Incidence and prevalence of microscopic colitis across the United Kingdom

Yvonne Nartey, Mariam Ratna, Timothy R Card, Colin J Crooks, Joe West, Laila J Tata

Division of Epidemiology & Public Health, School of Medicine

University of Nottingham

December 2021

Jointly funded by

Coeliac UK and Crohn's & Colitis UK

University of Nottingham

UK | CHINA | MALAYSIA

Report correspondence & queries: Laila.Tata@nottingham.ac.uk

Summary of report

This is the final project report on microscopic colitis for the project 'Contemporary epidemiology of coeliac disease, dermatitis herpetiformis, Crohn's disease and ulcerative colitis in the United Kingdom' which is jointly funded by Coeliac UK and Crohn's and Colitis UK to explore incidence and prevalence over the past 20 years.

This report summarises the incidence of microscopic colitis across the United Kingdom between 2000 and 2020, and the prevalence in the year 2020. We used health care records from 38.3 million people (7.6 million under age 18 years) registered with general practices contributing anonymised information to the Clinical Practice Research Datalink (CPRD). The CPRD population sample showed good representativeness of the total population from Office of National Statistics (ONS) information. Incidence analyses were adjusted for all population characteristics within each model. Prevalence estimates were provided along with age-standardised prevalence to compare across devolved nations.

Most of the English population (85%) had their general practice records linked to Hospital Episode Statistics (HES) records; this sub-group population had more detailed information for analysis.

Microscopic colitis incidence between 2000 and 2020

Whole population: Between 2000 and 2020, 10,226 people received a new diagnosis of microscopic colitis during 286,673,368 person-years of active general practice registration in the population. This represented a population incidence rate of 3.57 new cases per 100,000 person-years (i.e., 1 new case of microscopic colitis identified among every 28,011 people if they were each followed for 1 year). Incidence rates for England, Scotland, Wales and Northern Ireland were 3.67, 3.19, 2.72, and 2.43 new cases per 100,000 person-years respectively. Adjusted incidence was lower in Wales compared with England.

Adjusted incidence was 2.5 times higher in females compared with males. Incidence generally increased with age and was highest in 70-79 year olds. Incidence increased over time and was lower in groups with more socioeconomic deprivation. In England, adjusted incidence overlapped for most regions; London had the lowest adjusted relative incidence, whilst Yorkshire & The Humber and the North East had the highest.

For the English sub-group population with HES-linked records, the incidence rate was 3.68 new cases per 100,000 person-years (i.e., 1 new case of microscopic colitis identified among every 27,174 people if they were each followed for 1 year). Compared with the UK population and the overall English population, patterns of incidence were similar across age, gender, time and socioeconomic groups. Those with White ethnicity recorded had a higher incidence compared with all other ethnicity groups

Young people under age 18: Between 2000 and 2020, only 27 young people received a new diagnosis of microscopic colitis among 7.6 million followed-up across the UK, which was deemed too few to calculate stable estimates of incidence by characteristics. Cases were distributed across all characteristics.

Microscopic colitis prevalence in 2020

Whole population: From the CPRD population in 2020, we identified 8,575 people with microscopic colitis which represented a prevalence of 0.05%, equating to one in every 2,000 people with a diagnosis of microscopic colitis in the UK. Prevalence was marginally higher in England compared with the other devolved nations. Within England, age-standardised prevalence varied modestly between regions, with London having the lowest prevalence and Yorkshire & The Humber having the highest. Prevalence was considerably higher in females (0.08%) compared with males (0.03%) and increased with age for both groups. Prevalence decreased with increasing socioeconomic deprivation in the UK and England but not in the devolved nations. Prevalence in the English sub-group population with HES-linked records was the same as in the overall English population. Those with White ethnicity recorded had a higher prevalence compared with all other ethnicity groups and prevalence was higher in rural compared with urban areas.

Young people under age 18: From the CPRD population in 2020, we identified fewer than 20 young people with microscopic colitis, which was deemed too few to calculate stable estimates of prevalence.

Contents

List of Tables

List of Figures

List of Data Supplements with this report

The data supplements contain all tables in the report and additional tables (including separate tables for England, England sub-group with HES-linked data, Scotland, Wales and Northern Ireland)

datasheet_UoNreport2021_MC_incidence.xlsm

datasheet_UoNreport2021_MC_prevalence.xlsm

Acronyms & Abbreviations

CI	confidence interval
MC	microscopic colitis
CPRD	Clinical Practice Research Datalink
GP	general practice
HES	Hospital Episode Statistics
IBD	inflammatory bowel disease
ICD	International Classification of Diseases
IMD	Index of Multiple Deprivation (a measure of socioeconomic categorisation)
IRR	incidence rate ratio
ONS	Office of National Statistics
UK	United Kingdom

1 Introduction

This report summarises the incidence and prevalence of microscopic colitis (MC) across the UK. Figures are reported for the United Kingdom as a whole and separately for England, Scotland, Wales, and Northern Ireland. For England, figures are also reported separately for 10 geographical regions.

