to
severe IBD. Failure of medical therapy is a common indication for surgery, but surgery is also an option without having failed medical therapy. There can be different opinions on the right course of action, which makes IBD susceptible to treatment variation. In Paper I we aimed to examine the use of surgery and anti-tumour necrosis factor (TNF) among IBD patients in Norway’s four health regions.

**Paper II:** IBD is a chronic disease with low mortality, which is why the prevalence is expected to increase over time due to an aging population\(^9\). However, there is somewhat conflicting evidence on the development of IBD incidence\(^{10–12}\). The aim of Paper II was to estimate the incidence and prevalence for Norway. Furthermore, due to uncertainty of the accuracy of the data, we aimed to assess the impact of different definitions of an IBD patients.

**Paper III** introduces a novel method to estimate the incidence of CD using hospital data. Based on our findings from Paper II & III, we provide recommendations for how an incident patient should be defined when utilizing registry data, and different ways to determine incidence and prevalence dependent on the data available.

**Rationale for order of Papers**

When Paper I was written we only had access to hospital data and required the patients to have at least two IBD-related hospital visits to be included as incident patients. This is a common definition of an incident IBD patient in the literature when using health registry data\(^{13–18}\). After receiving prescription data we explored the possibilities of improving the incident definition (Paper II). It is not common to use prescription data when defining an incident IBD patient. This can either be due to data availability or simply what researchers are accustomed to. We also wanted to develop a method that allowed for the estimation of incidence when only hospital data are available (Paper III).
Structure of thesis

As we aim to explore both the nature of the disease and the data, the thesis is structured as follows: First, we present an overview of the disease and its associated treatments. Next in Chapter 3 follows an introduction to the data used. Chapter 4 includes statistical methods. Chapter 5 summarizes the results from each paper, followed by Chapter 6 where we discuss the main findings and compare our results with the existing literature. Next in Chapter 7, we further discuss methodological challenges and choices. Chapter 8 concludes the thesis.
2 Background

2.1 IBD

CD and UC overlap in clinical and pathological features, but the diseases have clear
differences in clinical presentation, underlying genetic factors and treatment response. UC occur in the colon, damaging the colonic mucosa, which often leads to erythema, erosions, ulcers and bleeding. CD can affect the entire gastrointestinal tract from the mouth to the anus and is characterized by skip lesions and transmural inflammation.

Both CD and UC usually involve symptoms like severe diarrhea, abdominal pain, fatigue and weight loss. In addition to systemic symptoms, mental health problems such as anxiety and depression is also common. IBD can have significant negative impacts on the quality of life of the patient, as well as incurring high costs to the healthcare system and society. The societal costs are caused by higher rates of unemployment, permanent work disability and sick leave. The aim of treatment is to induce and then maintain remission, i.e. normalisation of the inflamed gastrointestinal tract. The management of patients with IBD is complex and patients need personalised care. The number of treatment options and treatment strategies, including different types of medications and surgery, is increasing and there is an ongoing debate on what the optimal treatment is.

2.2 Medical treatment

Traditionally it has been common to follow a ‘step-up’ treatment pathway for mild to moderate IBD patients. The ‘step-up’ treatment involves four levels of medical treatment; 5-aminosalicylic acids (5-ASA) (primarily for UC), steroids, immunomodulators and biologics. The patient starts with the lower steps (5-ASA and/or steroids) and if treatment does not improve symptoms, you go on to the next step. Mild to moderate UC patients commonly start with 5-ASA as their first-line therapy. Systemic steroids are superior to 5-ASA for induction of remission, but due to
its side effects it should be reserved for patients with failure of response to 5-ASA\textsuperscript{31}. If conventional therapies fail, biological therapy is recommended. Currently anti-TNFs (infliximab, adalimumab and golimumab) are the first line biologic drugs of choice. If anti-TNF therapy fails, other biological therapies are indicated (e.g. anit-integrins (vedolizumab) or anti IL12/IL23 (ustekinumab)). If medical treatment do not suffice, surgery is the next step\textsuperscript{32,33}. Some patients undergo surgery at the time of diagnosis if they have acute and severe disease.

For mild to moderate CD patients, budesonide, a locally acting steroid, is recommended for induction of remission\textsuperscript{34}. Budesonide is less effective than systemic steroids, but has fewer side-effects\textsuperscript{35,36}. In moderate-to-severe CD, systemic steroids are recommended for induction of remission. However, long term steroid use should be avoided due to its side-effects\textsuperscript{37}. Patients with moderate-to-severe disease who do not respond to conventional therapy usually receive an anti-TNF drug as their first biological therapy. It is recommended to combine infliximab with thiopurines when initiating treatment, but some advice against it for adalimumab\textsuperscript{34,38,39}. Whether or not patients should get thiopurines in addition to their biologic treatment is dependent on the type of biologic, and the current evidence is deemed insufficient to recommend this strategy for all IBD patients\textsuperscript{34,40}. As with UC, medical treatment failure is the main reason leading to surgical treatment in CD patients\textsuperscript{41}. However, recent studies have shown that early surgery can have equivalent outcomes compared to biological therapy in localized ileal CD\textsuperscript{42}.

The optimal treatment strategy for IBD depends on subtype, previous response to treatment and disease severity, activity and location\textsuperscript{5}. There is evidence suggesting that “top down” therapy (i.e. starting advanced medical therapies, such as biologicals, early after diagnosis (Figure 1) can alter disease progression and prevent future complications\textsuperscript{43,44}. Earlier use of anti-TNFs have mounting evidence in CD, but there is more uncertainty regarding UC\textsuperscript{45–47}. In addition, there is risk of overtreatment and concurrent side effects in patients with milder disease\textsuperscript{48}. To what extent the now more common “top down” approach will decrease the surgery and hospitalization rates and increase the quality of life over long periods of time remains unknown. This is a central and ongoing discussion within the IBD field today.
2.3 Surgery

The lifetime risk of surgery ranges from 50% to 80% in CD and 10% to 30% in UC\textsuperscript{50-52}. Surgery rates vary substantially over time and between countries\textsuperscript{53}. A European multicentre population-based cohort reported a 10.4% 10-year cumulative risk of colectomy for UC patients in Northern European centres and 3.9% in the Southern European centres\textsuperscript{54}. In the Norwegian IBSEN (Inflammatory Bowel Disease in South Eastern Norway) cohort from 1990-1994, the cumulative colectomy rate for UC was 3.5%, 7.6%, 9.8% and 13% 1, 5, 10 and 20 years after diagnosis\textsuperscript{55,56}. Population-based studies from Denmark and Sweden published in the early 1990s reported 10-year colectomy rates of 24% and 28% for UC\textsuperscript{57,58}.

In Europe, the rate of surgery for CD ranged between 12% and 27% five years after diagnosis\textsuperscript{24}. In IBSEN, the cumulative probability of surgery for CD patients at 1, 5 and
10 years from diagnosis was 14%, 27% and 38%. This suggests that Norway had among the highest surgery rates for CD in the world, but no updated numbers are available. Even though there are large variations between countries of the reported surgery rates in IBD, the surgery rates have decreased during the last decades. The decline of surgery rates began prior to the introduction of anti-TNFs, which makes it difficult to accredit the reduction in surgery only to the improvement of treatment.

2.4 Surgery & biologics

There are numerous studies addressing the relationship between surgery and medical therapies, and especially anti-TNFs, in recent times. A systematic review with meta-analysis of studies with patients from 1980-2016 found that anti-TNF treatment reduced surgery compared to placebo with an odds ratio of 0.23 for CD and 0.67 for UC. Other studies reported similar findings where they found increased use of biologics and a subsequent decrease in surgery rates and/or time to surgery. Still, existing studies have failed to demonstrate how biologics influence the long-term course of the diseases and the need for surgical treatment.

The use of anti-TNF before and/or after surgery is controversial. There is conflicting evidence on postoperative complications rates from abdominal surgery during or after anti-TNF treatment. In 2014 a study dedicated to determine risk factors for postoperative IBD was initiated. To date, the results available are from an abstract which found no effect of anti-TNFs on postoperative complications for IBD patients. The use of biologics seems safe when used around the time of surgery, but it is not established how biologics use pre- or post-surgery affect postoperative complications.
2.5 Treatment variation

Typically, patients who end up with surgery while on anti-TNF therapy are considered cases of treatment failure. This is because the success of anti-TNFs is often defined as their capability to induce and maintain remission and reduce surgery and hospitalization rates. It is argued that surgery should rather be seen as an important complement in a multidisciplinary approach to improve the quality of life in patients. However, there are different opinions on the right course of action, which makes IBD susceptible to treatment variation. It is difficult to weigh the benefits of medical versus surgical therapy, concerns about side effects of treatment, costs, morbidity and patients' quality of life. In general, everyday clinical practice have treatment variations which can lead to inadvertent differences in health care outcomes and costs. The combination of heterogeneous disease phenotypes and disease course, different physician opinions, uncertain evidence and an increasing arsenal of available treatment make IBD susceptible to treatment variation.
3 Data

Norway has a long history of health registries and one of the world’s first national patient registry for any disease was founded in Norway in 1856, the Leprosy registry\(^\text{87}\). Today, Norway has 17 central health registries and 52 medical quality registries\(^\text{88}\). Since 1964, every Norwegian resident is registered with a unique personal identification number in the centralized civil registration system. This allows all residents to be followed over time and enables linkage between registries. This opened up new possibilities for health research, which is why Norway’s health data and its registries are sometimes referred to as the “new oil”\(^\text{89}\).

This thesis utilized data from The Norwegian Patient Registry (NPR), the Norwegian prescription Database (NorPD) and Statistics Norway (SSB).

