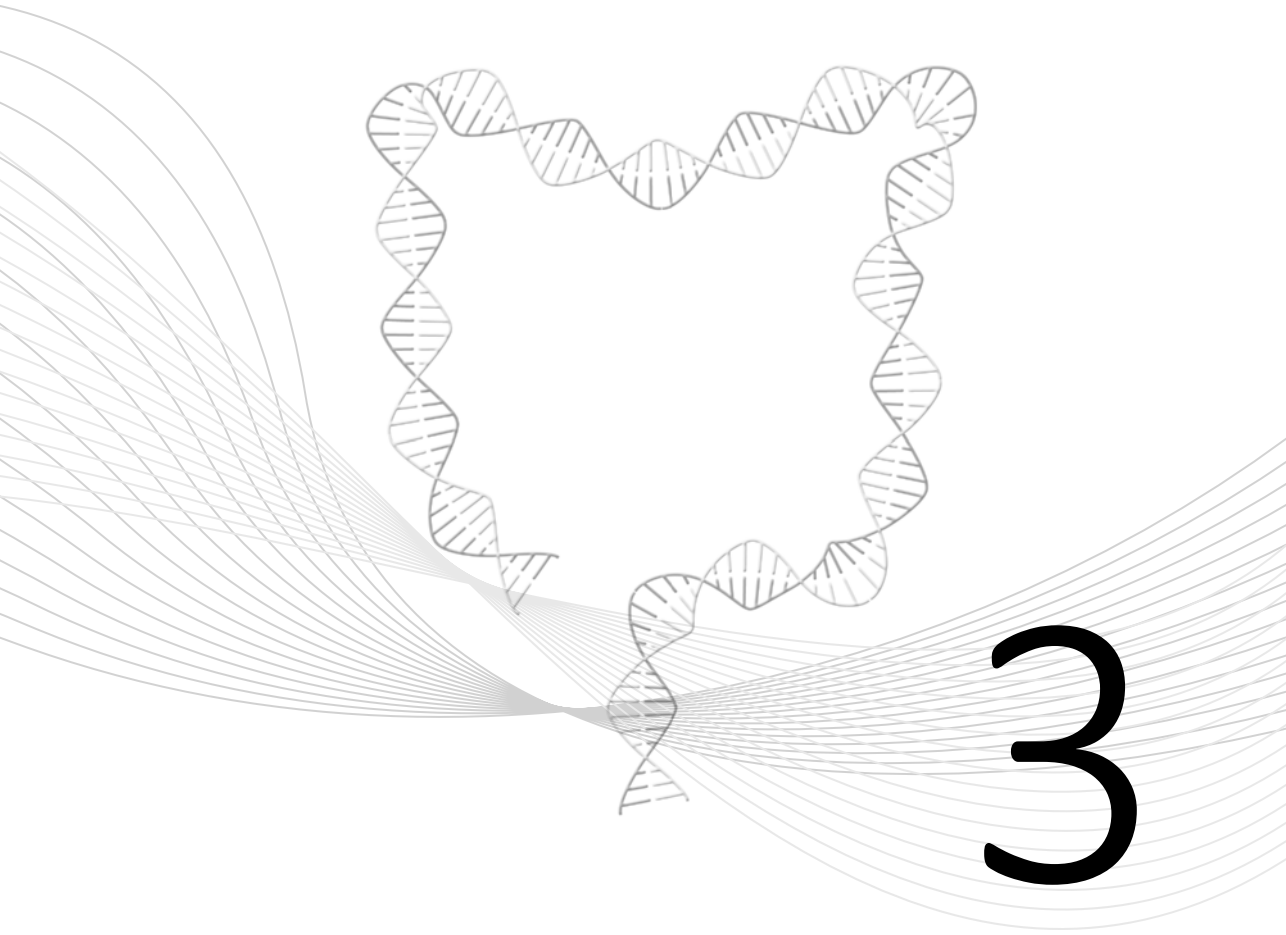
Clinical and genetic factors associated with disease course in inflammatory bowel disease

Spekhorst, Lieke Maaike



3

Cohort profile: design and first results of the Dutch IBD Biobank: a prospective, nationwide biobank of patients with inflammatory bowel disease

LM Spekhorst*, F Imhann*, EAM Festen, AA van Bodegraven, NKH de Boer, G Bouma, HH Fidder, G D'Haens, F Hoentjen, DW Hommes, DJ de Jong, M Löwenberg, PWJ Maljaars, AE van der Meulen-de Jong, B Oldenburg, MJ Pierik, CY Ponsioen, PC Stokkers, HW Verspaget, MC Visschedijk, CJ van der Woude, G Dijkstra, RK Weersma; on behalf of the Parelsnoer Institute (PSI) and the Dutch Initiative on Crohn and Colitis (ICC)

*Authors contributed equally

BMJ Open 2017 Nov 8;7(11)

Abstract

Purpose: The Dutch IBD Biobank aims to facilitate the discovery of predictors for individual disease course and treatment response in patients with inflammatory bowel disease (IBD). In this paper, we aim to describe the establishment of the Dutch IBD Biobank, including the facilitators and barriers to establishment. Moreover, we aim to provide a complete overview of the content of the Dutch IBD Biobank.

Participants: Since 2007, every patient with IBD treated in one of the eight Dutch university medical centres is asked to participate in the Dutch IBD Biobank in which 225 standardised IBD-related data items and biomaterials, such as serum, DNA, biopsies and a stool sample, are collected.

Findings to date: As of June 2014, the Dutch IBD Biobank had enrolled 3388 patients with IBD: 2118 Crohn's disease (62.5%), 1190 ulcerative colitis (35.1%), 74 IBD- unclassified (2.2%) and 6 IBD-indeterminate (0.2%). The inclusion of patients with IBD is ongoing. The quality of the biomaterials is good and serum, DNA and biopsies have been used in newly published studies.

Future plans: The genotyping (750,000 genetic variants) of all participants of the Dutch IBD Biobank is currently ongoing, enabling more genetic research. In addition, all participants will start reporting disease activity and outcome measures using an online platform and mobile app.

Introduction

Inflammatory bowel disease (IBD) is a chronic inflammatory disease of the gut comprising Crohn's disease (CD) and ulcerative colitis (UC). Of the 17 million inhabitants in the Netherlands, 39,000 individuals have been diagnosed with CD and 48,000 individuals with UC.[1] Approximately 39 new individuals per 100,000 are newly diagnosed with IBD every year. This incidence rate continues to rise, posing an increasing burden on society.[2] The clinical symptoms of IBD consist of diarrhoea, abdominal discomfort, weight loss, fatigue and rectal bleeding. However, these symptoms vary greatly both between individuals and in time. Some patients with IBD have a relatively mild disease course, requiring only limited therapeutic intervention, while others have a severe disease course with frequent flares requiring expensive medical and surgical interventions.

In recent years, many case–control studies have been performed to identify factors that can explain the *onset* of IBD. Genome-wide association studies (GWAS) have identified 200 genomic loci that are involved in the onset of IBD.[3] Epidemiological studies have identified environmental risk factors including smoking, appendectomy, infections, antibiotics, diet and lifestyle (stress, lack of sleep and/or exercise) that could trigger the onset of IBD.[4] Studies on the bacterial composition of the gut (the gut microbiota) have identified distinct microbial compositions associated with IBD.[5,6] Unfortunately, these studies provide little insight into reasons for the heterogeneous clinical presentation and *disease course* of patients with IBD. As a consequence, limited progress has been made in translating basic science into personalised treatment. Predicting individual disease outcome and tailoring IBD treatment requires prospective patient data on disease activity, complications and treatment, as well as biomaterials and *-omics* data (genome, transcriptome and gut microbiome), in order to link biomarkers to disease. To this aim, the prospective Dutch IBD Biobank was created. A new national institute to facilitate the biobank and other national biobanks was founded by the Dutch Federation of University Medical Centres (NFU) in 2007 and called the Parelsnoer Institute (PSI).[7] Gastroenterologists who specialised in treating patients with IBD in all eight Dutch university medical centres (UMC), together with a team of information architects and laboratory experts, built up the Dutch IBD Biobank.

