

Data linkage can reduce the burden and increase the opportunities in the implementation of Value-Based Health Care policy: a study in patients with ulcerative colitis (PROUD-UC Study)

John Walshe¹, Ashley Akbari¹, A. Barney Hawthorne², and Hamish Laing^{3,*}

Submission History

Submitted:	20/09/2021
Accepted:	26/09/2022
Published:	24/11/2022

¹Population Data Science, Swansea University Medical School, Faculty of Medicine, Health & Life Science, Swansea University, Wales, UK

²Cardiff and Vale University Health Board, Cardiff, Wales UK

³Value-Based Health and Care Academy, Faculty of Humanities and Social Sciences, Swansea University, Wales, UK

Abstract

Introduction

Healthcare systems face rising demand and unsustainable cost pressures. In response, health policymakers are adopting Value-Based Health Care (VBHC), targeting available resources to achieve the best possible patient outcomes at the lowest possible cost and actively disinvesting in care of low-value. This requires the evaluation of longitudinal clinical and patient reported outcome measures (PROMs) at an individual-level and population-scale, which can create significant data challenges. Achieving this through routinely collected electronic health record (EHR) data-linkage could facilitate the implementation of VBHC without an unacceptable data burden on patients or health systems and release time for higher-value activities.

Objectives

Our study tested the ability to report an international, patient-centred outcome dataset (ICHOM-IBD) using only anonymised individual-level population-scale linked electronic health record (EHR) data sources, including clinical and patient-reported outcomes, in a cohort of patients with moderate-to-severe ulcerative colitis (UC), receiving biopharmaceutical therapies ("biologics") in a single, publicly funded, healthcare system.

Results

We identified a cohort of 17,632 patients with UC in Wales and a cohort from two Health Boards of 447 patients with UC receiving biologics. 112 of these patients had completed 866 condition-specific PROMs during their biologics treatment. 44 out of 59 (74.6%) items in the ICHOM-IBD could be derived from routinely collected data of which a primary care source was essential for eight items and desirable for 21.

Conclusions

We demonstrated that it is possible to report most but not all the ICHOM-IBD outcomes using routinely collected data from multiple sources without additional system burden, potentially supporting Value-Based Health Care implementation with population data science. As digital collection of PROMs and use of condition-specific registries grow, greater utility of this approach can be anticipated. We have identified that the availability of longitudinal primary and secondary care data linked with PROMs is essential for this to be possible.

Keywords

data science; health policy; value-based health care; colitis, ulcerative; patient reported outcome measure; routinely collected health data

*Corresponding Author:

Email Address: hamish.laing@swansea.ac.uk (Hamish Laing)

Introduction

Health systems globally are under increasing pressure from rising demand and unsustainable costs, which is delaying access to new treatments and technologies. In response, healthcare payers and policymakers are seeking to increase the value created from the available resources by achieving better outcomes at the same or reduced cost. This approach, termed Value-Based Health Care (VBHC) [1] is an alternative to traditional cost containment, which gives little consideration to real-world outcomes. Importantly, in VBHC the outcomes measured should be those that matter most to patients; typically, these will be clinical outcomes supplemented by patient reported outcome measures (PROMs), such as the impact on normal activities or fatigue, which are most readily recorded by patients themselves [2]. This is a cultural shift for healthcare professionals familiar with a relentless drive for cost-improvement with little emphasis on patient-centred outcomes and coproduction [3]. It is a technical challenge because of its dependence on longitudinal patient outcome data [4]. The requirement for patient-centred outcomes and the historic paucity of capture of these as part of routine care means that implementing VBHC can be difficult because these data, if available, may be distributed across multiple systems [5] creating a significant data burden for patients and their clinical teams, which is a barrier to widespread adoption of VBHC [6].

The National Health Service (NHS) in Wales comprises seven Health Boards serving 3.2M citizens and is publicly funded by the Welsh Government, whose current health policy [7] includes a commitment to achieve the greatest value from the resources invested and makes achieving better value for patients a design principle for the future.

Ulcerative colitis (UC) is an inflammatory bowel disease (IBD) that affects 1 in every 284 people in Wales [8]. A chronic condition with periods of relapse and remission, UC may be controlled by medication and sometimes surgery. Patients with more severe UC have higher healthcare resource utilisation with more complications and emergency admissions and often are treated with biological therapies [9], which are expensive [10]. There is considerable interest in understanding the value of these medicines in the context of the outcomes they achieve for patients.

The International Consortium for Health Outcomes Measurement (ICHOM) has developed patient-centred outcome sets for many conditions using a standardised methodology that includes patient involvement and patient reported outcomes [11]. This standardisation potentially allows the value of patient care to be compared between healthcare providers and health systems. For IBD the ICHOM-IBD dataset comprises measures of symptoms, disease control, function, quality of life, survival, complications and hospitalisations [12]. While the data items are specified, the method of collecting them is not.

Although adoption is accelerating rapidly and they have been designated as part of the clinical record, PROM collection is not yet universal in Wales. One health board in our study had collected a condition-specific questionnaire, the IBD-Control PROM [13] regularly since 2016 from their patients with UC treated with biologics in a designated clinic. This PROM comprises thirteen questions, eight of which are used

to produce a numeric IBD-Control score quantifying how well a patient feels their condition is controlled at the time of completion. The scored questions and one non-scored question form most of the Symptoms, Function and Quality of Life domain of the ICHOM-IBD dataset.

The Secure Anonymised Information Linkage (SAIL) Databank (<https://saildatabank.com>) holds individual-level population-scale electronic health record (EHR) data for the population of Wales. This comprises demographic data, secondary care data for in-patient hospital admissions, outpatient and emergency department (ED) attendances and the primary care data from general practitioners (GPs) for 78% of the population of Wales [14, 15]. For this study, we also acquired and linked new data; pseudonymised lists of patients receiving biologics in two Health Boards and the IBD-Control PROM data from one of these.

We aimed to evaluate the level of completeness of the ICHOM patient-centred outcome set for patients with IBD that could be achieved just by linking routinely collected health data. If satisfactory, data-linkage in this way could underpin VBHC adoption more widely without imposing a significant data burden for health systems and patients.