The data included all patients in Norway who were registered in NPR with an ICD-10 code for IBD (ICD-10 code K50 (CD) or K51 (UC)) during the available data periods. The data from NPR and NorPD were at an individual level, meaning that every hospital contact and prescription was in separate lines of data for all Norwegians ever registered with IBD in NPR.

3.1 The Norwegian Patient Registry (NPR)

The main purpose of the NPR is to form the basis for administration, management and quality assurance of specialist health services, including financing\(^\text{90}\). In addition, it also serves to contribute to medical and health research. NPR was established in 1997 and is owned and managed by the Norwegian Directorate of Health\(^\text{91}\). Health care in Norway is mainly provided by tax-based public services and it is mandatory to report patient administrative data to NPR. NPR started to include personal identification numbers in 2008, which made it possible to follow individuals over time and enabled the linkage of data between different registries.
NPR encompasses all public specialist health-care services in Norway, as well as medical specialists and private institutions contracted to regional health authorities. Registrations in NPR can include up to two main diagnoses and 20 secondary diagnoses as ICD-codes. Furthermore, NPR contains procedure codes which provide additional information on the hospital visits. These include the Norwegian Classification of Medical Procedures (NCMP), Norwegian Classification of Surgical Procedures (NCSP), Norwegian Classification of Radiological Procedures (NCRP) and the Anatomical Therapeutic Chemical (ATC) classification system to classify drugs. Both intravenous and subcutaneous administered drugs can be included as ATC-codes in NPR.

3.2 The Norwegian Prescription Database (NorPD)

NorPD was established in 2004 and is owned and managed by the Norwegian Institute of Public Health. NorPD is a pseudonymised registry, meaning that the patient's personal identity number is replaced with a pseudonymised identifier. This is why it was not possible to include patients with IBD-prescriptions or diagnosis from NorPD if they did not have an IBD diagnosis in the NPR.

NorPD stores prescription data at an individual patient level from all Norwegian pharmacies. Each line in the NorPD is a unique dispensation from a pharmacy, it is not the entire prescription made by a physician. NorPD includes ATC-codes and its associated variables (e.g. drug strength, daily defined dose, number of packages, tablets, pens and syringes and active ingredient). NorPD includes subcutaneous administered drugs since they are collected from a pharmacy.
3.3 Statistics Norway (SSB)

Annual population by age and sex was obtained from Statistics Norway\textsuperscript{93}. The population is measured 1st January each year. These numbers were used in Paper II & III to make estimates of incidence and prevalence.

3.4 Data quality

When conducting registry-based research, it is important to have knowledge of the data quality. An indicator of data quality is the proportion of patients who received a link to a valid birth number or D-number. In 2019, this was 99.5\% for somatic contract specialists and 99.7\% for somatic hospitals\textsuperscript{94}. The fact that most patients can be followed over time reduce potential bias in study results when using patient level data.

The NPR is generally considered to be of good quality\textsuperscript{95}. The Norwegian directory of health publish annual reports on the validation and completeness of NPR by comparing data from medical quality registries to corresponding NPR data\textsuperscript{96}. Completeness “is a measure of the amount of available data from a statistical system compared to the amount that was expected to be obtained”\textsuperscript{97}. In 2019, nine medical quality registers had a NPR completeness analysis performed. With the exception of hip dysplasia (67.9\%), the values ranged from 91.9\% to 98.7\%.

In 2017, the Office of the Auditor General investigated the medical code practice in the health trusts\textsuperscript{98}. They concluded that the medical coding was of poor quality, but that they could not make general conclusions based on the results. They only investigated two diagnoses, hip fractures and pneumonia, for 600 hospital stays. One of the biggest flaws in the data was the fact that 41\% of the pneumonia hospital contacts had the wrong diagnosis listed as the main diagnosis. Among these registrations, a distinction was made between contacts that received a new main diagnosis (16\%), and stays that changed at the third sign level (25\%). In the case of IBD, a change at the third sign level could be a change from K50 (CD) to K51 (UC). This can indicate some lack in precision.
of the IBD diagnosis, which would explain some of the methodological challenges faced in this thesis.

Table 1 Data sources, years and variables used in each paper

<table>
<thead>
<tr>
<th>Paper</th>
<th>Data source</th>
<th>Years</th>
<th>Variables used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper I</td>
<td>NPR</td>
<td>2008-2015</td>
<td>ID, date, birthyear, sex, health region, and NCSP-, ATC-, ICD-codes</td>
</tr>
<tr>
<td>Paper II</td>
<td>NPR</td>
<td>2008-2017</td>
<td>ID, date, birth cohort (10-year intervals), sex, NCSP-codes, ATC-codes, ICD</td>
</tr>
<tr>
<td></td>
<td>NorPD</td>
<td>2004-2018(August)</td>
<td>ID, date, ATC-, ICD-, ICPC-codes</td>
</tr>
<tr>
<td></td>
<td>SSB</td>
<td>2010-2017</td>
<td>Population, year, age and sex</td>
</tr>
<tr>
<td>Paper III</td>
<td>NPR</td>
<td>2008-2017</td>
<td>ID, date, ICD-code</td>
</tr>
<tr>
<td></td>
<td>SSB</td>
<td>2009-2017</td>
<td>Population, year, age and sex</td>
</tr>
</tbody>
</table>

ATC, Anatomic Therapeutic Chemical classification system; ICD, International Classification of Diseases; ICPC, International Classification of Primary Care; NCSP, Norwegian Classification of Surgical Procedures; NorPD, Norwegian prescription database; NPR, Norwegian Patient Registry; SSB, Statistics Norway
4 Methods

4.1 Statistical methods

Most of the statistics in this thesis is of descriptive nature in that we describe something that has already happened without predicting or establishing causal relationships. Common descriptive studies include incidence and prevalence. These type of studies play an essential role as they provide the scientific basis for future studies and influence decisions about priority setting, investments and possible interventions.

All data handling and analysis was done with Python 3.X.

Paper I

The cumulative incidence of anti-TNF exposure and major surgery was estimated using the Kaplan-Meier method. The cumulative incidences were compared between health regions using the log-rank test. In order to test for differences in the proportions of patients receiving anti-TNFs before and after surgery, and for patients not receiving surgery, the Chi-Square test was used. Categorical variables were described as number (n) and percentages (%), and Continuous variables were described as median and interquartile range (IQR). The p-value had a threshold of 0.05 to determine statistical significance.

Paper II

Incidence and prevalence of IBD was estimated for different case definitions. Prevalence was presented as point-prevalence proportion with 95% CI at December 31st each year. Incidence was presented as incidence per 100,000 person-years with 95% CI. Incidence was calculated by dividing new IBD patients each year by the estimated mid-year population in the same year.
Paper III

This study quantified the effect different washout and look-ahead periods had on the incidence of CD. A non-linear regression was fitted to estimate the relationship between the washout period and the share of true incident patients. We required patients to have at least two IBD hospital contacts and removed patients with only one contact. This was done by using a non-linear regression to estimate the proportion of patients who only had one visit. Subtracting those with only one IBD-visit from the incidence estimate adjusted for washout, led to an estimate of CD incidence adjusted for both washout and look-ahead bias.

4.2 Ethical considerations

The study was approved by the Regional Committees for Medical Health Research Ethics (REC) (# 2016/113), the Data Protection Services, the Norwegian Institute of Public Health and the Norwegian Directorate of Health.

The source data used in this thesis contains information of over 50,000 people, making informed consent impossible. In registry-based research it is usually not required to obtain informed consent. When receiving approval from REC, it is assumed that the study participants do not object to the research. This is an assumed agreement between the individual and the state given the virtually free supply of health care. Since approval from REC replaces individual approval it is required that the well-being of the study participants is considered and that the benefit of the study outweighs the potential harms. The utilization of pre-existing data entails no inconvenience for the patients. The biggest risk with this sensitive information is that the patient may be identified from the data. In addition to being illegal, you would need prior knowledge of the person in order to identify the person, and as a researcher you should have no interest in finding the person behind the data.
5 Summary of results

5.1 Paper I

The aim of Paper I was to determine the use of anti-TNFs and major surgery in IBD patients across the four health regions in Norway (West, South-East, Central and North). The paper utilized NPR data from 2008-2015. IBD patients were followed for three years from their first IBD hospital visit in 2010-2012. For CD patients, the cumulative incidence of surgery was 13.8%, while it was 5.2% for UC (Figure 2). Across the four health regions, the surgery rates in CD ranged between 11.4% (North) and 17.1% (Central). For UC, the surgery rates ranged between 4.6% and 6.9%. Among the CD patients, 28.8% received anti-TNF treatment, while it was 11.8% for UC. The anti-TNF exposure ranged between 20.9% and 31.4% for CD and between 8% and 13.5% for UC (Figure 3).

Figure 2 Cumulative incidence of first major surgery by diagnosis and region.
Both the cumulative incidence of anti-TNF exposure and surgery varied significantly across Norway’s health regions during the first three years after their first IBD registration. The higher the proportion of patients who received anti-TNFs prior to surgery, the lower was the proportion of patients who underwent surgery. The regions that used the most anti-TNFs prior to surgery had the lowest cumulative incidence of surgery (e.g. South-East gave the most biologics prior to surgery and had the lowest surgery rates). The proportion of patients receiving anti-TNFs were more similar in patients not undergoing surgery; between 20.9% and 28.5% for CD and from 7.2% to 11.9% for UC (Figure 4).

*Figure 3 Cumulative incidence of anti-TNF exposure by diagnosis and region.*
Figure 4 Proportion of CD and UC patients receiving anti-TNFs before and after surgery and if they received no surgery

5.2 Paper II

Paper II show how the incidence and prevalence of IBD in Norway vary with different case definitions. The base case definition in the study defines incident patients as patients with two IBD hospital visits or one IBD hospital visit and two IBD prescriptions. The base case incidence ranged between 14.1 and 16.0 per 100,000 person-years for CD and between 24.7 and 28.4 per 100,000 person-years for UC patients in the years 2010 to 2017. The incidence was stable over time and suggests that Norway has among the highest incidence of IBD in the world. The incidence of the more common definition of two IBD hospital visits was lower for both UC and CD in 2017 (Figure 5 & Figure 6). However, it was higher in the preceding years.