The main objective of the biobank is to facilitate the discovery of predictors (both epidemiological risk factors and biomarkers) for individual disease course and treatment response, by:

1. providing full clinical records of patients describing their individual disease course over a prolonged period of time;
2. providing high-quality biomaterials;
3. standardising patient data collection and questionnaires during outpatient clinic visits and thereby improving clinical care.

The aim of this paper is to inform the IBD research community about the existence of the Dutch IBD Biobank and to give an elaborate overview of the establishment process as well as the content.

Cohort description

Design, participating centres and the Dutch healthcare setting

The Dutch IBD Biobank is a prospective, nationwide biobank in which both data and biomaterials are collected. In the Netherlands, there are approximately 80 hospitals and 8 UMCs (tertiary referral centres), where patients with complex IBD are referred to. All eight Dutch UMCs participate in the Dutch IBD Biobank. The Dutch UMCs are: the Amsterdam Medical Centre in Amsterdam, the Erasmus Medical Centre in Rotterdam, the Leiden University Medical Centre in Leiden (LUMC), the Maastricht University Medical Centre in Maastricht (MUMC), the Radboud University Nijmegen Medical Centre in Nijmegen, the University Medical Center Groningen in Groningen (UMCG), the University Medical Centre Utrecht in Utrecht and the VU (*Vrije Universiteit*) University Medical Centre in Amsterdam. PSI and the Dutch IBD Biobank are part of the Biobanking and Biomolecular Resources Research Infrastructure of the Netherlands (BBMRI-NL). This is the Dutch national node of BBMRI-ERIC, the largest research infrastructure project in Europe.[8]

Standardised data collection: the information model

Gastroenterologists from each of the eight UMCs convened to design the information model based on literature review and clinical standards. A working group of gastroenterologists made a longlist of data items including a definition for each data item. This longlist was subsequently discussed during a meeting in 2006, where one or more representatives from each Dutch UMC were present. Data items and definitions were accepted, modified if deemed necessary, or rejected if deemed not part of the core data set. This process was repeated until consensus was reached. The Dutch IBD Biobank prospectively collects 225 standardised data items on various topics, including patient demographics, family history, diagnosis, disease activity, disease localisation, results of physical examinations, radiographic imaging results, laboratory and endoscopy results, previous and current treatment, as well as a wide array of disease and treatment complications. Validated questionnaires and scores, such as the Harvey-Bradshaw Index (HBI), the Simple Clinical Colitis Activity Index (SCCAI) and the Montreal classification are incorporated in the information model. This model contains both the IBD-related items as well as instructions on how to score these items. It has been shown that clinicians score subphenotypes of IBD similarly, with a good to excellent interobserver agreement.[9] The information model is provided in English in online **Supplementary Table 1** and can be downloaded in Dutch on the PSI website: www.parelsnoer.org. The Dutch IBD

Biobank information model is regularly updated. The latest version is based on the coding system called Detailed Clinical Models (http://www.detailedclinicalmodels.nl/dcm-en) and is called PRISMA (Parelsnoer Repository for Information Specification, Modelling, and Architecture).

Local databases and infrastructure

Each UMC has implemented the information model and collects and stores their patient information locally. As stated by the NFU, research data should be collected and registered directly at the source, that is, during the patient visit. Therefore, the data collection process should be incorporated into the clinical care structure.[10] This approach has been gradually implemented in each UMC depending on the capacities of their electronic health record (EHR) system. At the moment, each UMC has a procedure to extract, transform and upload pseudonymised information of participants to the PSI central database (Figure 1). The UMCs are in different stages of having implemented the 'at the source' approach. In some UMCs it is already fully implemented, whereas in other UMCs this process has not yet begun. The first visit is prepared by a trained research nurse and since most of the 225 data items do not change during every visit, for example, family history, medical doctors usually only need to register a subset of items during visits.

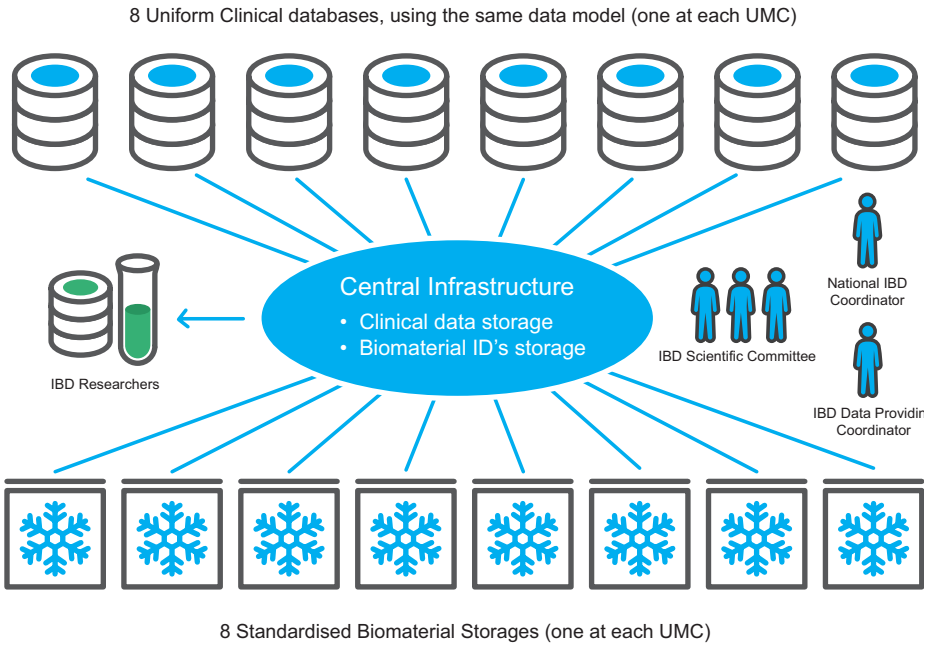


Figure 1 Overview of the data and biomaterial infrastructure of the Dutch IBD Biobank, built by the Parelsnoer Institute in collaboration with all eight university medical centres (UMC) in the Netherlands.
IBD: inflammatory bowel disease; ID: identifier.

Central Database and Central Data Infrastructure

Pseudonymised information about study participants is stored in the Central Database, managed by the Advanced Data Management (ADM) section of the Department of Medical Statistics and BioInformatics of the LUMC.

The software ProMISe, a web-based relational database management system for the design, maintenance and use of clinical data management, is used to store the Central Database. (https://www.msbi.nl/promise/). Researchers can access data in the Central Database following approval of their research proposal in secure web-based environment. Together, the Central Database and the web application form the Central Data Infrastructure (Figure 1).[7]

Data upload and pseudonymisation

In each UMC, data are automatically uploaded from the Local Database to the Central Database at least once a month. During the upload process, pseudoanymisation is performed by a trusted third party (TTP). Only the TTP has access to key containing both the local identifiers and the Dutch IBD Biobank identifier. Prior to the upload, data validation is performed locally on a set of essential data items. If necessary, corrections are made locally and subsequently included in the next upload. A full audit trail is in place for the entire process.

Privacy and information security audits

ADM, the Central Database and the Central Data Infrastructure software are audited according to Dutch NEN7510[11] international ISO 27.001[12] information security guidelines. ADM is audited twice per year while its software is periodically audited by Lloyds Register Quality Assurance, a certified independent auditor.