Methods

Data sources

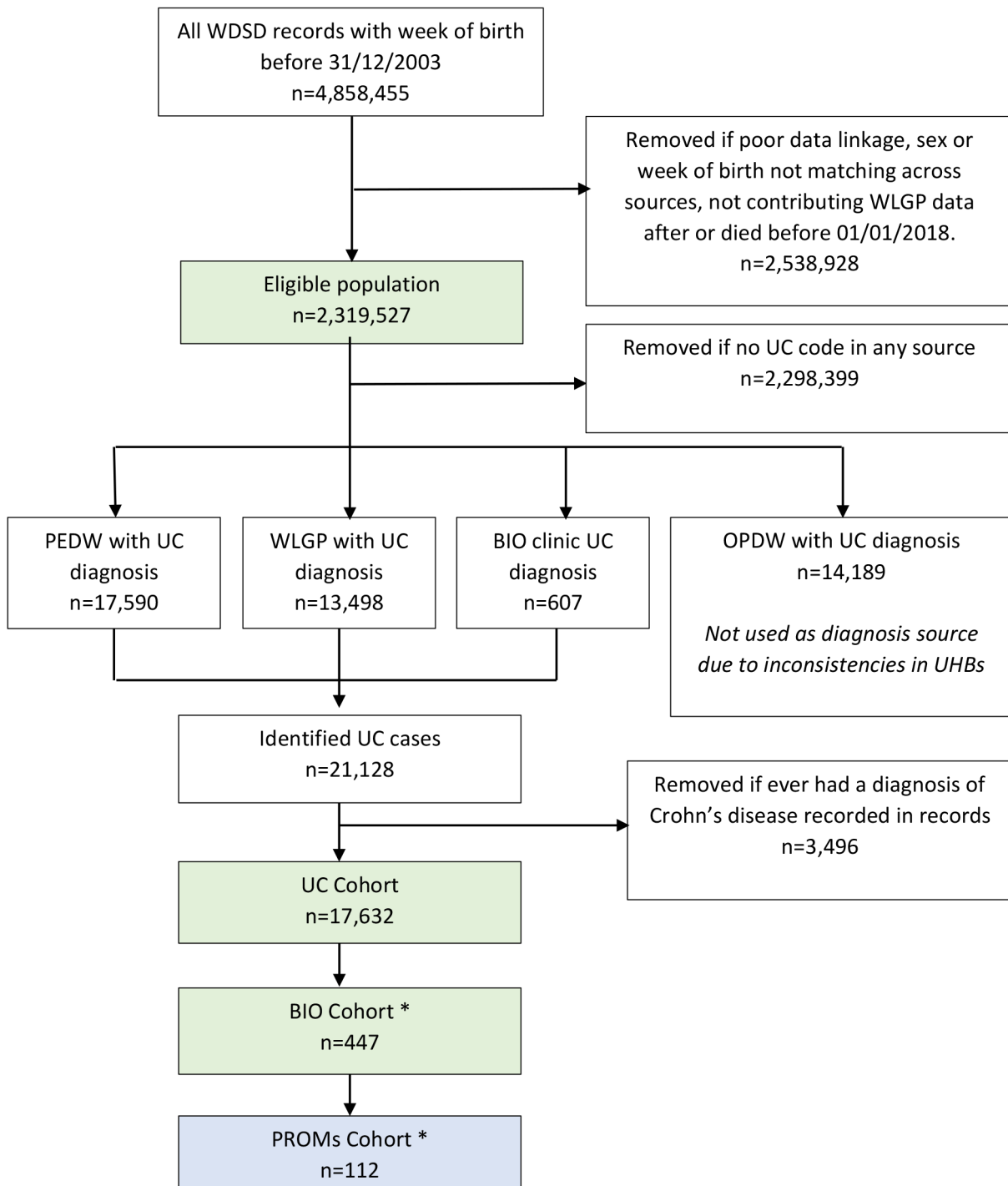
Data within the SAIL Databank trusted research environment (TRE) were accessed with extensive use of primary care Welsh Longitudinal General Practice (WLGP) data and the secondary care Patient Episode Database for Wales (PEDW), Emergency Department Data Set (EDDS), the Welsh Index of Multiple Deprivation Data (within WSDS) as well as newly acquired biologics clinic (BIO) and IBD-Control PROM data sources. A full list is provided in Appendix 1. Pseudonymisation and Data-Linkage was conducted according to standard SAIL methodology [16–18]. Only Welsh residents over the age of sixteen on 31st December 2019 (defined by week-of-birth \leq 31/12/2003) and with at least some GP data recorded after 1st January 2018 were included in the eligible population. Data were accessed back to 2000 and were excluded after 31st December 2019 to avoid COVID pandemic effects. For the ICHOM timeline analyses, data were included from 2013 onwards.

UC and biologic cohort generation

Tables of diagnostic codes for UC and Crohn's disease were prepared: Read codes in WLGP primary care data and International Classification of Diseases version 10 (ICD-10) codes in PEDW secondary care (see Appendix 2). These were reviewed by an experienced clinical informatician and senior NHS clinical coders and validated against coding tables used in another study [19].

The "UC Cohort" ($n=17,632$) was derived from all University Health Boards (UHBs) in Wales to identify the total eligible population ($n=2,319,527$) for those diagnosed with or having treatment for UC in at least one of three data sources (WLGP, PEDW, BIO) without any evidence in their records of having been diagnosed with Crohn's Disease. Figure 1.

Figure 1: STROBE diagram outlining the steps completed for the cohort selection and data cleansing from the multiple data sources used



*The STROBE diagram down to and including the UC cohort considers all seven UHBs; the BIO cohort two UHBs; and the PROMs cohort is from one of these UHBs.

(WSD: Welsh Demographic Service Dataset, WLGP: Welsh Longitudinal General Practice, UC: Ulcerative Colitis, PEDW: Patient Episode Dataset for Wales, BIO: biologics, OPDW: Outpatient Database for Wales, PROM: Patient Reported Outcome Measure, UHB: University Health Board).

Diagnostic coding of outpatient activity is inconsistent in Wales, and so Outpatient Dataset for Wales (OPDW) was not used as a diagnostic source.

A “biologics” cohort (BIO) was created for all patients with UC who attended biologics clinics, from the two UHBs from

which data were available (n = 447) representing a subset of the UC cohort. Of these, one UHB also had PROMs data available from 112 patients, who had completed a total of 866 IBD-control PROM questionnaires during their period of biologic therapy (38% of their patients). A small number of

UC patients were identified in the BIO cohort, who had not been identified in WLGP or PEDW. Inspection showed that most had insufficiently precise diagnostic coding in the other data sources.