In 2017, the base case prevalence was 0.27% for CD and 0.50% for UC (Figure 7 & Figure 8). The prevalence of CD and UC increased linearly from 2010-17. Nevertheless, we cannot conclude that there was an increase in prevalence in Norway due to the relatively short data inclusion period.
Figure 5 CD incidence in 2017 according to different case definitions. *Base case definition.

Figure 6 UC incidence in 2017 according to different case definitions. *Base case definition.
Figure 7 CD prevalence in 2017 according to different case definitions. Actively treated was defined as one or more IBD visit(s) in 2017 and at least one prescription of IBD-related drugs in 2017. *Base case definition.

Figure 8 UC prevalence in 2017 according to different case definitions. Actively treated was defined as one or more IBD visit(s) in 2017 and at least one prescription of IBD-related drugs in 2017. *Base case definition.
5.3 Paper III

Paper III uses a novel method to estimate the incidence of CD patients utilizing hospital data with a limited data period. The study quantifies the impact of limited washout periods and look-ahead bias and introduces a method to adjust for these limitations. When we required an incident patient to have at least two records of IBD diagnosis in the NPR, not taking washout period and look-ahead bias into account, the incidence was reduced from 30.1 per 100,000 person-years in 2009 to 13.8 in 2017. The numbers adjusted for washout and look-ahead biases showed an increase from 12.1 per 100,000 person-years in 2009 to 15.4 per 100,000 person-years in 2017 (Figure 9).

Figure 9 Estimated incidence of CD patients with two hospital events before and after adjusting for both washout and look-ahead biases.

The figure gives an example of why the incident definition of two hospital diagnoses is not optimal when it does not adjust for differences in the length of the observed time period before and after the patient was first observed in the data. After adjusting the incidence estimate, the study indicates that there has been a small increase in incidence over time, instead of reporting a declining incidence.
6 Discussion of main findings

6.1 Paper I

Paper I estimated the regional variation of surgery and anti-TNF exposure and concluded that there were regional treatment differences. The impact of biologics and surgery is still a topic of discussion in the literature on the treatment of IBD patients\textsuperscript{100,101}. After the introduction of biologics gastroenterologists have an additional treatment option before deciding whether the patient should undergo surgery or not. There has been debates on whether biologics have the potential to reduce the need for surgery in IBD patients, or if it just postpones the surgery.

Although the paper shows a correlation, it should not be interpreted causally in the sense that biologics necessarily postpone the need for surgery, or reduces the lifelong rates of surgery for IBD patients. Multiple studies have addressed this question, and this paper contributes by establishing a pattern that future studies can use.

The results also suggest that there may be different professional opinions on which patients should receive biologic treatment. This becomes clear when looking at the proportion receiving biologics before and/or after surgery. In the North and South-East regions more patients received biologics prior to surgery, 41% and 46% for UC and 37% and 35% for CD. This suggests that the treating physicians were more lenient to try biologics on more IBD patients instead of having them undergo surgery. One interpretation is that the surgery rates are lower as biologics postpones surgery if the decision is made to give biologics instead of performing surgery. For patients who did not undergo surgery the anti-TNF exposure was more similar between the regions.

A nationwide cohort study from Denmark showed a decrease in major and minor surgery over time alongside an increase in the use of anti-TNFs\textsuperscript{102}. Also in this study, no causal connection between the two was established. Another Danish study found that the increased use of biologics did not reduce surgical rates, but prolonged the time from diagnosis to first time intestinal resection among patients with CD and UC\textsuperscript{103}. Other
studies have found no change in surgery rates in the biologic era, so the final verdict on this topic is still up\textsuperscript{104–106}. Biologics is expected to play a role in the decrease in surgery, but the decrease was ongoing prior to the introduction of biologics\textsuperscript{100}. The reason is likely to be multifactorial, with additional factors being less smoking, earlier diagnosis and treatment and improved clinical practice.

6.2 Paper II and III

6.2.1 Incidence

Paper II estimated the incidence and prevalence of CD and UC in Norway of different IBD case definitions. The base case definition defined an incident patient as someone with at least one IBD hospital visit and two IBD prescriptions or two IBD hospital visits. The base case incidence was 14.6 per 100,000 person-years for CD and 25.7 per 100,000 person-years for UC in 2017. Paper III estimated the incidence of CD to be 15.4 per 100,000 person-years. These results suggest that Norway has among the highest incidence and prevalence rates in the world.

The incidence of CD and UC has increased worldwide during the last decades\textsuperscript{12}. However, in North America and Europe it seems like the increase is halted, and in some countries even decreasing\textsuperscript{10}. By contrast, the incidence in newly industrialized countries started to increase around the shift to the 21\textsuperscript{st} century\textsuperscript{10}. In Europe, the incidence is characterised by an East-West and North-South gradient \textsuperscript{107,108}. Northern Europe and the United Kingdom have the highest incidence rates of IBD in Europe, while the lowest rates are seen in Eastern and Southern Europe\textsuperscript{109}. The highest reported incidence of IBD is in the Faroe Islands with 74 per 100,000 person-years (CD: 10, UC 44, IBDU: 20)\textsuperscript{110}.

In Norway, the first published studies on the incidence of IBD from the 1960’s reported an annual incidence of 0.26 per 100,000 for CD patients and 1.05 per 100,000 for UC patients\textsuperscript{111,112}. The mean annual incidence in Western Norway was 5.3 per 100,000 for CD and 14.8 per 100,000 for UC in 1984 and 1985\textsuperscript{113,114}. Similar rates were found in
The IBSEN study enrolled 843 patients with IBD in South-Eastern Norway from 1990 to 1994. The study reported an annual mean incidence of 5.8 per 100,000 for CD and 13.6 per 100,000 for UC (IBDU: 2.4)\textsuperscript{117,118}. This was the previous incidence estimate of IBD in Norway prior to Paper II, suggesting that the incidence has increased substantially in 25 years.

An important aspect is the effect of different definitions on the incidence over time. If we were to use the definition of at least two IBD diagnoses we would have concluded that there was a decreasing incidence in Norway. Instead the results from Paper II suggest that the incidence is stable, while Paper III suggests a slight increase. A potential explanation for the difference between the two is the uneven follow-up time for patients to receive another hospital visit or an IBD prescription in Paper II.

A UK study using a primary care database required at least two IBD diagnosis codes or one IBD diagnosis in addition to a prescription for a drug commonly used to treat IBD\textsuperscript{119}. To address the look-ahead bias they required a minimum of one year of follow-up. The minimum washout period was nine months. Even though the data source was from a primary care database and the results from this thesis does not directly apply, the exclusion periods seem short. They reported a decrease in incidence which was potentially caused by methodological limitations since patients can go extended periods of time without using IBD medication\textsuperscript{120}.

A decreasing trend of incidence has also been reported elsewhere\textsuperscript{121,122}. These two studies were commented on by Kaplan\textsuperscript{123}, where he questioned the true trend. The requirement of multiple IBD registrations makes incidence estimates susceptible to underestimation, especially in the latter years of a study. This, in combination with a too short washout period, increases the likelihood of concluding that there has been a decreasing incidence. A common problem with the validated algorithms for IBD is the fact that they are sensitive to the follow-up time. As an example, if a study shows a sensitivity of 80\%, this means that they would miss out on 20\% of the IBD patients. One
can expect that the number of false negatives (IBD patients not captured) are more prevalent in the later years. Validation studies are further discussed in section 7.4.

A Danish study required at least two IBD diagnoses for incident patients, and naturally found a reduction in the incidence in the last years of the study\(^\text{13}\). However, they acknowledge that the sharp drop in incidence was due to methodological limitations and concluded that the incidence was increasing. A systematic review of population-based studies reported that 73% of CD and 83% of UC studies estimated decreasing or stable incidence in Europe and North America\(^\text{10}\). In conclusion, it can be difficult to determine if the trend reported in incidence studies reflect the true trend due to limitations of administrative health data.

### 6.2.2 Prevalence

It is estimated that 1.5-3 million people in Europe have IBD, giving a prevalence of 0.2-0.4\%\(^\text{24,124}\). The prevalence in Europe ranges between 0.0015\% and 0.32\% for CD and 0.0024\% and 0.51\% for UC\(^\text{24,124}\). One of the highest prevalence estimates has been reported from Norway in 2009, 0.26\% for CD and 0.51\%\(^3\). The estimates are very close to the base case estimates of 0.27\% for CD and 0.50\% for UC from Paper II. For patients being actively treated (patients with an IBD visit and IBD prescription in 2017), the prevalence was 0.15\% for CD and 0.21\% for UC (Figure 7 & Figure 8). With stable incidence, early age of onset, low mortality and increased lifespan, the prevalence of IBD will increase\(^8\).

A study using data from the Swedish National Patient Registry reported an IBD prevalence of 0.65\% in 2010 (requiring two IBD visits)\(^\text{125}\). The previous prevalence estimates from Denmark were 0.15\% for CD and 0.29\% for UC in 2002\(^\text{126}\).
7 Discussion of methods

7.1 Observational versus interventional studies

In clinical evidence generation, there are two broad categories of study design: interventional and observational studies. A non-exhaustive list of observational studies include case-control, case-crossover, cross-sectional, prospective and retrospective studies. A common interventional study design is the randomised control trial (RCT), which has been the gold standard for clinical evidence generation for decades\textsuperscript{127,128}. However, they are expensive, take many years to complete and may not last long enough to assess the long-term effects. Observational registry based studies are increasingly used as more information is routinely collected from health care facilities\textsuperscript{129}. Since the data already exists, data collection is faster and less costly. Moreover, as observational registry studies do not alter patient care, you typically do not need informed consent if you have an approval from an ethics committee that ensures data protection. Still, the data collection is not necessarily a quick process due to bureaucratic constraints. When linking data with the NPR it is not uncommon to wait a year or two.