This work is part of a grant entitled 'Contemporary epidemiology of coeliac disease, dermatitis herpetiformis, Crohn's disease and ulcerative colitis in the United Kingdom' which is jointly funded by Coeliac UK and Crohn's and Colitis UK to explore incidence and prevalence over the past 20 years.

The research uses information from the Clinical Practice Research Datalink (CPRD), an internationally recognised source of real-world data, to provide estimates of UK-wide and country specific population incidence and prevalence. Information that is recorded prospectively as part of routine clinical practice for individuals across the United Kingdom (UK) is available in CPRD for research. All electronic health records are anonymised so individual patients cannot be identified by the researchers. Approval to carry out the research for this grant was obtained from the CPRD Independent Scientific Advisory Committee (Protocol approval No: 19_266A) in December 2019.

2 Study population and analysis

2.1 Developing the baseline population

The study population included all individuals who were actively registered with general practices (GP) at any time between January 2000 and July 2020, and who contributed anonymised data to CPRD. Prevalence was calculated on July 1st, 2020.

Individual patient records had to meet a minimum standard of data quality to be included in the study, defined by CPRD as 'acceptable' and 'up-to-standard' for research. There are two current GP databases provided by CPRD: CPRD GOLD, containing data contributed by practices using Vision® software, and CPRD Aurum, containing data from practices using EMIS Web® electronic patient record system software (https://www.cprd.com/). Data from CPRD GOLD include practices from across the UK whilst CPRD Aurum includes practices from England and Northern Ireland only.

This study used the whole CPRD population, by combining both CPRD GOLD and CPRD Aurum practices and retaining the most recent records of those patients who had transferred between GOLD and Aurum. For 85% of the English population, patients' general practice records were individually linked with their Hospital Episode Statistics (HES) records, which include all inpatient and outpatient hospital admissions in England. The English population with HES-linked data provided a sub-group baseline population with more details for analysis including high data completeness on recorded ethnicity.

Information on age and gender was obtained from each person's general practice record. Exact date of birth is not available in CPRD as a method of maintaining patient anonymity; age was thus available to the nearest month for those under age 16 and to the nearest year for those age 16 and older. We used CPRD's 3-option gender field to categorise people as male or female and excluded a small number of people (<0.01% and thus insufficient for analyses) recorded as having indeterminate gender. We did not use further information to attempt to comprehensively capture non-binary sexuality as the validity of this information in CPRD is not known.

The location of a person's household and their general practice is not available in CPRD to maintain patient anonymity. In a secure linkage process before data are supplied to researchers, CPRD uses a trusted third party (NHS Digital) to link the Index of Multiple Deprivation (IMD) with the location of contributing general practices across the UK. The IMD is an area-level measure of socioeconomic deprivation based on information collected from the national census. For England only, IMD and a binary rural-urban classification are also available for patient households. Available for analysis were quintiles of IMD for GPs across the UK (i.e., the GP-level socioeconomic deprivation quintile using the postcode for the practice where the patient was registered) and for households in England (household-level socioeconomic deprivation quintile, according to the patient's home postcode), and rural-urban classification for households in England (also using postcode).

Information on ethnicity was only obtained for the sub-group baseline population of people in England with individually linked HES data, by combining ethnicity coding from both GP and hospital records. For Scotland, Wales and Northern Ireland, ethnicity information was available from GP records, however, it was missing in 99%, 99% and 92% of the populations respectively and was thus considered unusable for analysis.

2.2 Defining microscopic colitis

For all individual patients in the baseline population, cases with microscopic colitis were defined as those who had a relevant clinical diagnosis, coded as a CPRD medcode (Appendix A) in their electronic general practice record, with or without codes for indeterminate IBD or IBD codes that did not specify Crohn's disease or ulcerative colitis. If MC was coded but an individual also had medcodes for ulcerative colitis or Crohn's disease, they were defined as cases with macroscopic IBD, not MC, and were not included for analysis. Individuals with MC will have been diagnosed in secondary care following blood tests and endoscopy that do not show macroscopic IBD, but biopsy samples show microscopic changes in gut tissue.