ICHOM-IBD Outcome set generation

The ICHOM-IBD comprises 59 measures, some with multiple options, giving 131 separate data items. These cover four high-level domains: Symptoms, Functions & Quality of Life, Disutility of Care, Healthcare utilisation and Survival & Disease Control; along with baseline demographic and disease data (see Appendix 3). A full data dictionary was obtained from ICHOM and was mapped to over 2,000 Read and ICD-10 codes following an experienced clinical review. Algorithms were developed in DB2 SQL and R V4.1.0 to obtain each item from the data sources within SAIL.

For simpler data items, (e.g., COMORB: the presence of a comorbidity), patients were considered positive if any instances of the specified Read or ICD-10 codes were identified in WLGP or PEDW: absence of any codes in both was taken as negative. These items are universally available. For measures where all patients could have data, but some do not (e.g., HEIGHT), the absence of a recorded value was taken as null. For more complex items, (e.g., HOPADMUNP: the number of unplanned, IBD-related hospital admissions), a combination of clinical diagnostic and procedure coding and filtering on multiple data sources in SAIL was used.

The ICHOM-IBD includes nine items derived directly from the IBD-control PROM (see appendix 3). Of 447 patients treated with biologics at two of the Health Boards, 294 attended the one inviting patients to record the IBD-control PROM. Of these 294, 112 (38%) had completed PROMs. For the other patients the fields were left blank in the ICHOM-IBD.

The ICHOM-IBD set is collated for each patient for a specified period, varying by data item (generally the previous year or five years), allowing temporal trends to be studied. For PROMs data that had been collected more frequently, the most recent questionnaire was used for the 'yearly' ICHOM-IBD scores.

We attempted to derive the full ICHOM-IBD set for patients in the Biologics cohort. We analysed the number of patients with valid data in each item, and the number with each response value. Some ICHOM items relate to fixed or long-term metrics (e.g., height, age of onset), but many specifically reference the previous 12 months (e.g., medication use), or can be compiled over the previous 12 months rather than as an all-time indication (e.g., is a comorbidity current, rather than ever been present). Trends over several years can be compiled for a wide variety of metrics for both the UC cohort and Health Board matched populations for comparison.

Deprivation analysis

Deprivation scores were obtained from the Welsh Index of Deprivation (2019) derived from linkage to the individual's Lower-layer Super Output Area (LSOA, 2011). Where appropriate, these were standardised for the age-adjusted deprivation demographics of the two Health Board populations the patients were drawn from, rather than using all-Wales

comparators, as there is significant geographical variance in deprivation across Wales.

Results

The ICHOM-IBD comprises baseline demographics and disease data and four high-level domains (see Appendix 3). A full set of results for the BIO cohort is in Appendix 4.

Baseline demographics

As seen in Table 1, the age distribution of the UC cohort is generally older than the eligible population from which they are drawn, however the biologic cohort has a similar distribution to the eligible population. Biologic therapies are being administered to younger UC patients.

Identification of patients with UC

Figure 2 shows that the linkage of secondary care data (PEDW) with primary care health record data (WLGP) identified a greater number of patients with a diagnosis of UC (21,128) than either source on its own (17,590; 13,498). This analysis also identified 6,268 patients with a confirmed diagnosis of UC in secondary care with no record anywhere in their GP record.

Symptoms, Function and Quality of Life Domain

This domain of the ICHOM-IBD is highly dependent on the IBD-control PROM

IBD-Control PROMs: completion rates

Data from 866 IBD-Control PROM questionnaires were obtained from 112 Biologic patients (mean 7.7 per patient) from one health board, representing 25 % of the overall Biologics cohort and 38% of the one Health Board from where they were collected.

Of the 112 patients who returned IBD-Control PROMs: 17 completed only one questionnaire, 72 completed 5 or more, 51 submitted IBD-Control PROMs for longer than 12 months, 69 were still receiving treatment and reporting IBD PROMs at the study end. The average interval between questionnaire completion was 1.98 months, reflecting typical 2 monthly clinic attendances. From the first to last report, the average time was 15.7 months (for patients who completed more than one IBD-Control PROM). 43 patients stopped submitting PROMs during the study; most appeared to have completed their biologics therapy. For these, the average time from first to last report is 11.44 months.

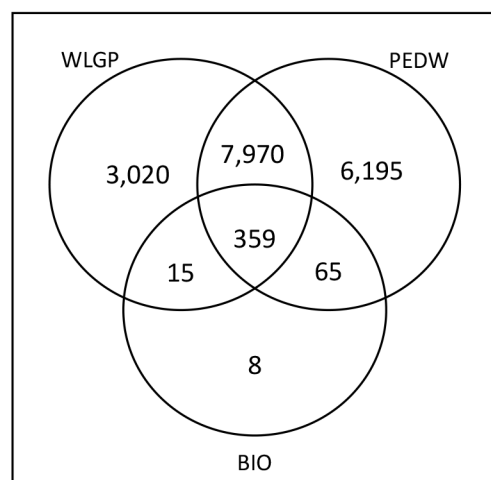
Disutility of care domain

Identification of steroid use reflecting disease severity and exacerbation was possible if prescribed in primary care (from WLGP) however we are aware that many patients are prescribed steroids by their secondary care specialist to be used as *required*, and these are not yet captured digitally in Wales.

Table 1: Population and cohort composition, stratified by: Age, Sex at birth, Health Board of Residence, and Deprivation Quintiles. 28 patients treated at one Health Board in our study resided in a different Health Board

Characteristic	Eligible population		UC Cohort		Biologics cohort	
	n	%	n	%	n	%
Total	2,319,527	100.00	17,632	100.00	447	100.00
Sex						
Female	1,200,432	51.75	9,117	51.71	209	46.76
Male	1,119,095	48.25	8,515	48.29	238	53.24
Age						
0-16	61,599	2.66	59	0.33	–	–
17-40	860,956	37.12	3,383	19.19	186	41.61
41-60	722,441	31.15	5,704	32.35	164	36.69
61-80	549,077	23.67	6,950	39.42	89	19.91
80+	125,454	5.41	1,536	8.71	–	–
mean	47 years		58 years		45 years	
median	47 years		60 years		45 years	
Health Board of residence						
Aneurin Bevan	392,637	16.93	2,776	15.74	16	3.58
Betsi Cadwaladr	472,468	20.37	3,996	22.66	–	–
Cardiff and Vale	386,743	16.67	2,690	15.26	266	59.51
Cwm Taf Morg.	365,626	15.76	2,998	17.00	12	2.68
Hywel Dda	277,101	11.95	2,115	12.00	–	–
Powys	49,847	2.15	395	2.24	–	–
Swansea Bay	318,374	13.73	2,436	13.82	141	31.54
not known	56,731	2.45	226	1.28	–	–
WIMD quintile						
1. Most deprived	460,783	19.87	3,000	17.01	98	21.92
2	458,772	19.78	3,470	19.68	73	16.33
3	449,911	19.40	3,425	19.42	66	14.77
4	427,093	18.41	3,663	20.77	68	15.21
5. Least deprived	466,236	20.10	3,848	21.82	139	31.10
not known	56,732	2.45	226	1.28	–	–