The Scandinavian countries have nationwide coverage with near complete coverage of all health care utilization. Patients are followed until death or emigration, which minimise patients lost to follow-up. Data over extended periods of time in combination with a nationwide population provide a good basis for high external validity, i.e. generalizable results. The use of nationwide registries allows for the inclusion of the entire population, compared to other studies where study participants are recruited from clinics and/or hospitals. This has the potential to reduce selection bias, but the use of administrative databases also poses some challenges in this regard. Since the registries are not primarily made for research, they are at an increased risk of misclassification and missing data. Misclassification bias occurs when patients are assigned to a different category than they should, e.g. IBD instead of non-IBD or incident IBD instead of prevalent IBD. The correct classification of an IBD-patient is a common challenge when using administrative databases and is further explored in Chapters 7.2-7.4\textsuperscript{130}.  

37
Participants in RCTs are randomly assigned to an intervention group or a control group. It is random in order to ensure equal characteristics of the groups and to minimize biases and confounding. However, there are generally differences between participants in a clinical trial and the real-world users, e.g. age, comorbidity, previous treatment and disease severity\textsuperscript{131}. If the RCT population differ from the general population, the results from the RCT will not necessarily replicate\textsuperscript{132,133}. This is because RCTs typically have more stringent inclusion criteria and several exclusion criteria. Moreover, when clinical effectiveness is measured, it is usually based on a rigorous dosing interval, e.g. 8 weeks interval for infliximab maintenance treatment. The adherence of patients in a natural setting do not necessarily follow the same dosing regimen as the RCT\textsuperscript{134,135}. Nationwide registries enable analyses of a broader range of patients and how they are treated in everyday clinical practice\textsuperscript{136}. Registry based studies can be particularly useful when other study designs are unethical, for instance when assessing the risk of fetal death after vaccination against influenza during pregnancy or whether abortions increase breast cancer risk\textsuperscript{137,138}. Furthermore, large populations with coverage over long periods allow for studies of rare exposures, diseases and outcomes to find small differences.

The range of information available from different health registries in Norway is vast, but it does not necessarily provide all the required confounders. For IBD, the registries lack some relevant variables which would have been readily available in a RCT (e.g. disease severity and phenotype) and other confounding variables (e.g. smoking and weight). The lack of potential confounder information combined with a large study population should be navigated carefully\textsuperscript{139}. Observational registry based studies are complementary to RCTs as some research questions are better answered by one method compared to the other\textsuperscript{140}. They enable innovative and efficient ways to answer important research questions. However, there are certain challenges for the use and interpretations of these data. Most methodologies in clinical research have limitations, and the limitations of administrative health data should not discourage its use. It is important to be aware of the strengths and limitations of the data in order to employ a valid and effective study design.
7.2 Change of IBD subtype and IBD unclassified

A single reference standard for the diagnosis of CD or UC does not exist\textsuperscript{141}. The diagnosis is based on a combination of clinical, endoscopic, imaging, and histologic findings\textsuperscript{5,6}. Diagnostic assessment guidelines recommend ileocolonoscopy with a minimum of two biopsies from the inflamed regions for a reliable diagnosis of UC and CD\textsuperscript{141,142}.

Even with careful diagnostics, it is not always possible to make a certain IBD diagnosis, nor a correct distinction between the IBD subtypes. IBD patients are categorised as IBD unclassified (IBDU) if endoscopic and histological assessments or other features cannot distinguish between CD and UC\textsuperscript{143}. The term indeterminate colitis (IC) has also been used interchangeably with IBDU, but the term IC is now reserved for patients who cannot be classified due to overlapping features of CD and UC after reviewing the histology of surgical specimens\textsuperscript{31,144,145}. In addition, terms as “possible IBD” and “uncertain colitis” are also used, making comparisons between unclassified IBD more difficult\textsuperscript{143,146}.

In the IBSEN study 5% received an IC/IBDU diagnosis\textsuperscript{146}. Five years after the initial IC diagnosis, 43% were diagnosed with UC, 13% with CD and 23% as non-IBD, while the remainder died or were lost to follow-up. In the same study, 3% of the initial UC patients were reclassified as CD, and 3% of the initial CD patients were reclassified to UC after five years\textsuperscript{146,147}. In a more recent European study, 9% were initially diagnosed with IBDU, of whom 18% changed to UC and 7% to CD, while 75% remained IBDU patients after five years\textsuperscript{148}. In other studies, 23–84% of the initial IBDU patients were reclassified as CD or UC\textsuperscript{149}.

A Swedish national registry study with more than 44,000 patients assessed the change in IBD subtype from 2002–2014\textsuperscript{149}. In total, 18% changed diagnosis during a median follow-up of nearly four years. At end of follow-up, 11% of adults were classified as IBDU (18% in children).
7.3 UC vs CD

To determine whether a patient should be classified as a CD or UC patient in the analysis, we used the last observed IBD diagnosis (the main diagnosis if both UC and CD was present at the last observation). Other studies have used the first diagnosis, the first two diagnosis, the last diagnosis, the majority of the last observations, the most recent diagnosis if they had only UC or CD in the last 5 years or a scoring based on the number of different diagnoses received and whether the source was hospital or physician database. Many of these studies also classify patients who had a diagnostic or procedure code typical of CD as CD-patients (e.g. small bowel resection).

The discrepancy of the different definitions demonstrates the uncertainty and complexity of the problem at hand. One should be aware of the limitations and consequences of the methodological choices and adjust the methods based on the data available and the research questions at hand since the optimal case definition depends on the research question and length of follow-up (e.g. differentiating between UC and CD using the last diagnosis can be imprecise when you want to assess the association between IBD subtype and future outcomes).

In Paper I we conducted a sensitivity analysis of the surgery and anti-TNF use when patients with both CD and UC was excluded (Figure A 1 & Figure A 2). The regional relationship were largely unaffected by the exclusion of the CD/UC patients, but resulted in a lower proportion of both CD and UC patients receiving anti-TNFs, in addition to a slight decrease in surgery for UC patients. A possible reason for this can be that the patients with more severe disease and complications have more hospital contacts within different specialities which increase the probability for coding differences and potential errors.

In paper II we also addressed the challenge of patients being registered with both CD and UC (14.6% were registered with both in NPR using our base case definition). The reasons for receiving both diagnoses are multiple: IBDU, clinical follow-up with investigations resulting in change from UC to CD, misdiagnosis or coding error. We did
not include them as a separate group since we found that nearly three times more of the patients had both CD and UC if they were first observed in 2010 compared to 2017\textsuperscript{156}. Since the probability of being registered with both CD and UC increased with more follow-up time, we did not want to give the impression that the patients registered with both CD and UC were one coherent group. If we were to exclude them, it would underestimate the true incidence of CD and UC. Moreover, it would bias the trend towards an increase in CD and UC incidence, as more patients would be excluded from the earlier year’s estimates.

7.4 Definition of an incident IBD patient

In Paper I we had data from NPR for the years 2008 to 2015. When we received the new data set, it included information from NorPD from 2004 to April 2018 and NPR 2008-2017. As stated in Paper I: Two or more diagnoses were required based on a recent Swedish validation study which reported a positive predictive value (PPV) for correct IBD diagnosis of 93\% using this definition\textsuperscript{157}. The PPV for IBD of 93\% is satisfactory, however it is important to note that the PPV relates to whether or not the patient has IBD and does not tell anything about the time of diagnosis. Moreover, the patients in the study were randomly selected from the year 1987 and forward so it is likely that most of these patients had a long follow-up to accrue visit number two. Thus, considering the data available in our study, the definition of an incident patient requiring two hospital visits did not appear optimal.

This prompted us to investigate other options, alongside the consequences of different approaches. First, we excluded everyone who had an IBD prescription prior to their first hospital visit. When looking at the time from prescription to first hospital visit it was a substantial amount of patients who received prescriptions shortly before their first hospital visit with IBD (Figure A 3 & Figure A 4). After looking at the visits prior to the first IBD visit, it became apparent that many of the patients received other ICD-codes (K55-K64) which is for “other diseases of intestines” (Table A 1). So we concluded that these patients most likely were examined for differential diagnoses to IBD and upon further investigations it became evident that they had UC or CD. The inclusion window
prior to the first hospital visit was set to 60 days, with a sensitivity analysis to show the effect of varying the window from between 0 and 180 days (Figure A 3 & Figure A 4).

There was a reduction in incidence when removing patients with at least two IBD diagnoses who had an IBD prescription 60 days or more prior to their first IBD-visit (Figure A 5). There is a decreasing trend, which is to be expected as the proportion of true incident patient’s increase with longer washout periods. Still, there is a substantial amount of patients who receive at least one IBD prescription prior to their first IBD hospital visit. This suggests that patients can get treatment outside of specialist healthcare since they receive IBD prescriptions without being registered at the hospital with IBD. This results in an overestimation of incidence if only hospital data is available, especially in the years with shorter washout period.

A Danish nationwide registry study reported that the median time between a patient’s first and second IBD visit was 430 and 654 days for CD and UC. This effect pulls towards an underestimation as there is not enough follow-up time for the patients to have their second IBD visits. Compared with the base case definition from Paper II, the requirement of two IBD visits resulted in higher incidence in the years 2010-2016, but not in 2017. The base case definition utilizes prescription data, for which we had data until April 2018. From NPR we had data until 2017. Seeing as most patients receive treatment shortly after their initial diagnosis, it expected that the base case definition resulted in a higher incidence (Figure A 6 Cumulative probability of receiving an IBD prescription).