CPRD medcodes align with the International Classification of Disease (ICD) diagnostic coding system used in HES, however there is no specific ICD-10 code for MC and thus it is not possible to identify MC cases from coding in hospital records.

2.3 Data management and statistical analysis

Data were received from CPRD via secure download as flat text files and were transferred into Stata format. Stata16MP statistical software was used for all data management and analyses.

The measure of incidence provides information on the number of people who are being newly diagnosed with MC. Incident cases of MC were defined by the earliest date on which they had a clinical diagnosis or a prescription for aminosalicylates after the first 12 months of their actively registered time with the general practice. Recorded diagnoses within the first 12 months of registration are more likely to reflect prevalent cases so their inclusion may overestimate true incidence.

For the sub-analysis of young people under age 18, a less stringent restriction for defining incidence was used. This considered that diagnoses made very early in life are more likely to be incident regardless of proximity to registration, and that recording of important existing medical conditions should be made close to new GP registration as they are more likely to represent active disease. Therefore, incident cases of MC in young people under age 18 were defined by the earliest date on which they had a clinical diagnosis or a prescription for aminosalicylates after the first 6 months of their actively registered time with the general practice (no children under age 2 years had an MC diagnosis).

Incidence was defined as the number of incident cases per 100,000 person-years contributed by people in the baseline population. Person-years is a measure that captures how long each person has been actively registered at their GP which is the time during which we can measure a new diagnosis of MC. It is calculated by adding together the total time each person in the population has been actively registered (e.g., if 3 people were registered for 1 year, 2.5 years and 7 years respectively, the total person-years would be 10.5). Incidence (also called incidence rate) with 95% confidence intervals (CI) was calculated using the strate command.

The measure of prevalence provides information on the number of people who are living with MC at a certain point in time. Prevalent cases were defined by having a clinical diagnosis on or before July 1st, 2020. For prevalent cases, we used the entirety of a person's health record, which included any recorded diagnoses before they joined their current general practice or diagnoses recorded in their medical history.

Prevalence was defined as the number of prevalent cases per 100 people in the baseline population on July 1st, 2020. Prevalence with 95% CI was calculated using the proportion command.

Incidence and prevalence were calculated for the United Kingdom as a whole and separately for England, Scotland, Wales, and Northern Ireland. For England, incidence and prevalence was separately calculated for 10 regions, based on the boundaries of Strategic Health Authorities which were part of the National Health Service structure between 2002 and 2013.

For the UK as a whole and for each country, incidence and prevalence were calculated for the total population and separately by age group, gender, GP-level socioeconomic deprivation quintile and calendar period. A sub-analysis of people in England who had HES-linked information was conducted which included age group, gender, ethnicity group, GP-level socioeconomic deprivation quintile, household-level socioeconomic deprivation quintile, rural-urban classification, and calendar period.

For incidence analyses, Poisson regression was used to calculate incidence rate ratios (IRR) with 95% CI to provide comparisons between groups (e.g., the ratio of the incidence of MC in females compared with males). Unadjusted IRRs and fully adjusted IRRs were calculated. Fully adjusted IRRs were adjusted for all factors in the model, e.g., an IRR for females compared with males in England was adjusted for age, ethnicity, socioeconomic quintile, urban-rural classification, and calendar period.

Case and baseline population numbers on which incidence and prevalence was calculated are presented in the appendix (Appendix B). To prevent identification of individuals whose health records are used for CPRD research, numbers are suppressed as "<5" wherever fewer than 5 people populate a table cell.

Population pyramids (Appendix C) were created to compare the age and gender distribution of the CPRD baseline populations for the United Kingdom, England, Scotland, Wales, and Northern Ireland with Office of National Statistics (ONS) *Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2019, using April 2020 local authority district codes* (https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates). The dstdize command was then used to calculate overall age-standardised prevalence for each country (gender proportions were highly similar), using direct standardisation to the UK ONS population. This was also used to calculate overall age-standardised prevalence for each of the 10 regions of England.

It is important to note that coronavirus disease 2019 (COVID-19), a respiratory illness from a newly discovered coronavirus infection (SARS-CoV-2), was declared as a pandemic in early 2020. A 'lockdown' in the UK to contain the spread of SARS-CoV-2 meant that the population could not attend general practices or hospitals for non-urgent medical care. In comparison with other years in this study (2000-2019), only January to July follow up data were contributed in 2020. This follow-up time coincided with the declaration of the pandemic and national lockdown; it therefore may have affected diagnosis and follow-up of MC and thus measured incidence and prevalence.