Figure 2: Numbers of patients in the UC cohort identified from each of the available data sources after those also diagnosed with Crohn's Disease had been removed



Complication of intervention, such as following endoscopy resulting in hospitalisation or death, could be identified. No such events occurred. Other outcomes of complications could not be identified reliably.

Healthcare utilisation domain

Hospitalisation events were available through PEDW and were classified as planned or unplanned and if they were IBD

related. Data were available from the Emergency Department Data Set (EDDS) for unscheduled, all-cause attendances at an Emergency Department (ED) in Wales. We looked for individuals in our cohorts who had attended ED for any reason and found that 17% of the eligible population, nearly one quarter (24%) of the UC cohort and 27% of the Biologic Cohort attended an Emergency Department in Wales at least once in the calendar year 2019.

Survival and disease control domain

The presence of anaemia was identified from WLGP in 12% of the Biologics cohort. The preferred measure for disease activity and remission in ICHOM-IBD is the Manitoba IBD Index. No data were available in our sources. Fewer than 2% of Biologics patient had been diagnosed with colorectal cancer and fewer than 2% had a cause of death that included IBD (absolute numbers redacted).

Summary of overall ICHOM-IBD outcome set completeness

The ICHOM-IBD comprises 59 main measures, which are derived from 131 individual items. We identified and utilised 44 out of 59 (75%) of the main measures and 94 out of 131 (72%) of individual items in creating the ICHOM-IBD, which were then available for our analysis. For those without PROM data, this fell to 35 measures (59%) and 85 (65%) items. For items where data completeness could be assessed (e.g., height, weight, smoking status), valid data were available in 80%–95% of individuals.

Compilation of the ICHOM-IBD was attempted for each cohort. As expected, different levels of completion were seen, with those patients in the Biologic (BIO) Cohort who had returned PROMs being the most complete (Table 2).

Appendix 4 lists the availability of data for each ICHOM-IBD outcome measure for the BIO cohort.

The contribution from primary care data (WLGP) is very significant, with 21 of the main measures being populated mostly and eight solely from these data. As seen in Table 3, the contribution to each item is also important; for example, 54% of comorbidity items recorded in ICHOM-IBD in 2019 were from WLGP and 46% from secondary care data (PEDW), with 45% of patients having specific comorbidities identified solely from WLGP, and 24% solely from PEDW.

A selection of available data are presented below to illustrate the benefits of a data-linkage approach.

Comorbidity

The major co-morbid conditions with UC have been selected.

The percentage of patients with no comorbidities recorded shows a decreasing trend in all groups. The Biologics cohort had a steeper decline of about 20% over 8 years. The most common comorbidities all showed a moderate increase over time.

The UC cohort generally shows a higher prevalence of chronic conditions (Arthritis, Lung disease, and Diabetes; Figure 3) than the age-standardised population group, and a higher rate of increase. The Biologics cohort initially shows a

lower rate than the overall UC cohort but then increases more steeply.

Deprivation

The WIMD 2019 [20] was used at LSOA (2011) level [21]. Patients who could not be linked with an LSOA were excluded from this analysis. For the Biologic cohort, the comparison was adjusted to reflect the WIMD for the Health Boards in which they resided.

As has been described elsewhere in the UK [22], there was a greater prevalence of UC overall in Wales in less deprived populations, and this effect was seen even after adjustment for the relative affluence of the health boards in which they reside.

Discussion

In this study we have shown that whilst many domains of the ICHOM-IBD person-centred outcome set can be extracted successfully from routine clinical data, especially when data are linked from multiple data sources, some elements, notably those on quality of life and humanistic outcomes can only be captured by patients completing PROMs and linking those to their clinical outcomes. These findings suggest that to implement Value-Based health Care, validated PROMs need to be captured from patients alongside their clinical outcomes; repeated over time throughout the care cycle and linked to clinical interventions and input costs. From these data can be derived estimates of which interventions and treatments are of high value and which are of lower or no value. This allows resources to be directed towards the highest value care and released from low-value activities, supporting the twin aims of better outcomes and financially sustainable health systems.

For outcomes to be captured at scale, digital platforms are required, and these are becoming widely available. Nonetheless, the collection and aggregation of clinical and patient reported outcomes longitudinally throughout a pathway of care and from different care settings with siloed data sources can create an unmanageable data burden [23], potentially preventing the implementation of VBHC. This provides an opportunity for population data scientists to support VBHC policy adoption through the aggregation, linkage, analysis and visualisation of these data in standardised form [24]. Our approach could be replicated for other conditions and automated within the NHS for routine use. However, as we have demonstrated, it will require access to GP (WLGP) and specialist clinical datasets for the required level of data completeness.

Outcome data can inform shared decision-making for patients and clinicians and, if standardised measures are used, benchmarking individual patients' progress against similar patients or between services and even countries for the same condition [25]. The creation of internationally developed person-centred outcome data sets by ICHOM is valuable in this regard, provided the data can be collected consistently and easily. We have included our tables in Supplementary Appendices for others to use.