7.5 Regional differences

Based on the results from Paper II & III, it is safe to assume that some prevalent patients were included as incident due to the relatively short washout period in Paper I. Nevertheless, the impact of the incident definition is not expected to have a substantial impact on the differences between the regions. This is because there is no reason to believe that the prevalence-incidence mix would differ between the four health regions.
Patients with moderate to severe IBD is more likely to fulfil the requirement of two IBD-visits within three years because of the need for hospital visits, especially the ones starting anti-TNF treatment or undergoing surgery. Conversely, considering the relative short washout period, milder prevalent patients might have been included as incident patients since it can go years between hospital visits. If we were to compare rates over time, e.g., 2010 vs 2012, the different prevalence-incidence mix between the years would have been an issue (e.g., more prevalent patients included in 2010 compared to 2012). Since we compared health regions for the years 2010-2012 as a whole, the difference between the regions should not be largely affected. However, one cannot be certain about the rates of surgery and anti-TNF use due to the mix of prevalent and incident patients.

7.6 Validation

A Swedish validation study reported a PPV of 93% for IBD with the definition of \( \geq 2 \) hospital visits\(^\text{157} \). Out of 146 patients, 17 were excluded since they only had one registration of IBD. Of these 17, eight did not fulfil any criteria for IBD, suggesting a lack of precision when requiring only one hospital registration. The requirement of two hospital registrations gave a high PPV for IBD, but when looking at CD and UC separately the PPV was lower (72% CD, 79% UC). The study also limited the analysis to patients with only CD or UC, which resulted in increased PPVs for CD (81%) and UC (90%), with a 92% in PPV for any IBD. The study did not report any other validity measure except PPV.

Another Swedish study aimed to validate algorithms for identification of IBD subtypes\(^\text{158} \). The medical records from 1403 patients with IBD (CD: 854, UC: 519, IBDU: 30) were reviewed. The study used a prevalent definition where they used the ICD-codes within 5 years preceding the review of the medical records. Patients registered with both CD and UC were classified using an algorithm based on ICD and NCSP codes\(^\text{17} \). The study only included patients treated with biologics which is likely to have overestimated the
overall accuracy since patients receiving biologics are at the hospital more often. In the Swedish National Patient Registry, 784 patients were assigned as CD, 405 as UC and 212 as IBDU. This resulted in a PPV of 97% for CD, 98% for UC and 8% for IBDU. The PPV was very high for CD and UC since most patients who were classified incorrectly ended up as IBDU. As a consequence, the sensitivity was lower than the PPV; 90% for CD and 77% for UC. A potential reason for UC having lower sensitivity than CD can be that the classification algorithm to differentiate between CD and UC included more codes specific to CD compared to UC. This highlights the value of not only looking at PPV when assessing the validity and reliability of a patient registry.

A Danish study reported the validity of two case definitions; at least one or at least two records of CD/UC in Danish National Patient Registry\textsuperscript{159}. For CD, sensitivity, specificity, negative predictive value (NPV) and PPV were all above 90% for both at least one and at least two diagnosis. For UC, the mentioned validity measures ranged between 77\% and 94\% when at least one diagnosis was required, while it was between 76\% and 89\% for at least two diagnoses. The numbers are interesting as they touch upon an important consequence of different case definitions. If one were to require at least one diagnosis one would include more IBD patients, but also more patients without IBD. When one increases the number of IBD diagnosis required the sensitivity will go down, and the specificity up. For UC, the sensitivity was 94\% with one diagnosis and 81\% when you require at least two diagnosis. Even with a median follow-up of 10 years, 13\% of the patients did not have a second hospital visits. This coincides with a point made in Paper II; that IBD patients go longer periods without having hospital contacts.

A validation study from a Canadian province found high validity and reliability of a case definition of at least two hospitalizations or four physician claims or two outpatient visits with IBD within two years (specificity 99.8\%; sensitivity 83.4\%; PPV 97.4\%; NPV 98.5\%\textsuperscript{160}). Another Canadian study tested more than 5,000 algorithms and found that the most accurate case definition was five physician contacts or hospitalizations within 4 years for 18 to 64 year-olds at diagnosis\textsuperscript{155}. They only had data on pharmacy claims for patient’s ≥65 years at diagnosis, but requiring an IBD-related medication improved accuracy of the definition. In the same study, less than 5\% of IBD patients had more
than 8 years between consecutive visits and they concluded that 8 years was a sufficient washout period.

Everhov\textsuperscript{149} used the case definition and subclassification (majority of last nine diagnoses) from Canada and found that only 38\% of the Swedish IBD population could be correctly classified. The main reason for this should be the fact that the Swedish study did not include primary care so many patients would not accrue the five visits within 4 years. It also underlines that it is difficult to find a one-size-fits-all when it comes to the optimal case definition of an IBD patient from administrative databases.

The validation studies from Denmark\textsuperscript{159} and Sweden\textsuperscript{157,158} suggests that the requirement of at least two IBD visits leads to high validity estimates. Even though one can assume that the patient registries are similar in the Scandinavian countries, the Swedish and Danish registries have longer data periods which reduces the washout problem.

We believe to have improved the case definition of an incident IBD patient in Norway, but we cannot conclude that this is an optimal definition. Most validation studies have not included prescription information, which we believe holds great value when identifying incident IBD patients. However, without a validation study, it is not possible to define the optimal case definition for a registry study. Even if there was a validation study, the validity of different case definitions depends on the number of years with data, what sort of data, registration practice and health care system. Ideally, the results should be updated regularly since treatment and registration practice can change, which will affect the validity of results. Still, while we wait for a validation study in Norway, the different case definitions in Paper II give suggestions of the true incidence and prevalence ranges.
8 Concluding Remarks

The central topic of this thesis is the use of nationwide registry data and IBD with special focus on incidence, prevalence and regional treatment variation. It shows that IBD patients receive different treatment dependent on their geographical location. Furthermore, it provides more up-to-date estimates of incidence and prevalence in Norway, raising questions about the previously held view that an incident IBD patient should be defined as two IBD hospital visits. Lastly, we proposed a novel method to estimate CD incidence based on hospital data.

This thesis include the first studies to utilize data from NPR and/or NorPD to estimate the incidence and prevalence of IBD. The results from this thesis are relevant for IBD researchers and physicians treating IBD patients. The results of Paper II is of special importance for future register-based IBD research as they provide insights and potential solutions to common challenges faced when working with registry data. The consequences of washout period and look-ahead bias are relevant for most diseases and should be thoroughly investigated, alongside the number and type of contacts required. This thesis’s contributions enable future register-based research on IBD to be based on more informed methodological choices with an increased awareness of the strengths and limitations of two of Norway’s largest health registries.
9 Appendix

Figure A 1 Cumulative incidence of anti-TNF exposure by diagnosis and region. Patients registered with both CD and UC was removed.

Figure A 2 Cumulative incidence of first major surgery by diagnosis and region. Patients registered with both CD and UC was removed.
Figure A 3 Change in CD incidence when varying the inclusion window of IBD prescriptions. In the base case definition patients were included if they had an IBD prescription two months or less prior to their first IBD visit.

Figure A 4 Change in UC incidence when varying the inclusion window of IBD prescriptions. In the base case definition patients were included if they had an IBD prescription two months or less prior to their first IBD visit.
Figure A 5 Annual reduction in incidence when removing patients with at least two IBD diagnosis who had an IBD prescription 60 days or more prior to their first IBD-visit.

Figure A 6 Cumulative probability of receiving an IBD prescription
<table>
<thead>
<tr>
<th>ICD-code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>K55</td>
<td>Vascular disorders of intestine</td>
</tr>
<tr>
<td>K56</td>
<td>Paralytic ileus and intestinal obstruction without hernia</td>
</tr>
<tr>
<td>K57</td>
<td>Diverticular disease of intestine</td>
</tr>
<tr>
<td>K58</td>
<td>Irritable bowel syndrome</td>
</tr>
<tr>
<td>K59</td>
<td>Other functional intestinal disorders</td>
</tr>
<tr>
<td>K60</td>
<td>Fissure and fistula of anal and rectal regions</td>
</tr>
<tr>
<td>K61</td>
<td>Abscess of anal and rectal regions</td>
</tr>
<tr>
<td>K62</td>
<td>Other diseases of anus and rectum</td>
</tr>
<tr>
<td>K63</td>
<td>Other diseases of intestine</td>
</tr>
<tr>
<td>K64</td>
<td>Hemorrhoids and perianal venous thrombosis</td>
</tr>
</tbody>
</table>
10 References


54


94. Selfors M. Årsrapport 2018, Norsk pasientregister. Published online 2019:50.


120. Melesse DY, Targownik LE, Singh H, Blanchard JF, Bernstein CN. Patterns and Predictors of Long-term Nonuse of Medical Therapy Among Persons with Inflammatory Bowel Disease. Inflamm Bowel Dis. 2015;21(7):1615-1622. doi:10.1097/MIB.0000000000000418


159. 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
Regional differences in anti-TNF-α therapy and surgery in the treatment of inflammatory bowel disease patients: a Norwegian nationwide cohort study

Sandre Svatun Lirhus, Marte Lie Høivik, Bjørn Moun, and Hans Olav Melberg

Institute of Health and Society, University of Oslo, Oslo, Norway; Institute of Clinical Medicine, University of Oslo, Oslo, Norway

ABSTRACT

Background and aims: During the last decades, substantial progress has been made in both medical and surgical treatment of inflammatory bowel disease (IBD). The aim of this study was to determine the use of anti-TNFs and surgery during the first 3 years after diagnosis in IBD patients across the four health regions in Norway using nationwide patient registry data.