Biomaterial collection

In addition to the data items, biomaterials are collected from all patients with IBD: including DNA, serum, faeces, mucosal biopsies and resection specimens when surgical procedures were required. Laboratory experts of all eight university hospitals convened to create uniform biomaterial collection and processing protocols. The biomaterials are stored in one of the eight local biobanks (Figure 1). The biomaterial identifiers are uploaded to the Central Database and linked to the clinical data. Neither the local biomaterial identifiers nor the stickers on the biomaterial vials contain identifiable patient information. During the upload process, a unique additional biomaterial identifier is added to the local biomaterial identifier in case multiple UMCs have a biomaterial with the same identifier. When a research project is approved, all eight local biobanks will send the required biomaterials to the researcher while the biomaterial identifiers linked to the clinical data can be downloaded using the secure web portal of the Central Infrastructure. If a biomaterial sample does not meet

the required standards, the sample will be disposed. A brief summary of the biomaterial protocol is provided in Table 1.[7] The entire biomaterial protocols can be downloaded from www.parelsnoer.org, but are only available in Dutch.

Coordination

The Dutch IBD Biobank has two national coordinators and an assistant coordinator, who manage updates of the information model and the delivery of data and biomaterial to researchers (Figure 1).

Table 1 Sample collection[7]

Sample	Volume/ number	Processing	Time	Aliquoting	Storage	Additional information
Serum	10 ml clotted blood	2000xg at room temperature or 4°C for 10 minutes	Within 2-4 hours	≥ 5 x 0.5 ml	–80°C	Deviations
DNA	10 ml EDTA blood	Cell pellet, to UMC specifications	Within 4 weeks (4°C) or 3 months (< -20°C)	≥ 2 stock aliquots	4°C or lower	OD-ratio 260/280 and concentration in µg/ml
Faeces	Not defined	Direct storage or after homogenization	Within 12 hours	≥ 5 x 5 gr	–80°C	None
Intestinal biopsy	2 per localization: 'normal' and 'affected/ inflamed'	Formalin fixation and paraffin embedding	Immediate	Per set	Room temperature	None
Resection specimen	2 per localization: 'normal' and 'affected/ inflamed'	Formalin fixation and paraffin embedding	At Pathology	0.5 cm^3 samples	Room temperature	Only if feasible
Resection specimen	2 per localization: 'normal' and 'affected/ inflamed'	Snap frozen in isopentane	At Pathology	0.5 cm^3 samples	–80°C	Only if feasible

OD: optical density; UMC: university medical centre.

Informed consent

All patients with IBD who are treated in the Dutch UMCs are asked to participate in the Dutch IBD Biobank by their gastroenterologist during a visit to the outpatient department of their UMC. If they are willing to participate, they are asked to sign an informed consent form (English translation in

online Supplementary Document 1). Patients who choose to participate may revoke their consent at any point, after which their data and biomaterials will be removed from the Dutch IBD Biobank. Data and biomaterials that have already been sent to a researcher cannot be revoked, which is clearly stated in the patient informed consent form.

Patient enrolment

Patient enrolment started in January 2007 and is ongoing (Table 2).

Not all patients were asked to join at once, but they were asked in batches so gastroenterologist and research nurses could manage the initial data registration.

Every patient with IBD enrolled has a proven IBD diagnosis according to the Lennard-Jones criteria.[13] Diagnosis is confirmed by endoscopy, radiology and/or histology.

Table 2 Demographic characteristics of patients with IBD after the first data download on 17 July 2014, per university medical centre

	Total	MUMC	VUMC	AMC	UMCG	UMCU	EMC	LUMC	UMCN
n	3388	373	369	405	625	524	260	458	374
CD	2118	219	206	264	344	337	194	310	244
UC/IBD-U /IBD-I	1270	154	163	141	281	187	66	148	130
Sex (f/m%)	59/41	54/46	64/36	57/43	59/41	58/42	64/36	58/42	64/36
Age at diagnosis*	26 (20-37)	31 (22-44)	28 (21-37)	26 (20-35)	27 (21-39)	25 (19-35)	23 (18-30)	26 (20-34)	27 (20-37)
Disease duration*	12 (5-20)	8 (2-17)	11 (6-20)	13 (6-22)	8 (4-15)	14 (6-24)	12 (6-20)	15 (7-23)	14 (7-24)

*Median years with 25%–75% IQR.

AMC: Amsterdam Medical Centre; CD: Crohn's disease; EMC: Erasmus Medical Centre; F: female; IBD: inflammatory bowel disease; IBD-I: inflammatory bowel disease-indeterminate; IBD-U: inflammatory bowel disease-unclassified; LUMC: Leiden University Medical Centre; M: male; MUMC: Maastricht University Medical Centre; UC: ulcerative colitis; UMCG: University Medical Center Groningen; UMCN: Radboud University Nijmegen Medical Centre; UMCU: University Medical Centre Utrecht; VUMC: VU (Vrije Universiteit) University Medical Centre (Amsterdam).

Definitions

To create an overview of the content of the biobank, the characteristics of the patients were assessed. The following clinical and demographic items reported in this study are registered at the time of inclusion in the Dutch IBD Biobank and are referred to as *baseline*: first diagnosis, disease localisation, smoking status, employment status, gender, ethnicity, presence of a stoma or pouch, disease activity (modified HBI and modified SCCAI score) and date of birth. Disease localisation is scored according to the Montreal classification, which describes the maximum disease extent during entire disease course, and is registered at *baseline*. Disease localisation has to be confirmed by radiology, endoscopy or histology assessment. The items dysplasia, bowel

cancer, family history of IBD, current diagnosis and medication use described in this study were registered *during the last follow-up visit before the data download* in July 2014. Items describing disease behaviour, surgery, appendectomy, extraintestinal manifestations (EIM) and complications were registered *over the entire disease course up to baseline*. The definitions *baseline*, *last follow-up visit before the data download* and *over the entire disease course up to baseline* are graphically explained in online **Supplementary Figure 1**.

Statistical analyses

All descriptive statistics and statistical analyses are performed using Stata software V.13.1 (http://www.stata.com/). Continuous variables are expressed as medians and IQRs 25 and 75. Qualitative variables are presented as counts and frequencies. We compared outcomes between patients with CD and UC. Qualitative variables were analysed using the Pearson's X^2 test. Quantitative variables were analysed using the Mann-Whitney U test. We performed a multivariate analysis of the effect of smoking on different outcomes in all patients with IBD. We corrected for covariates with $P < 0.20$ in the univariate analyses (age, gender, diagnosis, disease duration and prior anti-tumour necrosis factor use). The statistical models were built using backward selection: covariates that were not statistically significantly influencing the outcome variable ($P > 0.05$) were removed from the model. We then applied the same strategy to patients with CD and UC separately to correct for disease activity. A P value < 0.05 was considered statistically significant.

Follow-up

Clinical and demographical follow-up data are collected at every visit to an outpatient department. Usually, patients with IBD in the Netherlands are seen by a gastroenterologist twice a year. This is standard clinical care following treatment protocols used in every UMC. The disease course is heterogeneous, as a consequence, data available on follow-up can be extensive for one patient but more limited for another. If requested by the gastroenterologist, a blood sample is taken. Furthermore, if required, intestinal mucosal biopsies are collected during endoscopy and resection specimens are obtained during surgery.

Findings to date

Consent rate and differences between participants and non-participants

We first assessed possible differences between patients with IBD willing to participate in the Dutch IBD Biobank and patients with IBD who did not want to participate. To do so, a subset at one UMC (UMCG) was downloaded and analysed. This subset was used because privacy guidelines do not

allow data of participants not wishing to take part to be uploaded to the PSI central database. On 17 July 2014, after the first data download, 786 patients were asked to participate in the UMCG. Of these, 742 patients with IBD gave their informed consent while 44 patients with IBD declined to participate. The consent rate was 93.4%. Table 3 provides an overview of the characteristics of those who consented to participate and those who did not. Of the 742 patients who consented, 625 were used in the analysis of the 2014 data because they met the selection criteria (clear IBD diagnosis, known date of birth and gender, informed consent and isolated DNA available including a biomaterial identifier). The characteristics of the consenting and non-consenting patients were similar. Only disease location according to the Montreal classification was statistically significantly different between these two groups (P = 0.037, X^2 test).