We studied people with moderate to severe ulcerative colitis in this proof-of-concept study because they have greater

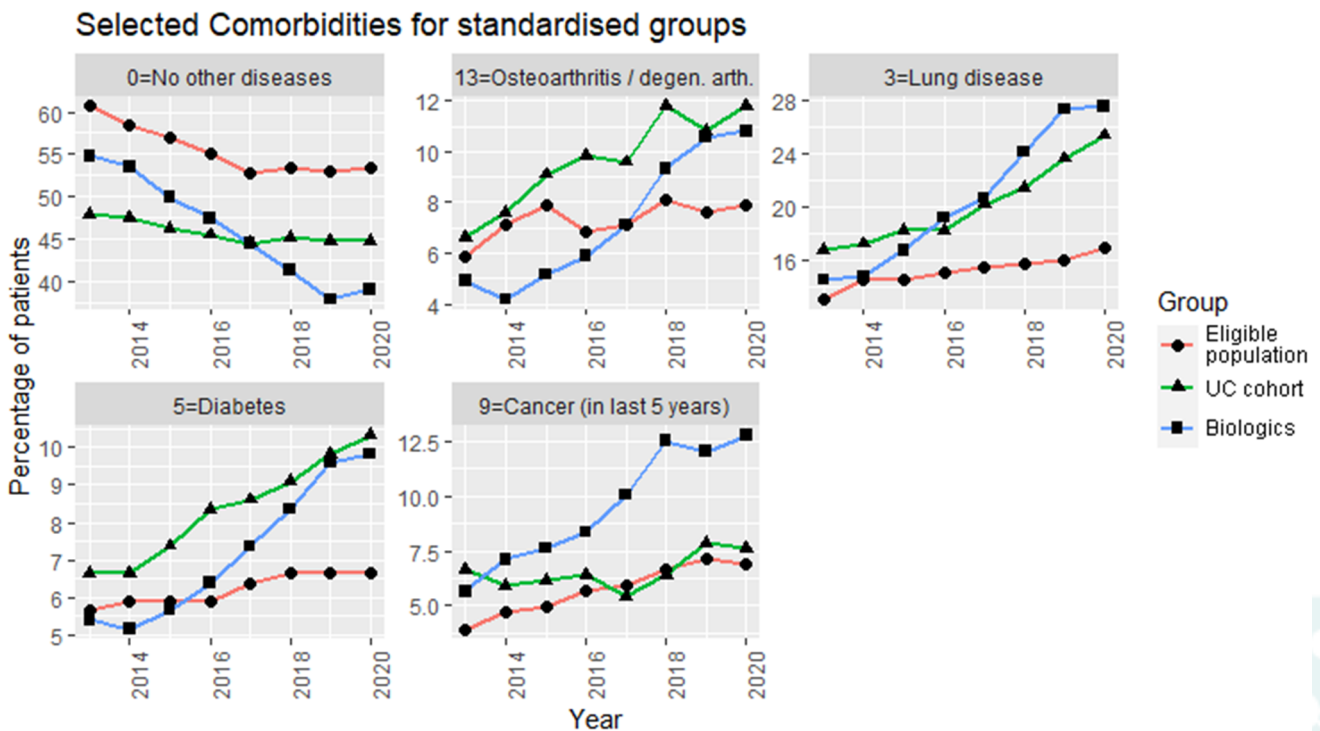
Table 2: Availability of data for patients in the BIO cohort. No data were available for any patient for Montreal Index, Educational Level, HIV status (redacted), Bowel cancer screening, and biologic disease activity/remission for any patients

Domain	Availability	Exceptions
Baseline demographic measures	95–100%	No education level
Baseline clinical measures	81–100%	No HIV status (redacted)
Baseline Condition measures	100%	
Disease localisation / behaviour	0%	Manitoba IBD Index and Montreal Classification not captured
Treatment factors	100%	No bowel screening data
Symptoms/functions QoL	25%	100% in those completing PROMs
Disutility of Care	100%	Incomplete outcome of complication data. Steroid data in primary care only
Healthcare Utilisation	100%	
Survival and anaemia	100%	
Disease control	0%	No Manitoba IBD Index

Table 3: The number of comorbidities for the patients in the BIO cohort identified from each data source demonstrates the importance of including both

Data source	Number of comorbidities identified from each source	Number of patients identified from each source
WLGP data only	810	253
PEDW data only	704	167
Present in both		66

Figure 3: Trends in comorbidity rates for the three age and Health Board standardised patient groups, for the presence of comorbidities in the preceding year (5 years for cancer). The year on the x-axis represents the year a condition is recorded as present



healthcare resource utilisation due to unscheduled admissions, higher rates of emergency and elective surgery and the use of biopharmaceutical therapies [26]. The relapsing and remitting nature of UC and its impact on quality of life makes the collection of PROMs over time, in addition to purely clinical

outcomes, very helpful in understanding the value of different treatments and interventions. Three of the 12 major outcome groups in the ICHOM-IBD set are derived from the IBD-Control PROM, so these patient reported data are critical to understanding the outcomes that matter most in UC.

There are some limitations in this study. We used biologic therapies as a proxy measure for people with moderate to severe disease, based on the licenced indications. More precise clinical classification requires data that are not yet routinely recorded in digital form in Wales. We were only able to include IBD-Control PROM data from one Health Board, however, this will become easier as the collection and sharing of PROMs becomes routine in the Welsh NHS [27] and PROM data are centrally stored in a standardised format for analysis in a National Data Resource.

We were not able to report accurately the use of steroid medication, which can be an indication of disease exacerbation. Whilst primary care prescriptions were included, many patients with UC are prescribed steroids in a secondary care setting where patient-linked prescribing data are not yet available in Wales. A national Electronic Prescription and Medicines Administration system is planned and will address this deficit.

The ICHOM-IBD person-centred outcomes set includes the Manitoba [28] and Montreal [29] Indices. These are important for cohort segmentation but are not consistently recorded in the health record in Wales and so could not be included. The Montreal Index is included in the IBD Registry (<https://ibdregistry.org.uk/>) and so could be linked in future if the use of the registry by clinicians increases. Other items without available data were HIV status (redacted in SAIL), educational level, bowel cancer screening status (this could be linked in future) and biological remission status.

Finally, as the data are drawn from routinely collected electronic health record (EHR) data sources and systems, changes in recording and investigating practices over time could potentially impact the quality and consistency of recording.

Despite these limitations, most of the ICHOM-IBD set was derived successfully for the BIO cohort using routinely collected linked health data from multiple sources and the IBD-Control PROM, where it had been collected, and could also be linked to clinical events using our methodology. The incorporation of additional clinical datasets would be required to enable its use in VBHC.