Methods: This study used nationwide data from the Norwegian Patient Registry. Cumulative incidence of anti-TNF exposure and major surgery was calculated for patients diagnosed in 2010–2012. The analyses were stratified by diagnosis and health region. All patients were followed for an equal period of 3 years from diagnosis.

Results: The study population included 8,257 IBD patients first registered between 2010 and 2012, of whom 2,829 were diagnosed with Crohn’s disease (CD) and 5,428 with ulcerative colitis (UC). Across Norway’s health regions, the cumulative incidence of major surgery after 3 years varied from 11.4% to 20.9% for CD and from 4.6% to 6.9% for UC. The cumulative incidence of anti-TNF exposure varied from 20.9% to 31.4% for CD and from 8.0% to 13.5% for UC. The region with the lowest anti-TNF use had the highest surgery rates for both UC and CD.

Conclusions: Cumulative incidence of anti-TNF exposure and surgery varied significantly across Norway’s health regions during the three first years after IBD diagnosis.

Introduction

During the last two decades, several new drugs, including anti-TNF-α agents (anti-TNFs) and anti-integrins, have been introduced for treatment of inflammatory bowel disease (IBD). Biologic therapies have changed the management of IBD through their ability to induce and maintain remission [1–3]. However, the effect on the long-term disease course is not sufficiently clarified and remains an active research area [4–6]. A substantial proportion of patients do not achieve remission with biologics; others lose response or experience adverse reactions [1,7,8]. Failure of medical therapy is the most common indication for surgery, but surgery can also be an option without having failed medical therapy [9–11].

Research conducted on variations in the use of healthcare highlights that considerable variation is unwarranted as it is not explained by illness or patient preference [12,13]. Treatment of IBD patients can be classified as ‘preference-sensitive care’ when more than one generally accepted treatment option is available, such as elective surgery or medical therapy for patients with moderate to severe IBD. Variations in healthcare are found across geographic regions, institutions and even among individual physicians within single institutions [14]. Practice variation is important because of its potential impact on cost and outcomes, but there has been limited focus on such variation in the treatment of IBD [15].

Previous studies have described treatment patterns based on nationwide registries in other countries [16–18], but there is very limited research on intra-country variation.

The aim of this study was to determine the use of anti-TNFs and surgery during the first 3 years after diagnosis in IBD patients across the four health regions in Norway using nationwide patient registry data.

Materials and methods

Data source

In Norway, all citizens have near free access to a tax-supported healthcare system. Norway’s specialist healthcare system consists of four regional health authorities (Central, West, South-East and North) which are state enterprises responsible for specialist healthcare in the different health regions (Figure 1). All inpatient and outpatient hospital contacts are registered in the Norwegian Patient Registry (NPR) and it is mandatory to report diagnoses and clinical procedures. In addition, all biological drugs (both intravenous and subcutaneous preparations) can only be prescribed at hospitals and are registered in the NPR by their ATC codes. The registry dates back to 1997, and unique personal identification numbers were added in 2008, which made it possible to follow individual patients over time.
Defining the cohort

The dataset in this study included every inpatient and outpatient hospital event for all patients who received an IBD diagnosis (ICD-10 code K50 or K51) at least once between 2008 and 2015. Patients first observed with an IBD diagnosis in 2008 and 2009 were excluded to reduce the risk of misclassifying prevalent IBD cases as incident IBD. Patients first observed between 2013 and 2015 were not included since they had less than 3 years of observational information. The study cohort was defined as all patients with at least two registered K50 or K51 events during their 3-year follow-up. Two or more diagnoses were required based on a recent Swedish validation study which reported a positive predictive value for correct IBD diagnosis of 93% using this definition [19]. The date of diagnosis was set to the earliest record of an IBD diagnosis in the registry. For individuals who received both ulcerative colitis (UC) and Crohn’s disease (CD) diagnoses the last registered diagnosis was used as the main diagnosis. Patients with registered events in multiple health regions during the follow-up time were excluded in order to enable comparison between the health regions (Figure 2).

Anti-TNF therapy

The drugs included in the analysis were the TNF-α inhibitors infliximab, adalimumab and golimumab. Other biologics were not included as they were not registered as IBD treatment in Norway until after 2015. All patients with at least one registered event of anti-TNF use after their first IBD diagnosis were considered anti-TNF recipients (anti-TNF exposure).

Surgery

The Norwegian patient registry collects all surgical procedure codes for each individual hospital contact. The NOMESCO Classification of Surgical Procedures (NCSP) was used to define major surgery, which included resections, colectomies, strictureplasty and intestinal obstruction repair (Supporting Information Table S1).

Statistical analysis

Cumulative incidence of ever use of anti-TNFs and major surgery was estimated by constructing time-to-event curves (1 – Kaplan-Meier). All patients were followed for an equal period (3 years) and no censoring occurred as the data did not include time of death. The cumulative incidences were compared between regions using the log-rank test. The Chi-square test was used to test for differences in proportions of patients receiving anti-TNFs before and after surgery, and for patients not receiving surgery.

Ethical considerations

The study was approved by the Norwegian Patient Registry, the Norwegian Data Protection Authority and the Regional Committees for Medical and Health Research Ethics.
### Results

#### Study population

The study population included 8,257 IBD patients first registered between 2010 and 2012, of whom 2,829 (34.3%) were diagnosed with CD and 5,428 (65.7%) with UC (Figure 2). Crohn’s patients had a median age of 42 years at diagnosis and UC patients had a median age of 47 years. The analyses were stratified by diagnosis and health region (Table 1).

#### Surgery

**Crohn’s disease**

The cumulative incidence of major surgery in Norway was 13.8% after 3 years. Across the four health regions, the proportion receiving surgery varied between 11.4% (North) and 17.1% (Central) after 3 years (p = .033). The South-East and West regions had cumulative incidences of 13.0% and 15.6% (Figure 3).

**Ulcerative colitis**

After 3 years, 5.2% of UC patients received surgery in Norway. The cumulative incidence of surgery varied from 4.6% (South-East) to 6.9% (Central) between the health regions (p = .011). The North and West regions had cumulative incidences of 5.2% and 5.5% (Figure 3).

#### Anti-TNFs therapy

**Crohn’s disease**

Among the CD patients, 28.8% received at least one anti-TNF 3 years after diagnosis. The anti-TNF exposure in each of the four health regions ranged from 20.9% (Central) to 31.4% (South-East) (p = .001). The North and West regions had cumulative incidences of 26.1% and 28.9% (Figure 4). The Central region had administered anti-TNFs to 11.5% of the CD patients after 1 year, which was approximately 50% less compared to the other regions (20.1%–21.3%).

Of the 391 CD patients who underwent surgery, 173 (44.3%) had received anti-TNFs sometime during the follow-up period. The proportion of patients who received anti-TNFs prior to surgery varied from 11.3% to 36.8% across the health regions. After surgery between 15.5% and 36.8% received anti-TNFs. For patients who did not receive surgery, the proportion receiving anti-TNFs varied from 20.9% to 28.5% (Figure 5).

**Ulcerative colitis**

Among the UC patients, 11.8% received at least one anti-TNF after 3 years. The anti-TNF exposure in each of the four health regions ranged from 8% (Central) to 13.5% (South-East) (p < .001). The West and North regions had cumulative incidences of 10% and 12.2% (Figure 4).

Of the 282 UC patients who underwent surgery, 107 (38%) had received anti-TNFs sometime during the follow-up period. The proportion of patients who received anti-TNFs prior to surgery varied from 18.6% to 46% across the health regions. After surgery, between 1.9% and 8.8% received anti-TNFs. In patients who did not receive surgery, the proportion receiving anti-TNFs ranged from 7.2% to 11.9% (Figure 6).

#### Discussion

In this nationwide registry study of 8,257 IBD patients, we found significant variations in both surgery and anti-TNF exposure across Norway’s health regions. The region with the lowest anti-TNF use had the highest surgery rates for both UC and CD. There were also significant differences in whether or not patients received anti-TNFs before or after surgery.

The management of IBD patients is challenging because of a complex disease spectrum and rapidly changing treatment options, which make it prone to unwarranted variation. Anti-TNFs have shown to reduce the need for surgery in the short term, but their effect in the long term is uncertain [20–22]. Biologic agents and surgery are complementary therapies, and consensus on the role of different treatments is still being developed. This uncertainty may create differences in beliefs and subsequent variations in treatment practice.

Large hospitals treat the majority of IBD patients in all the Norwegian health regions. This makes the treatment landscape more susceptible to the treatment preferences of key opinion leaders. Lower utilisation of anti-TNFs and more surgery might indicate that the Central region was more restrictive in giving the severe cases of IBD anti-TNFs and instead opted for surgery or other medical therapy. This hypothesis was further strengthened by the variation in the proportion of patients receiving anti-TNF prior to surgery. The higher the proportion of patients receiving anti-TNFs prior to surgery, the lower was the proportion of patients undergoing surgery. This was true for both UC and CD. The rankings between the regions with respect to the use of anti-TNFs prior to surgery and the cumulative incidence of surgery were exactly reversed, but the
margin of difference was small for some of the regions. In CD patients, the Central region administered anti-TNFs to half as many patients as the other regions within 1 year after diagnosis. The proportion of patients receiving anti-TNFs was more similar in patients’ not undergoing surgery compared with the population receiving surgery.

In a prospective follow-up study from Norway (the IBSEN Study) conducted before anti-TNFs were introduced, the...
surgery rates for UC were 3.5% after 1 year and 9.8% after 10 years [23]. For CD, the surgery rates were 13.6%, 27% and 37.9% after 1, 5 and 10 years [24]. The cumulative probabilities of surgery were lower in our study than in the IBSEN study in all regions. Given that the IBSEN cohort was diagnosed between 1990 and 1994 (in the prebiologic era), our results indicate a reduction in IBD surgery over time in Norway. Our results are in line with a nationwide cohort study from Denmark showing a decrease in major and minor surgery over time alongside an increase in the use of anti-TNFs, although a causal connection between the two could not be established [17]. Other studies have found no change in surgery rates in the post-biologic era [4,5].