Table 3 Baseline characteristics of the responders and non-responders recruited through the University Medical Center Groningen on 17 July 2014

Responders			
	IBD (CD, UC, IBD-U)	CD	UC
n (%)	742 (100%)	411 (55%)	294 (40%)
Sex	742 (100%)	411 (100%)	294 (100%)
Male	305 (41%)	141 (34%)	142 (48%)
Female	437 (59%)	270 (66%)	152 (52%)
Age of onset median years (IQR 25-75)	26.8 (20-38)	24.5 (19-35)	30.6 (23-41)
Disease duration at inclusion median years (IQR 25-75)	8.2 (4-15)	9.3 (4-15)	7.6 (4-14)
Disease location (according Montreal)			
Crohn's disease		411 (100%)	
A1: diagnosis ≤ 16 years		58 (14%)	
A2: diagnosis 17-40 years		278 (68%)	
A3: diagnosis > 40 years		75 (18%)	
L1: ileal disease[a]		148 (37%)	
L2: colonic disease[a]		85 (22%)	
L3: ileocolonic disease[a]		163 (41%)	
L4: upper GI disease[b]		41 (10%)	
P: perianal		130 (32%)	
B1: non-stricturing, non-penetrating		211 (51%)	
B2: stricturing		134 (33%)	
B3: penetrating		66 (16%)	
Ulcerative colitis			288 (100%)
E1: proctitis			40 (14%)
E2: left-sided colitis			92 (32%)
E3: extensive colitis			156 (54%)

Table 3 *continued*

Non-responders			
	IBD (CD, UC, IBD-U)	CD	UC
n (%)	44 (100%)	25 (57%)	16 (36%)
Sex	44 (100%)	25 (100%)	16 (100%)
Male	16 (36%)	9 (36%)	5 (31%)
Female	28 (64%)	16 (64%)	11 (69%)
Age of onset median years (IQR 25-75%)	30.3 (19-42)	19.6 (17-39)	33.3 (25-42)
Disease duration at inclusion median years (IQR 25-75%)	8.1 (4-12)	7.2 (3-12)	8.8 (5-13)
Disease location (according to Montreal guidelines)			
Crohn's disease		25 (100%)	
A1: diagnosis ≤ 16 years		7 (28%)	
A2: diagnosis 17-40 years		12 (48%)	
A3: diagnosis > 40 years		6 (24%)	
L1: ileal disease*		4 (16%)	
L2: colonic disease*		10 (40%)	
L3: ileocolonic disease*		11 (44%)	
L4: upper gastrointestinal disease		0 (0%)	
P: perianal		9 (36%)	
B1: non-stricturing, non-penetrating		11 (44%)	
B2: stricturing		10 (40%)	
B3: penetrating		4 (16%)	
Ulcerative colitis			15 (100%)
E1: proctitis			5 (33%)
E2: left-sided colitis			5 (33%)
E3: extensive colitis			5 (33%)

[a]These percentages were calculated for 396 patients with CD (responders);
[b]These percentages were calculated for 402 patients with CD (responders);
*P = 0.037.
CD: Crohn's disease; GI: gastrointestinal; IBD: inflammatory bowel disease; IB-U: inflammatory bowel disease-unclassified; UC: ulcerative colitis.

The characteristics of the Dutch patients with IBD in UMCs

A download of data on 17 July 2014 was analysed to explore the demographic and clinical characteristics of the cohort recruited to that date. It included 3388 patients with IBD: 2118 CD (62.5%), 1190 UC (35.1%), 74 IBD-unclassified (2.2%) and 6 IBD-indeterminate (0.2%). The median age of patients with IBD at inclusion was 42 years old (IQR 32–54 years) (Tables 4-6). In all, 93% of patients are of Central European Caucasian descent and the other 7% are of African, Hindustani, Moroccan, Turkish, Asian, Jewish, other western, other non-western or mixed descent. Smoking status at the time of first IBD diagnosis was registered for 3021 patients with IBD (89%), and more patients with CD smoked compared with patients with UC (44% CD, 18% UC, P < 0.001). Patients with UC were more likely to have quit smoking in the 6 months prior to the first IBD diagnosis (1.0% CD,

4% UC, P < 0.001). Ileocolonic disease in patients with CD (46%) (Figure 2) and extensive colitis (E3) in patients with UC (56%) (Figure 3) are more common in our cohort than in other studies (Figures 4 and 5).[14-18] The high number of patients with extensive disease in our cohort can be explained by a selection bias (tertiary referral centres). The disease locations in CD were similar in men and women (Figure 6).

Table 4 Demographic characteristics of patients with inflammatory bowel disease in the Dutch IBD Biobank cohort on 17 July 2014

	IBD (CD, UC, IBD-I, IBD-U)	CD	UC
n (%)	3388 (100%)	2118 (62%)	1190 (35%)
Sex	3388 (100%)	2118 (100%)	1189 (100%)
Male	1377 (41%)	773 (36%)*	566 (48%)*
Female	2010 (59%)	1345 (64%)*	623 (52%)*
Age at inclusion median years (IQR 25-75)	42.5 (32-54)	41.1 (31-53)*	45.5 (34-56)*
Ethnicity	3323 (100%)	2073 (100%)	1170 (100%)
Caucasian	3090 (93%)	1930 (93%)	1084 (93%)
Other	233 (7%)	143 (7%)	86 (7%)
Non-IBD surgery			
Appendectomy†	394 (12%)	313 (15%)*	76 (6%)*
Smoking status at diagnosis	3021 (100%)	1910 (100%)	1037 (100%)
Current smoker	1052 (35%)	846 (44%)*	190 (18%)*
Former smoker (< 6 mth)	60 (2%)	19 (1.0%)*	40 (4%)*
Former smoker (> 6 mth)	601 (20%)	254 (13%)*	328 (32%)*
Never smoked	1308 (43%)	791 (42%)*	479 (46%)*

†Missing values were scored as absent;
*P < 0.001.
CD: Crohn's disease; GI: gastrointestinal; IBD: inflammatory bowel disease; IBD-I: inflammatory bowel disease-indeterminate; IBD-U: inflammatory bowel disease-unclassified; UC: ulcerative colitis.

Moreover, the most extensive disease during the entire disease duration (Montreal L (disease location) in patients with CD and Montreal E (disease extent) in patients with UC) is well documented in the Dutch IBD Biobank, while other studies often only report disease extent at the time of diagnosis (median disease duration in the Dutch IBD Biobank is 12 years). EIMs are more common in patients with CD than in patients with UC, which we corroborated in the Dutch IBD Biobank data (Figure 7).[19-21] We found that patients with UC who smoked more often suffered from ocular manifestations and arthropathy than those who did not smoke, matching previous findings.[22,23] An increased risk of EIM in patients with CD who smoked has previously been reported,[24] but we could not confirm this result in our cohort.