We also have identified the critical importance of sourcing information from both GP and secondary care systems in the completion of the ICHOM-IBD set. This was possible in our study because nearly all (78%) GP data in Wales are available in the SAIL trusted research environment (TRE). An agreement is required to make primary care data available to the NHS outside of a TRE, for insights into the value of different elements of care pathways; not only for risk stratification and comorbidity data, but also to ensure that all relevant patients are identified. The identification of over 6,000 patients whose diagnosis of UC in secondary care (PEDW and BIO) was not also recorded in their GP record is a concern and potential clinical risk, particularly because some of them were receiving biologic therapies. It may reflect poor communication or processes between the care sectors, adversely impacting patient care and service planning.

The inclusion of comorbidities in the ICHOM-IBD is important as these may impact the overall quality of life and function for patients. Our data show that the most common comorbidities demonstrate a moderate increase over time which may reflect the ageing of the cohorts. The

Biologic cohort showed a lower rate of comorbidities initially than the UC-population cohort, which may reflect a younger patient group, but this increases more steeply over time which warrants further investigation. The addition of deprivation through linkage of outcomes data to a non-clinical dataset allowed the confirmation of the increased prevalence of UC in more affluent communities, which may also impact on patient experience and what matters most to them.

Conclusions

Implementing a policy of Value-Based Health Care is a cultural, technical and data challenge. It requires the collection, analysis and visualisation of treatment outcomes that matter to patients as completely as possible with linkage to the whole care pathway. In this proof-of-concept study, we have shown that validated person-centred outcomes could be derived without creating an additional data burden for patients and clinicians through data linkage of multiple sources of routinely collected clinical and validated patient reported outcomes (IBD-Control PROM), placing population data science at the heart of policy implementation. However, for this to be implemented will require large scale collection of PROMs as part of routine practice and better use of and access to the data in disease registries for more granular clinical data to enable better cohort segmentation and baseline determination. Missing data for disease extent and clinical measures of disease control and incomplete PROM data, limited our ability to estimate that value of biologic interventions in UC in this study.

For PROMs to be meaningful in understanding value in long-term conditions, they must be collected throughout the care cycle and, at a minimum, before and after specific interventions or treatments of interest. Collecting and analysing them over time creates a powerful tool for coproduction of care decisions and for the comparative value of treatments, particularly where these are of high cost.

Standardised outcome sets that are developed with patients and globally applicable are helpful but may include data items that are not available or not easily mapped to coding systems in use in a particular health system. They can help to identify gaps in routinely collected data and data sources that should be included in linkage. Data that allow risk adjustment and cohort segmentation is important if outcomes are to be compared between subsets of patients or between clinicians, services or health systems.

Acknowledgements

This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who make anonymised data available for research.

The authors would like to thank Professor John Williams CBE and Wayne Lewis (Crohn's Colitis UK) for advice, Katherine Harrison and Yvette Lloyd (NHS Clinical Coders) for support with code mapping, Swansea Bay University Health Board gastroenterology team and Cardiff and Vale University Health Board for provisioning data, Digital Health and Care

Wales (DHCW) for pseudonymisation, Shannon Stevens for project management.

Statement on conflicts of interest

Swansea University is in receipt of funding from Pfizer UK which supports a programme of activities that is led by HL. AA, JW, and ABH have no conflicts to declare.

Consent for publication

No identifiable individual person data are included. All data are redacted to prevent re-identification.

Funding

JW, AA, and HL are employees of Swansea University, which received partial funding from Pfizer in connection with the development of this manuscript.

The research output relates to delivery of the European Regional Development Fund co-funded ACCELERATE project involving Operations supported by the West Wales & Valleys and South-East Wales ERDF EU Structural Funds Programme.

AA and this work were supported by Health Data Research UK, which receives its funding from HDR UK Ltd (HDR-9006) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and the Wellcome Trust. This work was supported by the ADR Wales programme of work. The ADR Wales programme of work is aligned to the priority themes as identified in the Welsh Government's national strategy: Prosperity for All. ADR Wales brings together data science experts at Swansea University Medical School, staff from the Wales Institute of Social and Economic Research, Data and Methods (WISERD) at Cardiff University and specialist teams within the Welsh Government to develop new evidence which supports Prosperity for All by using the SAIL Databank at Swansea University, to link and analyse anonymised data. ADR Wales is part of the Economic and Social Research Council (part of UK Research and Innovation) funded ADR UK (grant ES/S007393/1).

Availability of data and materials

The data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK, but as restrictions apply they are not publicly available. All proposals to use SAIL data are subject to review by an independent Information Governance Review Panel (IGRP). Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been

granted, it is gained through a privacy protecting safe-haven and remote access system referred to as the SAIL Gateway. SAIL has established an application process to be followed by anyone who would like to access data via SAIL at <https://www.saildatabank.com/application-process>.

Ethics approval and consent to participate

Approval for the use of anonymised data in this study, provisioned within the Secure Anonymised Information Linkage (SAIL) Databank was granted by an independent Information Governance Review Panel (IGRP) under project 1009. The IGRP has a membership comprised of senior representatives from the British Medical Association (BMA), the National Research Ethics Service (NRES), Public Health Wales and Digital Health and Care Wales (DHCW formerly NWIS). Usage of additional data was granted by the data owners through data sharing agreements. The SAIL Databank is General Data Protection Regulation (GDPR) and the UK Data Protection Act compliant. This study received IRAS approval: 276316 REC 20/HCRW/0016.

Authors contributions

JW was the primary data analyst and was responsible for analysing the data, producing the figures and drafting the manuscript, editing, proofreading and approving the final manuscript.

AA, was the senior data analyst who supervised the data analysis and reporting, proofread and approved the final manuscript for publication.

ABH provided specialist clinical advice, collated and prepared the PROM data for linkage, reviewed the draft, and approved the final manuscript.

HL conceived the study and is the principal investigator for PROUD-UC. He was responsible for the study design and writing, editing, proofreading, approving and submitting the final manuscript.