The data did not include time of death, but since the follow-up period is short, it is unlikely that this affected the cumulative incidences. Furthermore, the IBSEN study found no significant evidence of excess mortality for IBD patients after 20 years [25,26]. As with any diagnosis, there is a risk of coding errors during hospital admission. However, when requiring two IBD diagnoses within 3 years, we believe the number of false positive IBD diagnoses to be very low. No known validation studies of the IBD diagnoses in the NPR exists. Registration of anti-TNFs should be accurate since hospital trusts are reimbursed for each registered ATC-code.

In conclusion, this study has shown that the cumulative incidence of anti-TNF exposure and surgery varied across Norway's health regions during the first 3 years after IBD diagnosis. The regional variations may lead to differences in health outcomes and costs. Further research is needed to analyse the consequences of the differences in treatment patterns, for instance its effect on hospitalisations and sick-leave.

Acknowledgements

Data from the Norwegian Patient Registry have been used in this publication. The interpretation and reporting of these data are the sole responsibility of the authors, and no endorsement by the Norwegian patient registry is intended nor should be inferred. The authors thank Tom Glomsaker for deciding which surgical codes to include.

Disclosure statement

SSL reports grants from Takeda, during the conduct of the study; personal fees from Takeda, outside the submitted work. HOM reports unrestricted research grants from Tillotts, Ferring and Takeda. Advisory board Takeda. Speaker fees from AbbVie, Jansen, Hospira, Orion Pharma, Pfizer, Sandoz, Vifor Pharma. Speaker fees: Takeda, Pfizer, Orion Pharma, Hospira, Ferring, Cosmo Pharma.

ORCID

Sande Svatun Lirhus http://orcid.org/0000-0002-0895-7654
Marte Lie Helvik http://orcid.org/0000-0002-0104-465X

References


Incidence and Prevalence of Inflammatory Bowel Disease in Norway and the Impact of Different Case Definitions: A Nationwide Registry Study

Sandre Svatun Lirhus 1
Marte Lie Høivik 2,3
Bjørn Moun 2,3
Karoline Anisdahi 2,3
Hans Olav Melberg 1

1Department of Health Management and Health Economics, University of Oslo, Oslo, Norway; 2Department of Gastroenterology, Oslo University Hospital, Oslo, Norway; 3Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Background: Countries have different diagnostic procedures and treatment regimens for inflammatory bowel disease (IBD) patients. In addition to differences in population characteristics, completeness of data and health registries, different follow-up time and case definitions can have a large impact on estimates of the incidence and prevalence of IBD.

Aim: The aim of this study was to use hospital and prescription data to estimate incidence and prevalence of Crohn’s disease (CD) and ulcerative colitis (UC), using different case definitions.

Methods: This study used nationwide data from the Norwegian Patient Registry (2008 to 2017) and the Norwegian Prescription Database (2004 to April 2018). Incidence and prevalence were estimated using different case definitions of an IBD patient, varying the number of IBD-related hospital visits and IBD prescriptions required. The base case definition included patients with at least one IBD hospital visit and two IBD prescriptions or two IBD hospital visits.

Results: From 2010 to 2017, 16,758 incident IBD patients fulfilled our base case definition, with 6045 diagnosed with CD (36.1%) and 10,713 (63.9%) with UC. For CD, 47.2% of the patients were male while 53.8% of UC patients were male. The base case incidence varied between 14.1 and 16.0 per 100,000 person-years for CD and 24.7 and 28.4/100,000 person-years for UC patients in the years 2010–2017. When we required at least two IBD hospital visits, not utilizing the prescription data, the CD incidence was 22.3 per 100,000 person-years in 2010 and 13.9 per 100,000 person-years in 2017. For UC, the incidence was 47.4 and 20.6 per 100,000 person-years in 2010 and 2017. In 2017, the prevalence of CD was 0.27% (95% CI: 0.26–0.27) and 0.50% (95% CI: 0.490–0.502) for UC.

Conclusion: According to our base case definition, the incidence of IBD in Norway was stable from 2010 to 2017. Both the incidence and prevalence of IBD in Norway is among the highest in the world. Moreover, the study also highlights the consequences of different case definitions.

Keywords: inflammatory bowel disease, incidence, prevalence, real-world data

Introduction
In Europe, the reported incidence of Crohn’s disease (CD) ranges from 0.5 to 10.6 cases per 100,000 person-years and for ulcerative colitis (UC) the numbers are 0.9 to 24.3 per 100,000 person-years.1 The prevalence in Europe varies from 1.5 to 213 cases per 100,000 persons for CD and 2.4 to 294 cases per 100,000 persons for UC.1 The highest incidence and prevalence estimates have been reported in Northern Europe. The incidence of inflammatory bowel disease (IBD) has been
reported to increase in the western world in the latter half of the 20th century.\textsuperscript{2,3} However, a recent systematic review of population-based studies from 1990 or later reported that 73% of CD and 83% of UC studies show stable or decreasing incidence in North America and Europe.\textsuperscript{4}

The diagnosis of IBD is based on a combination of symptoms, endoscopy, radiology and histological findings in biopsies.\textsuperscript{5,6} Countries have different diagnostic procedures and treatment regimes, making comparisons of studies on the incidence and prevalence of IBD challenging. In addition to differences in population characteristics, completeness of data and health registries, different follow-up time and case definitions can have large impact for the estimates of IBD incidence and prevalence.

There are numerous studies that have estimated the incidence and prevalence of IBD using population-based registries.\textsuperscript{1,4} Unlike Scandinavia, most countries do not have nationwide patient registries and data on IBD epidemiology is often derived from tertiary referral centers or insurance databases.

Potential problems related to limited follow-up and case definitions are addressed to a varying extent in the literature, but few studies have quantified the problem. Given this, the aim of this study was to use nationwide hospital and prescription data to estimate incidence and prevalence of IBD, using different case definitions.

**Materials and Methods**

**Data**

In Norway, all in- and outpatient hospital contacts are registered in the Norwegian Patient Registry (NPR), and it is mandatory to report diagnoses and clinical procedures. NPR dates back to 2008. The Norwegian Prescription Database (NorPD) includes information of all dispensed drug prescriptions by Norwegian pharmacies, including date (at pick-up time), and the Anatomical Therapeutic Chemical (ATC), International Classification of Diseases (ICD-10) and International Classification of Primary Care codes (ICPC). Data from NorPD and NPR were linked using the unique personal identification number of every Norwegian resident, which made it possible to follow individual patients over time. The source population included all patients who received an IBD diagnosis (ICD-10 code K50 (CD) or K51 (UC)) at least once between 2008 and 2017 in NPR. Data from NorPD was from 2004 to 30 April 2018. Patients only registered with an IBD diagnosis in NorPD were not included as the data extraction was based on an IBD diagnosis in NPR.

**Definition of an IBD Prescription**

ICD- and ICPC-codes were not included in NorPD before March 2008 and were fully implemented in March 2009.\textsuperscript{7} For 5-aminosalicylic acid (5-ASA) prescriptions with a diagnosis code, 98.7% had an ICD-code (ICD K50/51, ICPC D94). For budesonide prescriptions with a diagnosis code, 99.2% had an IBD-code. These medications were therefore included as IBD prescriptions also when they were missing both an ICD- and ICPC-code in order to exclude prevalent patients prior to March 2008/2009. Other IBD-related prescriptions were only included if they had an IBD diagnosis code. Consequently, our definition of an IBD prescription included all pharmacy claims with ICD-code K50 or K51, ICPC-code D94 and/or prescriptions of 5-ASA or budesonide.

**Case Definitions**

Incidence and prevalence were estimated using different case definitions of an IBD patient, varying the number of IBD diagnosis required in NPR and IBD prescriptions from NorPD. For the incidence estimates including IBD prescriptions, patients with an IBD prescription more than 60 days prior to their first NPR registration were excluded. For prevalence, patients were defined as actively treated if they had at least one IBD visit in 2017 and at least one prescription for 5-ASA, corticosteroids, immunomodulators or biologic agents in 2017.\textsuperscript{8}

If a patient was registered with both UC and CD, the last observed ICD-code in NPR determined whether the patient would be classified as CD or UC. Incident patients were included in the year of their first IBD hospital visit. Prevalent patients were included from the year of their first IBD visit or IBD prescription, whichever came first.

**Base Case Definition**

To avoid the inclusion of patients without IBD and a concurrent overestimation of incidence and prevalence, registry studies of IBD usually require two IBD-related hospital visits (ICD K50/51).\textsuperscript{9-11} When requiring two IBD visits, patients might be excluded if the follow-up time is too short to capture the second visit, and the problem increases the closer you get to the end of the study period. To deal with this fact, patients with only one IBD-related hospital event were accepted if there were two IBD prescriptions to compensate for only one hospital event.
To distinguish incident and prevalent cases, a lookback period without an IBD diagnosis was required. Consequently, patients with their first IBD visit between 2008 and 2009 were excluded from the incidence estimates. Patients with an IBD prescription (in NorPD where we had available data going back to 2004) more than 60 days prior to their first NPR registration were excluded from the incidence estimates. See Figures S1 and S2 to see the effect of varying the 60-day window.

In the base case, incident IBD was defined as at least one IBD visit and two IBD prescriptions or two IBD visits. IBD visits included all in- and outpatient hospital visits with IBD.