Table 5 Clinical characteristics, extraintestinal manifestations and complications in patients with inflammatory bowel disease in the Parelsnoer Institute cohort

	IBD	CD	UC
n (%)	3388 (100%)	2118 (62%)	1190 (35%)
Disease Characteristics			
Age of onset median years (IQR 25-75)	26.4 (20-37)	24.6 (19-33)**	30.1 (22-41)**
Disease duration at inclusion median years (IQR 25-75)	11.5 (5-20)	12.2 (6-22)**	10.7 (5-19)**
Family history of IBD	932 (28%)	613 (29%)*	301 (25%)*
Disease location (Montreal classification)			
L1: ileal disease[a]		379 (23%)	
L2: colonic disease[a]		518 (31%)	
L3: ileocolonic disease[a]		780 (46%)	
L4: upper GI disease[†]		177 (8%)	
P: perianal[†]		563 (27%)	
E1: proctitis[b]			82 (8%)
E2: left-sided colitis[b]			357 (36%)
E3: extensive colitis[b]			558 (56%)
Pouch[†]	155 (5%)	38 (2%)	112 (9%)
Disease Activity at inclusion			
mHBI score[c]		1828 (100%)	
Remission 0-4		1218 (67%)	
Mild disease 5-7		314 (17%)	
Moderate disease 8-16		274 (15%)	
Severe disease > 16		22 (1.2%)	
mSCCAI score[d]			1016 (100%)
Remission < 2.5			752 (74%)
Active disease ≥ 2.5			264 (26%)
Liver disease due to IBD	3388 (100%)	2118 (100%)	1190 (100%)
Primary sclerosing cholangitis (PSC)[†]	71 (2%)	25 (1.2%)**	43 (4%)**
Liver disease other than PSC[†]	65 (1.9%)	42 (2.0%)	22 (1.8%)
Extraintestinal manifestations	3388 (100%)	2118 (100%)	1190 (100%)
Skin manifestations[†e]	336 (10%)	250 (12%)**	80 (7%)**
Musculoskeletal manifestations[†f]	731 (22%)	513 (24%)**	204 (17%)**
Ocular manifestations[†g]	147 (4%)	104 (5%)*	38 (3%)*
Complications	3388 (100%)	2118 (100%)	1190 (100%)
Osteopenia (T-score < -1)[†]	676 (20%)	496 (23%)**	169 (14%)**
Thromboembolic events[†]	119 (4%)	76 (4%)	42 (4%)

[a]Percentages calculated for 1677 patients with CD;
[b]Percentages calculated for 997 patients with UC;
[c]mHBI: modified Harvey-Bradshaw Index score; patients with CD were asked to rate their well-being on a scale from 1 to 10 (1: feeling terrible to 10: feeling very good) and to rate abdominal pain on a scale from 0 to 10 (0: no abdominal pain to 10: worst pain imaginable). Patients were also asked to provide data on diarrhoea frequency. In addition, patients were asked about the presence of oral aphthous lesions, active abscesses and fistulae as well as extraintestinal manifestations (arthralgia, uveitis, erythema nodosum, pyoderma gangrenosum). The physician assessed the presence of anal fissures and evaluated possible abdominal resistance through physical examination. mHBI data were available on 1828 patients (100%);
[d]mSCCAI: modified Simple Clinical Colitis Activity Index score; patients with UC were asked to rate their well-

being on a scale from 1 to 10 (1: feeling terrible to 10: feeling very good). In addition, patients were asked to describe the defecation frequency during the day and during the night, the defecation urgency (yes or no), the presence of blood in their stool (yes or no) and extracolonic manifestations (arthritis, uveitis, erythema nodosum, pyoderma gangrenosum);
[e]The following skin manifestations associated with IBD were scored: pyoderma gangrenosum, erythema nodosum, hidradenitis suppurativa, psoriasis or palmoplantar psoriasiform pustulosis and metastatic CD. Which type was not specified, only the presence of a skin manifestation;
[f]Musculoskeletal manifestations were divided into two groups: (1) arthritis (red and swollen joints), for example, dactylitis, reactive arthritis, gout; (2) arthropathy (not red or swollen joints, but symptoms with an inflammatory pattern; pain at night or at rest), for example, sacroiliitis, ankylosing spondylitis, enthesitis and inflammatory back pain;
[g]Ocular manifestations comprised uveitis and episcleritis diagnosed by a doctor. Which eye condition was not specified, only the presence of an ocular manifestation;
[†]Missing values were scored as non-present;
*P < 0.05; **P < 0.001.
CD: Crohn's disease; GI: gastrointestinal; IBD: inflammatory bowel disease (CD+UC+IBD-I (indeterminate)+IBD-U (unclassified)); UC: ulcerative colitis.

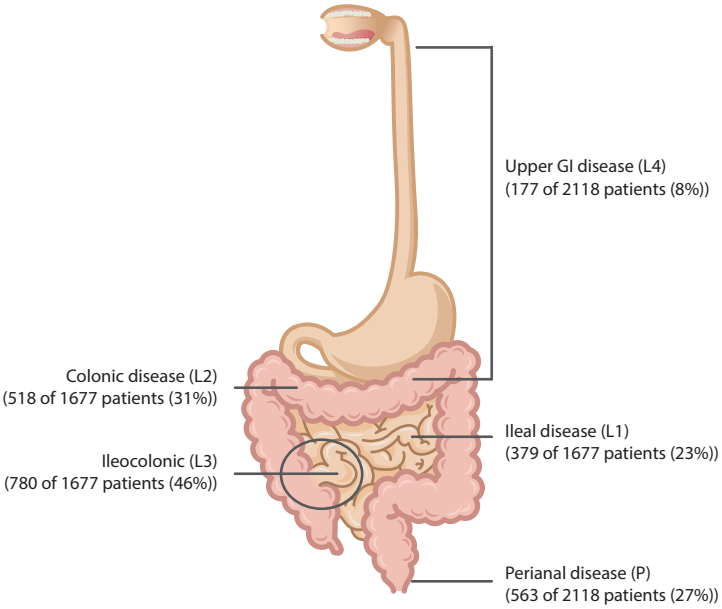


Figure 2 Disease localisation in patients with Crohn's disease in the Dutch IBD Biobank according to the Montreal classification.
GI: gastrointestinal.

Table 6 Malignancies, surgery and medication use of patients with inflammatory bowel disease in the Parelsnoer Institute cohort

	IBD	CD	UC
n (%)	3388 (100%)	2118 (62%)	1190 (35%)
Malignancy	3388	2118	1190
Dysplasia n[a]	131	62	63
Bowel cancer n[b]	15	9	5
Surgery	3388 (100%)	2118 (100%)	1190 (100%)
(Segmental) small bowel resection[†]	252 (7%)	242 (11%)	10 (0.8%)
Ileocaecal resection[†]	759 (22%)	758 (36%)	-
(Segmental) colon resection[†]	591 (17%)	368 (17%)	212 (18%)
Resection other[†]	168 (5%)	139 (7%)	28 (2%)
Strictureplasty[†]	99 (3%)	89 (4%)	-
Ileostomy/colostomy[†]	414 (12%)	283 (13%)	123 (10%)
Surgery for abscesses or fistulas[†]	494 (15%)	467 (22%)	27 (2%)
Outcome post-surgery	3388 (100%)	2118 (100%)	1190 (100%)
Stoma[†]	402 (12%)	270 (13%)	121 (10%)
Disease recurrence after IBD surgery			
Neoterminal ileum[c]	393 (52%%)	393 (52%)	-
Ileocolonic anastomosis[d]	56 (7%)	56 (7%)	-
Pouchitis[e]	93 (60%)	22 (58%)	67 (60%)
Surgical complication	1187 (100%)	959 (100%)	216 (100%)
Stricture anastomosis[f]	122 (10%)	107 (11%)	15 (7%)
Medication use during disease course	3306 (100%)	2068 (100%)	1158 (100%)
Immunomodulators[g]	2216 (67%)	1513 (73%)**	664 (57%)**
Biologicals[h]	1274 (39%)	1027 (50%)**	231 (20%)**
Azathioprine[i]	1374 (42%)	951 (46%)**	398 (34%)**
Mercaptopurine[j]	276 (8%)	199 (10%)**	73 (6%)**
Both azathioprine and mercaptopurine[k]	270 (8%)	172 (8%)	90 (8%)
Thioguanine[l]	114 (3%)	62 (3%)	50 (4%)
Methotrexate[m]	423 (13%)	363 (18%)**	52 (4%)**