Supplementary appendices

1. Data sources used
2. Clinical codes (ICD10 and Read) used for UC and Crohn's Colitis
3. Mapped codes for ICHOM-IBD outcomes set
4. Full ICHOM-IBD results for BIO cohort

References

1. Porter ME, Teisberg EO. Redefining Health Care, Boston Mass.: Harvard Business Review Press, 2006.
2. Knowles SR, Graff LA, Wilding H, Hewitt C, Keefer L, Mikocka-Walus A. "Quality of Life in Inflammatory Bowel Disease: A Systematic Review and Meta-analyses-Part I,"

- Inflammatory Bowel Diseases*, vol. 24, no. 4, pp. 742–751, 2018 <https://doi.org/10.1093/ibd/izx100>
3. Elwen G, Nelson E, Hager A, Price A. “Coproduction: when users define quality,” *BMJ Qual Saf*, vol. 29, pp. 711–716, 2020. <https://doi.org/10.1136/bmjqs-2019-009830>
 4. Hurst L, Mahtani K, Pluddemann A, Lewis S, Harvey K, Briggs A, Boylan A-M, Bajwa R, Haire K, Entwistle A, Handa A, Heneghan C. “Defining Value-based Healthcare in the NHS,” Oxford University, Oxford, 2019.
 5. EIT Health, “Implementing Value-Based Health Care in Europe: Handbook for Pioneers” (Director Gregory Katz), European Union, Brussels, 2020. <https://eithealth.eu/think-tank-topic/value-based-healthcare> [Accessed 1 Sep 2021]
 6. Bohm N, Bermingham S, Grimsey Jones F, Goncalves-Bradley DC, Diamantopoulos A, Burton JR and Laing H. “The Challenges of OutcomesBased Contract Implementation,” *PharmacoEconomics*, 2021 <https://doi.org/10.1007/s40273-021-01070-1>
 7. Welsh Government, “A Healthier Wales: our Plan for Health and Social Care,” 2021. [Online]. Available: <https://gov.wales/sites/default/files/publications/2021-09/a-healthier-wales-our-plan-for-health-and-social-care.pdf>. [Accessed 1 Sep 2021].
 8. Scanlon I, Lewis W, Wang T, Rees R, Berry S. “Inflammatory Bowel Disease in numbers: Understanding the Scale of Crohn’s and Colitis in Wales.,” Crohn’s and Colitis UK, 2020.
 9. National Institute for Health and Care Excellence, “Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy. Technology appraisal guidance TA329.,” 25 Feb 2015. [Online]. Available: <https://www.nice.org.uk/guidance/ta329>. [Accessed 25 Sep 2021].
 10. Beard J, Franco D, Click B. “The Burden of Cost in Inflammatory Bowel Disease: A Medical Economic Perspective and the Future of Value-Based Care. ” *Curr Gastroenterol*, vol. 22, no. 6, 2020. <https://doi.org/10.1007/s11894-020-0744-z>
 11. International Consortium for Health Outcomes Measurement, “ICHOM Patient Centered Outcome Measures” [Online]. Available: <https://www.ichom.org/standard-sets/>. [Accessed 2 Sep 2021].
 12. International Consortium for Health Outcomes Measurement, “IBD standard set,” [Online]. Available: <https://connect.ichom.org/standard-sets/inflammatory-bowel-disease/>. [Accessed 1 Sep 2021].
 13. Bodger K, Ormerod C, Shackcloth D, Harrison M. “Development and validation of a rapid, generic measure of disease control from the patient’s perspective: the IBD-Control questionnaire.,” *Gut*, vol. 63, pp. 1092–1102, 2014;. <https://doi.org/10.1136/gutjnl-2013-305600>
 14. Health Data Research UK, “HDR Innovation Gateway - WLGP dataset,” [Online]. Available: <https://web.www.healthdatagateway.org/dataset/33fc3ffd-aa4c-4a16-a32f-0c900aaea3d2>. [Accessed 1 Sep 2021].
 15. Lyons J, Akbari A, Agrawal U, Harper G, Azcoaga-Lorenzo A, Bailey R, et al. “Protocol for the development of the wales multimorbidity e-cohort (WMC): data sources and methods to construct a population-based research platform to investigate multimorbidity.,” *BMJ Open*, vol. 11, no. 1, 2021. <https://doi.org/10.1136/bmjopen-2020-047101>
 16. Ford DV, Jones KH, Verplancke J-P, Lyons RA, John G, Brown G, et al. “The SAIL Databank: building a national architecture for e-health research and evaluation.,” *BMC Health Serv Res.*, vol. 9:157, 2009. <https://doi.org/10.1186/1472-6963-9-157>
 17. Jones KH, Ford DV, Jones C, Dsilva R, Thompson S, Brooks CJ, et al. “A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation.,” *J Biomed Inform.*, vol. 50, no. Aug, pp. 196–204, 2014. <https://doi.org/10.1016/j.jbi.2014.01.003>
 18. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford DV, et al. “The SAIL databank: linking multiple health and social care datasets.,” *BMC Med Inform Decis Mak.*, vol. 9:3, 2009. <https://doi.org/10.1186/1472-6947-9-3>
 19. Crohn’s and Colitis UK, [Online]. Available: <https://www.crohnsandcolitis.org.uk/news/study-shows-over-50-more-people-in-wales-have-crohns-or-colitis-than-previo>. [Accessed 10 February 2022].
 20. Welsh Government, “Welsh Index of Multiple Deprivation,” [Online]. Available: <https://gov.wales/welsh-index-multiple-deprivation>. [Accessed 1 Sep 2021].
 21. Office for National Statistics, UK Gov., “Census geography areas,” [Online]. Available: <https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography#super-output-area-soa>. [Accessed 1 Sep 2021].
 22. Pasvol TJ, Horsfall L, Bloom S, Segal AW, Sabin C, Field CN, Rait G. “Incidence and prevalence of inflammatory bowel disease in UK primary care: a population-based cohort study,” *BMJ Open*, 2020. <https://doi.org/10.1136/bmjopen-2019-036584>
 23. Mjaset C, Ikram U, Nagra NS, Feeley TW. “Value-Based Health Care in Four Different Health Care Systems,” *NEJM Catalyst*, November 2020. <https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0530>
 24. McGrail K, Moran R, O’Keefe C, Preen D, Quan H, Sanmartin C, et al. “A position paper on population data science: the science of data about people,” Swansea, 2018.

25. International Consortium for Health Outcomes Measurement, "Collect, Process and Report Data - ICHOM Global Benchmarking Platform," [Online]. Available: <https://www.ichom.org/global-benchmarking-platform/>. [Accessed 25 Sep 2021].
26. Perera S, Yang S, Stott-Miller M, Brady J. "Analysis of Healthcare Resource Utilization and Costs after the Initiation of Biologic Treatment in Patients with Ulcerative Colitis and Crohn's Disease.," *JHEOR.*, vol. 6(1), pp. 96-112, 2018. <https://doi.org/10.36469/9791>
27. Welsh Government, "Value Based Health Care Programme – Data Requirements," Welsh Health Circular (2020)003, 2020.
28. Clara I, Lix LM, Walker JR, Graff LA, Miller N, Rogala L, Rawsthorne P, Bernstein CN. "The Manitoba IBD Index: Evidence for a New and Simple Indicator of IBD Activity," *Am J Gastroenterol*, vol. 104, pp. 1754–1763, 2009. <https://doi.org/10.1038/ajg.2009.197>
29. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, et al. "Toward an integrated clinical,

molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology.," *Can J Gastroenterol*, vol. 19 Suppl A:5A–36A, 2005.