**Statistical Analysis**

Statistics Norway (SSB) presents population statistics as the number of inhabitants on January 1st of every year. We calculated prevalence counting all incident and prevalent IBD patients alive on December 31st of each year, divided by inhabitants in Norway on January 1st of the succeeding year, presented as point-prevalence proportion with 95% CI. Incidence was estimated adding all incident patients, divided by the mid-year population in the same year, presented as incidence per 100,000 person-years with 95% CI (exact Poisson confidence limits). The mid-year population was estimated taking the average of the population on January 1st of each year and January 1st of the succeeding year (eg the mean between 50-year-olds in 2011 and 51-year-olds in 2012).

Due to anonymization requirements from the health authorities, information on age was provided by birth cohort in 10-year intervals. Mean age is therefore not reported and age-specific incidence and prevalence was only calculated for the last available year (2017) since changes in age-specific incidence and prevalence over time could not be determined (Figures S3–S8). The respective age-intervals were chosen to ensure correct age groups. Data handling and analyses were performed using Python 3.8.

**Ethical Considerations**

The study was approved by the Norwegian Data Protection Authority and the Regional Committees for Medical and Health Research Ethics, NPR and NorPD.

**Results**

The source population included 51,488 IBD patients who had at least one IBD visit during 2008–2017.

**Incidence**

From 2010 to 2017, 16,758 incident IBD patients fulfilled our base case definition, with 6045 diagnosed with CD (36.1%) and 10,713 (63.9%) with UC. Some of the incident patients were diagnosed with both UC and CD (14.6%) in NPR and these were classified according to their last diagnosis. For CD, 47.2% of the patients were male while 53.8% of UC patients were male. The base case incidence varied between 14.1 and 16.0 per 100,000 person-years for CD and 24.7 and 28.4 per 100,000 person-years for UC patients in the years 2010 to 2017 (Figure 1).

When we required at least two IBD hospital visits, not excluding patients with an IBD prescription prior to their first IBD hospital visit, the CD incidence was 22.3 per 100,000 person-years in 2010 and 13.9 per 100,000 person-years in 2017. For UC, the incidence was 47.4 and 20.6 per 100,000 person-years in 2010 and 2017 (Figure 2). The incidence of IBD in Figure 2 was reduced when patients with at least two IBD visits were removed if they had received an IBD prescription more than 60 days prior to their first IBD visit (Figure 3).

When we required only one IBD visit, the incidence of CD increased from 14.6 (CI: 13.6–15.7) to 27.7 (CI: 24.3–27.1) per 100,000 person-years in 2017 (Figure 4). The incidence decreased to 13.9 (CI: 12.9–14.9) per 100,000 person-years when we required two IBD visits (Figure 4). With the strictest definition, requiring two IBD visits and one IBD prescription, the CD incidence was 10.1 (CI: 9.2–10.9) per 100,000 person-years. For UC, the base case incidence was 25.7 (CI: 24.3–27.1) per 100,000 person-years in 2017 (Figure 5). The incidence decreased to 20.6 (CI: 19.4–21.8) when we required two IBD visits. When we required two IBD visits and an IBD prescription, the incidence decreased to 16.4 (CI: 15.3–17.5).

The incidence of IBD for patients with a single IBD hospital visit and prescription can be seen in Figure S9. See Figure S10 for the incidence for patients with a single IBD hospital visit and no IBD prescriptions.

**Prevalence**

In 2017, 40,900 prevalent IBD patients fulfilled our base case definition, where 14,352 were diagnosed with CD (35.1%) and 26,548 (64.9%) with UC. For CD, 46.9% of the patients were male, while 52.4% of UC patients were male.
The prevalence of CD was 0.27% (95% CI: 0.26–0.27) and 0.50% (95% CI: 0.49–0.50) for UC (Figures 6–8). When at least two IBD visits were required, the prevalence of CD decreased to 0.25% (95% CI: 0.25–0.26). For CD patients being actively treated (defined as one or more IBD visits in 2017 and at least one prescription of IBD-related drugs in 2017) the prevalence was 0.15% (95% CI: 0.15–0.15). For UC, the prevalence decreased to 0.42% (95% CI: 0.42–0.43) when two IBD visits were required and for patients being actively treated the prevalence was 0.21% (95% CI: 0.21–0.22).

**Discussion**

This study provides the first Norwegian estimate of incidence and prevalence of IBD based on nationwide data. According to this study, both incidence and prevalence of IBD in Norway is among the highest in the world. In 2017, the incidence of IBD was 40 (CD: 14.6, UC: 25.7) per 100,000 person-years and the prevalence was 0.77% (CD: 0.27%, UC: 0.50%). The base case definition resulted in higher incidence and prevalence estimates in 2017 compared to the more commonly used definition of at least two hospital visits.
Figure 3 The incidence of IBD for patients with at least two records in the Norwegian Patient Registry. Patients were removed if they had received an IBD prescription more than 60 days prior to their first IBD visit.

Figure 4 CD incidence in 2017 according to different case definitions. *Base case definition.

Figure 5 UC incidence in 2017 according to different case definitions. *Base case definition.
Prevalence

The prevalence of CD and UC increased linearly from 2010 to 2017. However, we cannot conclude that there is an increasing prevalence in Norway due to the relatively short data inclusion period. Previous Norwegian estimates from 2009 were 0.26% for CD and 0.51% for UC. Despite the use of a different methodology, the estimate from 2009 is very close to this study’s base case estimation for 2017.

A Swedish study using National Patient Registry data from 1987 to 2010 reported an IBD prevalence of 0.65% (CD: 0.35%, UC: 0.19% and IBDU: 0.11%) in 2010 requiring two IBD visits. Incidence

A population-based prospective study from the early nineties including one-fourth of Norway’s population estimated the incidence of IBD to 19.3 (CD: 5.1, UC: 10.6 and IBDU: 3.6) per 100,000 population. This would suggest an increasing incidence of IBD over the last 25 years in Norway. However, there was no clear trend in incidence over the eight years in our study. An increase could have been masked by an underestimation of the incidence in the later years of the study due to the case definition requirement of a second IBD visit and/or two IBD prescriptions.
A Danish study using nationwide registry data from 1977 to 2013 reported an incidence of 9.1 per 100,000 person-years for CD and 18.6 per 100,000 for UC in the period of 2010–13. However, when breaking down the incidence per year, there was a substantial decrease in the incidence from 2010 to 2011 (CD: 11.6, UC: 26.3) to 2012–13 (CD: 6.6, UC: 10.7) since patients were required to have two IBD visits (visually estimated from Figure 1 of Lophaven et al.⁸). The incidence estimates from 2010 to 2011 suggest that the IBD incidence is comparable between Norway and Denmark. The reported median time between the first two recordings in the Danish study was 430 days (CD) and 654 days (UC) which emphasizes the probability of an underestimation of incidence in the later years of a registry study.

The optimal case definition depends on the data period, registration practice and health care system. A Canadian study tested more than 5000 algorithms and found the most accurate case definition to be five physician contacts or hospitalizations within 4 years (for patients 18–64 years at diagnosis). If we would have used this definition, we would not have been able to calculate the incidence for the last years of our study period. It also introduces problems with the date of diagnosis for patients who have a GP visit with an IBD diagnosis prior to their first hospital registration with IBD. A Swedish study reported a positive predictive value of 93% for any IBD requiring two IBD hospital visits. The criteria of at least two hospital visits would lead to an overestimation in the earlier years due to the short lookback period and underestimation in the later years. The overestimation in earlier years is apparent in Figure 2, demonstrating the value of including prescription data to classify incident patients. In addition to reducing the misclassification of prevalent cases as incident cases, the difference between Figures 2 and 3 suggests that patients get treatment outside specialist healthcare for longer periods as they receive IBD prescriptions without having hospital contacts.

The main strength of the study is the use of a complete and representative dataset for the whole population across many years, for both hospital events and prescription events. This means that the results are not subject to statistical sample uncertainty.

A limitation of the study is the handling of patients registered with both UC and CD which includes 14.6% of the base case incidence. We did not included them as a third separate IBD unclassified (IBD-U) group since that would remove nearly three times more patients in 2010 compared to 2017. This is because the probability of receiving both an UC and CD diagnosis increases with longer follow-up time. This indicates that there are multiple reasons for a patient to be registered with both UC and CD: IBD-U, wrong ICD-code registered due to human error, misdiagnosis or a change from UC to CD. Since the proportion of incident patients who are registered with UC and CD varies over time, we did not want to claim or give the impression that this was one coherent group (IBD-U) by including it as a unified category. Excluding the group would also be a mistake, since this would underestimate the incidence. We acknowledge that there is no perfect way to solve this issue with patients registered with both UC and CD.

Figure 8: UC prevalence in 2017 according to different case definitions. Actively treated was defined as one or more IBD visits in 2017 and at least one prescription of IBD-related drugs in 2017. *Base case definition.
In summary, our base case definition tried to include all incident IBD patients and avoid classifying prevalent patients as incident cases. It is difficult to assess the accuracy of the definition, but we believe that our base case definition is a step forward compared to the other estimations discussed in this study.

**Conclusion**

According to our base case definition, the incidence of IBD in Norway was stable from 2010 to 2017. Both the incidence and prevalence of IBD in Norway is among the highest in the world. Moreover, the study also highlights the consequences of different case definitions.

**Funding**

This work was financially supported by Takeda.

**Disclosure**

SSL reports research funding from Takeda during the conduct of the study and personal fees from Takeda outside the submitted work. HOM reports research funding from Takeda during the conduct of the study and personal fees from Takeda outside the submitted work. MLH reports unrestricted research grants from Tillotts, Ferring and Takeda; advisory board for Takeda; and speaker fees from AbbVie, Meda and Takeda. BM serves in the advisory board for AbbVie, Jansen, Hospira, Orion Pharma, Pfizer, Sandoz, Vifor Pharma; reports speaker fees from Takeda, Pfizer, Orion Pharma, Hospira, Ferring, and Cosmo Pharma. KA reports personal fees from Takeda. The authors report no other conflicts of interest in this work.

**References**


7. Accepted February 20, 2020.