[a]Dysplasia had to be confirmed in an intestinal biopsy by a pathologist. All intestinal biopsies were included including those from polyps;
[b]Bowel cancer included colorectal cancer, small bowel cancer and anal cancer;
[c]Percentage of disease recurrence in neoterminal ileum calculated from total patients with an ileocaecal resection (n = 759 IBD, n = 758 CD);
[d]Percentage of disease recurrence in ileocolonic anastomosis (no disease recurrence in neoterminal ileum) calculated from total patients with an ileocaecal resection (n = 759 IBD, n = 758 CD);
[e]Percentage of pouchitis calculated from total pouches (n = 155 IBD, n = 38 CD, n = 112 UC).
[f]Total patients who underwent surgery (small bowel resection, Ileocaecal resection, colon resection or resection other) (n = 1187 IBD, n = 959 CD, n = 216 UC);
[g]Immunomodulators: patients used one of the following immunosuppressives: azathioprine, Imuran, mercaptopurine, Purinethol, methotrexate, Metoject, thioguanine, Lanvis;
[h]Biologicals: patients used one of the following anti-tumour necrosis factors: infliximab, adalimumab or certolizumab;
[i]Azathioprine: patients used azathioprine or Imuran;
[j]Mercaptopurine: patients used mercaptopurine or Purinethol;
[k]Both azathioprine and mercaptopurine: patients used azathioprine and/or Imuran and mercaptopurine and/

or Purinethol. It was unclear which one of the drugs was used first;
[l]Thioguanine: patients used thioguanine or Lanvis;
[m]Methotrexate: patients used methotrexate or Metoject;
[†]Missing values were scored as non-present;
**P < 0.001.
CD: Crohn's disease; IBD: inflammatory bowel disease (CD+UC+IBD-I (indeterminate)+IBD-U (unclassified)); UC: ulcerative colitis.

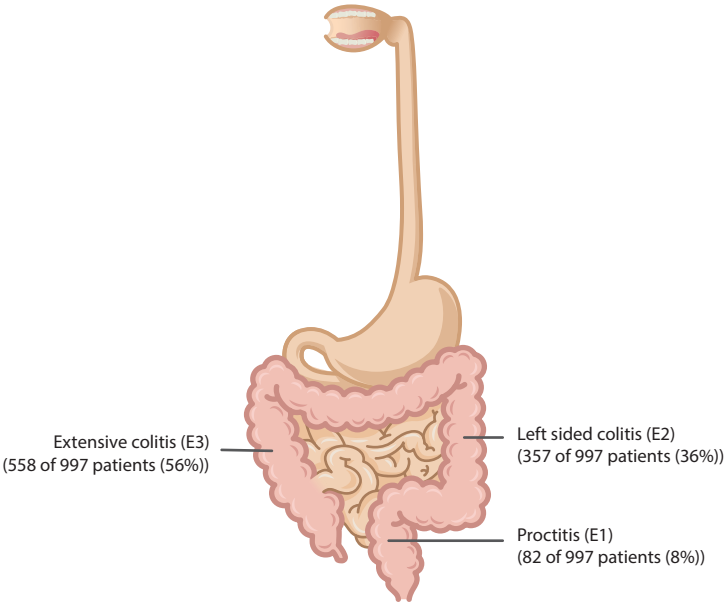


Figure 3 Disease localisation in patients with ulcerative colitis in the Dutch IBD Biobank according to the Montreal classification.

Genetic predictor of a fibrostenotic or inflammatory disease course in CD

The availability of genomic data and detailed clinical data in the Dutch IBD Biobank enabled a GWAS that aimed to find genetic predictors for recurrent fibrostenotic disease in patients with CD, by comparing the extremes of the clinical spectrum: (1) patients with CD with a mild disease course defined by inflammation without any signs of stricturing or penetrating disease during the last 5 years, versus (2) patients with CD who underwent ileocaecal resection due to confirmed intestinal strictures at least twice. We identified a genetic variant in the *WWOX* gene that regulates fibrosis through the SMAD pathway. The *WWOX* gene could therefore be an important signalling modulator involved in fibrostenotic CD (*Resubmitted to the Journal of Crohn's and Colitis*).

Previously published finding: Rare variants in MUC2 are associated with UC in the Dutch population. A subsequent study aimed to identify rare genetic variants with a large effect on UC susceptibility. Pooled resequencing of 122 genes in UC susceptibility loci in 1021 Dutch UC cases and 1166 Dutch controls revealed that rare variants in the *MUC2* gene were associated with increased

UC susceptibility (gene-based analysis with SKAT-O, nine variants in the *MUC2* gene: P value of 9.2×10^{-5}; threshold P = 0.0011 after Bonferroni correction). Interestingly, this association appeared to be population specific for the Netherlands.[25] Using the same approach and samples, a protein truncating variant in *RNF186* that protects against UC was also identified.[26]

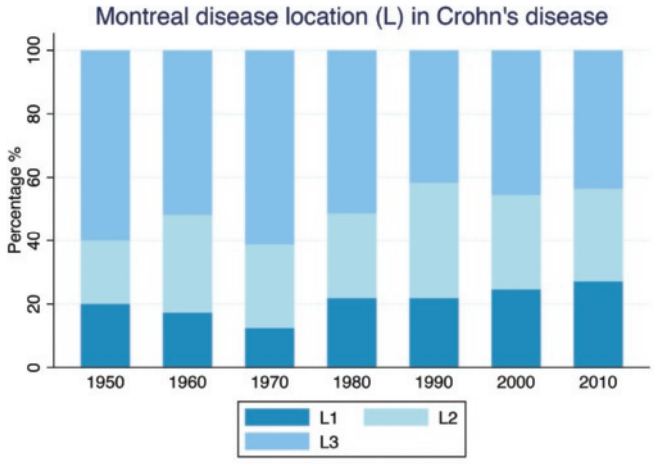


Figure 4 Date of Crohn's disease diagnosis and of disease location (L) according to the Montreal classification. L1: ileal; L2: colonic; L3: ileocolonic.

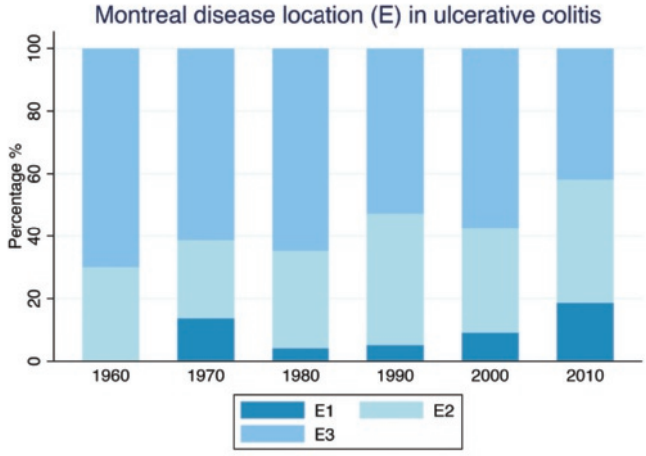


Figure 5 Date of ulcerative colitis diagnosis and of disease extent (E) according to the Montreal classification. E1: proctitis; E2: left-sided colitis; E3: pancolitis.