Abbreviations

VBHC:	Value-Based Health Care
PROM:	Patient Reported Outcome Measure
UC:	Ulcerative Colitis
IBD:	Inflammatory Bowel Disease
ICHOM:	International Consortium for Health Outcomes Measurement
WIMD:	Welsh Index of Multiple Deprivation
PEDW:	Patient Episode Dataset for Wales
WLGP:	Welsh Longitudinal General Practice
SAIL:	Secure Anonymised Information Linkage (Databank)
WDS:	Welsh Demographic Service Dataset
EDDS:	Emergency Department Data Set
ICHOM-IBD:	ICHOM outcome data set for IBD
SQL:	Structured Query Language
UHB:	University Health Board



Supplementary Appendix 1: data sources for cohort identification and ICHOM-IBD generation

	Name	Description
ADDE	Annual District Death Extract	Mortality records from the Office for National Statistics (ONS) including cause and date of death.
BIO	Biologics	Records of patients receiving biopharmaceutical treatments, their dates of treatment and medicine administered and in designated outpatient clinics in two health boards.
EDDS	Emergency Department Data Set	Records of attendance at Emergency Department with dates and ICD-10 clinical coding.
OPDW	Out-Patient Database for Wales	Out-Patient records for secondary care episodes including clinic attendance.
PEDW	Patient Episode Database for Wales	In-patient records for secondary care spells and episodes, including dates, locations and ICD-10 and OPCS clinical coding.
PROM	Patient Reported Outcome Measures	The IBD-control PROM questionnaire responses from patients attending biologics clinics at one health board.
WDSD	Welsh Demographic Service Dataset	Residence and registration history, including linkage of Lower-layer Super Output Area (LSOA) version 2011 to Welsh Index of Multiple Deprivation (WIMD) version 2019.
WLGP	Welsh Longitudinal General Practice	Primary care records with date and Read clinical coding for general practice attendances.



Supplementary Appendix 2: Diagnosis Read codes

Diagnosis	Read_cd	Code Description
0 = Crohn's disease	J40..	Regional enteritis -Crohns Disease
0 = Crohn's disease	J400.	Regional enteritis of the small bowel
0 = Crohn's disease	J4000	Regional enteritis of the duodenum
0 = Crohn's disease	J4001	Regional enteritis of the jejunum
0 = Crohn's disease	J4002	Crohns Disease of the terminal ileum
0 = Crohn's disease	J4003	Crohns Disease of the ileum unspecified
0 = Crohn's disease	J4004	Crohns Disease of the ileum NOS
0 = Crohn's disease	J4005	Exacerbation of Crohns Disease of small intestine
0 = Crohn's disease	J400z	Crohns Disease of the small bowel NOS
0 = Crohn's disease	J401.	Regional enteritis of the large bowel
0 = Crohn's disease	J4010	Regional enteritis of the colon
0 = Crohn's disease	J4011	Regional enteritis of the rectum
0 = Crohn's disease	J4012	Exacerbation of Crohns Disease of large intestine
0 = Crohn's disease	J401z	Crohns Disease of the large bowel NOS
0 = Crohn's disease	J402.	Regional ileocolitis
0 = Crohn's disease	J40z.	Regional enteritis NOS
0 = Crohn's disease	Jyu40	[X]Other Crohns Disease
0 = Crohn's disease	N0311	Arthropathy in Crohn s disease
0 = Crohn's disease	N0453	Juvenile arthritis in Crohn s disease
1 = Ulcerative colitis	J410.	Ulcerative proctocolitis
1 = Ulcerative colitis	J4100	Ulcerative ileocolitis
1 = Ulcerative colitis	J4101	Ulcerative Ulcerative Colitis
1 = Ulcerative colitis	J4102	Ulcerative rectosigmoiditis
1 = Ulcerative colitis	J4103	Ulcerative proctitis
1 = Ulcerative colitis	J4104	Exacerbation of Ulcerative Colitis
1 = Ulcerative colitis	J410z	Ulcerative proctocolitisNOS
1 = Ulcerative colitis	J411.	Ulcerative (chronic) enterocolitis
1 = Ulcerative colitis	J412.	Ulcerative (chronic) ileocolitis
1 = Ulcerative colitis	J413.	Ulcerative pancolitis
1 = Ulcerative colitis	J436.	Microscopic Colitis
1 = Ulcerative colitis	Jyu41	[X]Other Ulcerative Colitis
1 = Ulcerative colitis	Jyu42	[X]Other specified noninfective gastroenteritis+colitis
1 = Ulcerative colitis	N0310	Arthropathy in Ulcerative Colitis



Supplementary Appendix 2: Diagnosis ICD-10 codes

Diagnosis	Read_cd	Code Description
0 = Crohn's disease	K500	Crohn's disease
0 = Crohn's disease	K501	Crohn's disease
0 = Crohn's disease	K508	Crohn's disease
0 = Crohn's disease	K509	Crohn's disease
0 = Crohn's disease	M074	Crohn's disease
0 = Crohn's disease	M091	Crohn's disease
1 = Ulcerative colitis	K510	Ulcerative colitis
1 = Ulcerative colitis	K511	Ulcerative colitis
1 = Ulcerative colitis	K512	Ulcerative colitis
1 = Ulcerative colitis	K513	Ulcerative colitis
1 = Ulcerative colitis	K514	Ulcerative colitis
1 = Ulcerative colitis	K515	Ulcerative colitis
1 = Ulcerative colitis	K518	Ulcerative colitis
1 = Ulcerative colitis	K519	Ulcerative colitis
1 = Ulcerative colitis	K528	Ulcerative colitis
1 = Ulcerative colitis	M075	Ulcerative colitis
1 = Ulcerative colitis	M092	Ulcerative colitis
2 = Indeterminate IBD or colitis unclassified	K523	other IBD

Please see the website for Supplementary Appendices 3, 4.