Associations between genetic variants and subphenotypes of IBD

The Dutch IBD Biobank participated in a large study where the clinical characteristics of patients with IBD were associated to genetic variants. The discovery of genetic variants associated with specific disease location and disease behaviour was published in the *Lancet*.[27]

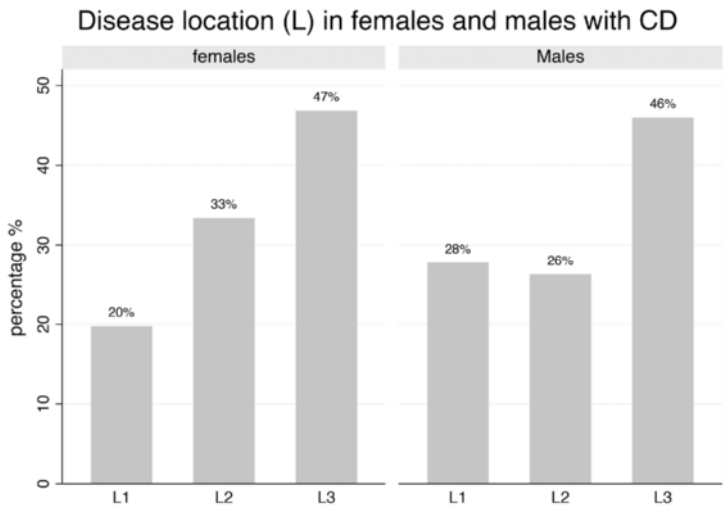


Figure 6 Disease location (L) according to the Montreal classification stratified by sex in patients with Crohn's disease (CD).
L1: ileal; L2: colonic; L3: ileocolonic.

GWAS and sequencing studies investigating the IBD diagnosis using DNA collections that were integrated in the Dutch IBD Biobank

For 1904 participants of the Dutch IBD Biobank genotype data are available consisting of ~200,000 single nucleotide polymorphisms (SNP) obtained using the Immunochip, an Illumina genotyping array focused on immune-mediated diseases. These genotype data were used in landmark genetic studies published in *Nature* and *Nature Genetics* investigating IBD pathogenesis.[3,28-30] These studies led to the discovery of 200 genetic loci associated with IBD, explaining 21.3% of the onset of IBD.

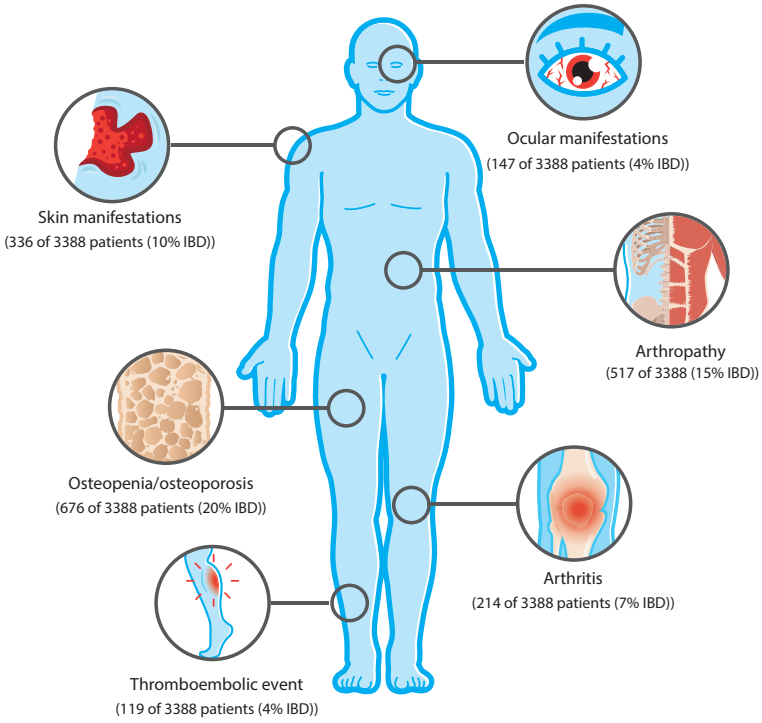


Figure 7 Extraintestinal manifestations and complications of patients with inflammatory bowel disease (IBD) in the Dutch IBD Biobank.

Discussion: strengths and weaknesses of the Dutch IBD Biobank

Strengths

A major strength of the Dutch IBD Biobank is its prospective design and extensive uniform information model comprising 225 data items, and the participation of all eight UMCs in the Netherlands. In addition, the biomaterials such as serum, DNA and a stool sample are collected at baseline, and, if available, biopsies from endoscopy and resection tissue are collected during follow-up, allowing the integration of subphenotypes enabling biomarker discovery research.

Since IBD is a chronic disease that requires lifelong treatment, patients treated in tertiary centres are rarely referred back to a general or local hospital and therefore loss to follow-up is uncommon.

Barriers to establishment and limitations

Setting up the Dutch IBD Biobank required a tremendous effort and there were many barriers to establishment. While some of these barriers were overcome, some limitations of the Dutch IBD Biobank remain. After a large initial grant provided by the Dutch government to the Netherlands Federation of University Medical Centres facilitating the establishment of the Dutch IBD Biobank and seven similar biobanks ended in 2011, the Dutch UMCs had to fund the continuation of the Dutch IBD Biobank themselves, meaning a reduction of staff that assisted in patient inclusion in some centres. As a consequence, the enrolment of patients has slowed down in these centres.

A major challenge was the establishment of the information technology (IT) infrastructure. In all UMCs, the local EHRs needed to be adapted so that the necessary information could be extracted. The gradual process of implementing data collection 'at the source' during the patient visit, and the renewal of EHRs in several hospitals means that adaptations to the local IT infrastructure continue to be necessary. Similar projects should be aware that the investments in the IT infrastructure will be ongoing after the establishment, and make sure they anticipate that continuous funding is required.

Data completeness, data similarity, data validation, quality control and feedback

A large majority of the data items were completely scored as can be seen in Tables 3-6. However, the different collection approaches by different UMCs sometimes lead to small differences in the clinical data, as some items were scored differently. Prior to completing this study, the authors reviewed all data and reported all inconsistencies to the national coordinators and to all UMCs. Several gastroenterologists, research nurses and IT departments improved the local data and a new upload to the Central Database was performed. Initially, very strict data validation steps were included in the Central Database software. However, these validation steps were too strict, and, because clinical patient records are often imperfect, very few patient records could be uploaded to the Central Database. After being aware of this problem, all data validation steps were removed from the Central Database software. Unfortunately, the lack of data validation steps led to errors in the data. Now, a small set of data validation protocols is in place. We recommend similar initiatives to start with simple data validation protocols and gradually expand these as the data quality and collection protocols improve.

Selection bias

Because all tertiary referral centres in the Netherlands participate in the Dutch IBD Biobank, the cohort will contain a large fraction of patients with IBD with a more severe disease course. This IBD cohort is not therefore suitable for studies that require a population-based cohort, for example, studies on the incidence and prevalence of IBD manifestations.

Collaboration

IBD researchers of the Dutch UMCs can access the Dutch IBD Biobank data and biomaterials after their research proposal has been approved by the Scientific Committee of the Dutch IBD Biobank. Other researchers can use the data and biomaterials of the Dutch IBD Biobank, but have to establish a cooperation with one or more Dutch UMCs.

Research proposal and application process

Research proposals can be submitted to the Scientific Committee and the Institutional Review Board. Proposals are judged against the following criteria:

1. It is reasonably plausible that the proposed research could lead to new insights.
2. The aims in the research proposal can be met using the proposed research methodology.
3. The proposed research is in concordance with the patient informed consent.
4. The proposed research will be conducted by people in institutes and facilities that are skilled and able to conduct the research.
5. The research proposal does not request more data and biomaterials than necessary.
6. The research proposal meets reasonable standards.
7. The proposed research does not unacceptably conflict or overlap with other research proposals.

After the Scientific Committee has approved a research proposal, the data manager will provide the pseudonymised research data in the web-based environment, and will facilitate the biomaterial delivery to the researcher. Applicants do not have to pay a fee.

The Dutch IBD Biobank can be contacted via email: IBDParel@umcg.nl. More information can also be found on the PSI website: www.parelsnoer.org. The Dutch IBD Biobank aims to cooperate with international IBD research groups. The information model and the list of biomaterials are publicly available and can be downloaded from the PSI website. The Dutch IBD Biobank encourages other biobanks to use the same information model and biomaterial collection standards to enable larger international studies on IBD and we encourage similar initiatives to contact us in an early stage.

Future developments

Genotyping the entire Dutch IBD Biobank

All DNA samples are in the process of being genotyped with a newly developed genome-wide genotyping array from Illumina, containing 750,000 SNPs. These data will be leveraged by

imputation against whole genome sequence data of 700 Dutch individuals studied in the Genome of the Netherlands project.[31] The availability of the genotype data will enable more genetic studies.

Web-based data access for researchers

The Dutch IBD Biobank is working on a multiomics data sharing portal called the *Molgenis Research IBD Portal*, based on Molgenis software.[32] This portal will make summary level statistics publicly available.

Mobile app for patients

The web-based follow-up of Patient-Reported Outcome Measurements including clinical disease activity scores is another project that the Dutch IBD Biobank is implementing. Patients will regularly fill in online questionnaires on disease activity, treatment response, quality of life and quality of care. Several UMCs are using the *app* My IBD Coach: http://www.sananet.nl/mijn-ibd-coach.html. The use of this *app* for IBD eHealth was extensively tested in a trial led by the MUMC, the Netherlands, where it was proven effective in reducing the number of hospital admissions.[33]

Conclusions

The Dutch UMCs have together created a biobank containing data and biomaterials of more than 3000 patients with IBD. The creation of the Dutch IBD Biobank took a very large multicentre multiyear effort, and new projects continue to improve the infrastructure and data collection. The main objective of the biobank is to facilitate the biomarker discovery. By now, studies using the Dutch IBD Biobank have led to the discovery a genetic predictor of a more severe disease course in patients with CD, showing that combining *-omics* data with prospectively collected clinical records can lead to useful results. Whether the standardising of patient data collection and during the patient visits and questionnaires online improves the clinical care of patients with IBD in the Netherlands is not yet known, but studies investigating the use of online disease activity scores and early detection of IBD exacerbations in the Netherlands are showing a reduction in hospitalisations.[33] We encourage researchers who want to establish similar biobanks to contact us, and to take our important recommendations, including the continuous IT funding, and the step-by-step implementation of data quality measures described in the discussion, into account.

Supplementary Data

Supplementary data are available at *BMJ Open* online.

References

1. Vektis. Insurance healthcare data in the Netherlands. https://www.vektis.nl
2. van den Heuvel TRA, Jeuring SFG, Zeegers MP, *et al*. Evolution of IBD incidence, disease phenotype, and mortality; 20 years of epidemiologic research in the Dutch population based IBDSL cohort. *Epidemiology* 2016.
3. Liu JZ, van Sommeren S, Huang H, *et al*. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979-86.
4. Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol* 2015;12:205-17.
5. Gevers D, Kugathasan S, Denson LA, *et al*. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;15:382-92.
6. Morgan XC, Tickle TL, Sokol H, *et al*. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13:R79.
7. Manniën J, Ledderhof T, Verspaget HW, *et al.* The Parelsnoer institute: a national network of standardized clinical biobanks in the Netherlands. *Open J Bioresour* 2017;25:1-8.
8. van Ommen GJ, Törnwall O, Bréchot C, *et al.* BBMRI-ERIC as a resource for pharmaceutical and life science industries: the development of biobank-based Expert Centres. *Eur J Hum Genet* 2015;23:893-900.
9. Spekhorst LM, Visschedijk MC, Alberts R, *et al.* Performance of the Montreal classification for inflammatory bowel diseases. *World J Gastroenterol* 2014;20:15374-81.
10. Centres NF of UM. Collecting data at the source' (Dutch title: Registratie aan de bron). Utrecht: NFU Nederlandse Federatie van Universitair Medische Centra.
11. NEN. Health Informatics - Information security management in healthcare. 7510:2011 NL. Vlinderweg, Delft, Netherlands: NEN.
12. International Organization for Standardization. Information technology - Security techniques - Information security management systems - Requirements. Geneva Switzerland: International Organization for Standardization; ISO/IEC 27001:2013.
13. Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol* Suppl 1989;170:6-9.
14. Gasche C, Scholmerich J, Brynskov J, *et al.* A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000;6:8-15.
15. Burisch J, Pedersen N, Čuković-Čavka S, et al. East-West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. *Gut* 2014;63:588-97.
16. Sjöberg D, Holmström T, Larsson M, *et al.* Incidence and natural history of ulcerative colitis in the Uppsala Region of Sweden 2005-2009 - results from the IBD cohort of the Uppsala Region (ICURE). *J Crohns Colitis* 2013;7:e351-7.
17. Moum B, Vatn MH, Ekbom A, *et al*. Incidence of ulcerative colitis and indeterminate colitis in four counties of southeastern Norway, 1990-93. A prospective population-based study. The Inflammatory Bowel South-Eastern Norway (IBSEN) Study Group of Gastroenterologists. *Scand J Gastroenterol* 1996;31:362-6.
18. Lakatos L, Kiss LS, David G, *et al.* Incidence, disease phenotype at diagnosis, and early disease course in inflammatory bowel diseases in Western Hungary, 2002-2006. *Inflamm Bowel Dis* 2011;17:2558-65.
19. Isene R, Bernklev T, Høie O, *et al*. Extraintestinal manifestations in Crohn's disease and ulcerative colitis: results from a prospective, population-based European inception cohort. *Scand J Gastroenterol* 2015;50:300-5.
20. Veloso FT, Carvalho J, Magro F. Immune-related systemic manifestations of inflammatory bowel disease. A prospective study of 792 patients. *J Clin Gastroenterol* 1996;23:29-34.
21. Bernstein CN, Blanchard JF, Rawsthorne P, *et al.* The prevalence of extraintestinal diseases in inflammatory bowel disease: a population-based study. *Am J Gastroenterol* 2001;96:1116-22.
22. Manguso F, Sanges M, Staiano T, *et al*. Cigarette smoking and appendectomy are risk factors for extraintestinal manifestations in ulcerative colitis. *Am J Gastroenterol* 2004;99:327-34.

23. Roberts H, Rai SN, Pan J, *et al.* Extraintestinal manifestations of inflammatory bowel disease and the influence of smoking. *Digestion* 2014;90:122-9.

24. Ott C, Takses A, Obermeier F, *et al.* Smoking increases the risk of extraintestinal manifestations in Crohn's disease. *World J Gastroenterol* 2014;20:12269-76.

25. Visschedijk MC, Alberts R, Mucha S, *et al.* Pooled Resequencing of 122 Ulcerative Colitis Genes in a Large Dutch Cohort Suggests Population-Specific Associations of Rare Variants in MUC2. *PLoS One* 2016;11:e0159609.

26. Rivas MA, Graham D, Sulem P, *et al.* A protein-truncating R179X variant in RNF186 confers protection against ulcerative colitis. *Nat Commun* 2016;7:12342.

27. Cleynen I, Boucher G, Jostins L, *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* 2016;387:156-67.

28. Franke A, McGovern DP, Barrett JC, *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118-25.

29. Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119-24.

30. Anderson CA, Boucher G, Lees CW, *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 2011;43:246-52.

31. Francioli LC, Menelaou A, Pulit SL, *et al.* Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 2014;46:818-25.

32. Swertz MA, Dijkstra M, Adamusiak T, *et al.* The MOLGENIS toolkit: rapid prototyping of biosoftware at the push of a button. *BMC Bioinformatics* 2010;11(Suppl 12):S12

33. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, *et al.* Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. *Lancet* 2017;390:959-68.