2.4 Notes on sample size and guidance on interpretation of findings

This report used the largest available population-based sample in the United Kingdom and its devolved nations on which the population incidence and prevalence of diagnosed MC could be reliably estimated between the year 2000 and 2020. No sample size calculations were conducted.

Because we are calculating incidence and prevalence based on a representative sample of the population (not the whole population), these should be interpreted as estimates. When interpreting these, the point estimate (labelled `Incidence' or `Prevalence' in tables) should be read with its accompanying CI (labelled `95% confidence interval' in tables). The 95%CI provides a range of values

(from the lowest estimate to the highest estimate) that is compatible with a p-value of 0.05. A p-value of 0.05 indicates the calculated estimate range would have occurred by chance fewer than 5 times in 100. We are therefore confident that it provides a good estimate of incidence or prevalence.

When comparing incidence or prevalence estimates between groups (e.g., age, gender, countries) the 95% CI should be compared alongside the point estimates. We describe estimates as similar where they have overlapping CIs, even if the point estimates are slightly different. For example, when we describe prevalence in one group as being higher or lower compared with another group, we have observed that confidence intervals for the estimates do not overlap.

Some CIs will be wider than others (i.e., the incidence or prevalence estimate will be less precise) and this is because of the size of the sample population and cases available for a particular calculation (wider CIs will usually be from smaller sample populations).

For an incident rate ratio (IRR), its 95% confidence interval is interpreted as statistically significant when it excludes '1.00'. Because the IRR is a ratio measure comparing two rates, when it includes '1.00', this means that incidence rate estimates have overlapping CIs.

Examples of interpretation:
A point prevalence of 5% with a 95%CI of (2%-6%) indicates that we estimate the prevalence to be between 2% and 6% and we would consider this to be similar to a point prevalence of 6% with a 95%CI of (1%-8%) because the CIs are overlapping. However, the latter is a less precise estimate likely because it is calculated using a smaller sample population.

An incidence rate ratio of 2.2 is calculated by dividing 2 rates, as follows: incidence 22 (in females) / incidence 10 (in males) = IRR 2.2. The incidence in females may have a confidence interval that overlaps with that in males, however, so the 95% CI of the IRR may cross '1.00' indicating no statistically significant difference in incidence between females and males (e.g., IRR 2.2, 95% CI 0.85-2.60).

In this report, the sample population of England is considerably larger than those of Scotland, Wales and Northern Ireland (mirroring the true relative populations sizes of these countries), so CIs for most estimates are narrowest for England. Estimates for Northern Ireland show the least precision as this is the smallest sample population, yet it is still calculated using a baseline population of >549,580 people; >342,839 in 2020.

3 Description of report data supplement

This report contains the incidence and prevalence tables and graphs for the UK population with an appendix (Appendix B) that includes the baseline sample population numbers and case numbers on which incidence and prevalence were calculated. **The accompanying data supplements contain all tables in the report and additional tables (including separate tables for England, England sub-group with HES-linked data, Scotland, Wales and Northern Ireland**):

datasheet_UoNreport2021_MC_incidence.xlsm
datasheet_UoNreport2021_MC_prevalence.xlsm

4 Findings on study population representativeness

A total of 38.3 million people (38,323,590) comprised the baseline study population followed up over time for calculating incidence between 2000 and 2020. A total of 16.1 million people (16,087,994) comprised the baseline mid-year population for calculating prevalence in 2020. Equivalent figures for young people under age 18 were 7.6 million followed up over time and 3.3 million in 2020.

The age, gender, and geographic breakdown for the total UK population is shown in Appendix B and is available in the data supplements by country. Population pyramids (Appendix C) of the age distributions for CPRD males and females are very similar to the total ONS UK, England, Scotland, Wales and Northern Ireland populations, showing excellent representativeness of the CPRD population for use in estimating disease prevalence.

5 Incidence of microscopic colitis in the United Kingdom 2000-2020

Between 2000 and 2020, 10,226 people received a new diagnosis of microscopic colitis during 286,673,368 person-years of active general practice registration in the population, representing a population incidence rate of 3.57 new cases per 100,000 person-years (i.e., 1 new case of microscopic colitis identified among every 28,011 people if they were each followed for 1 year) (Table 5-1).

Table 5-1 shows incidence rates stratified by different population characteristics with unadjusted and adjusted incidence rate ratios. Adjusted incidence was 2.5 times higher in females compared with males. Incidence generally increased with age and was highest in 70-79 year olds. Incidence increased considerably over time and was lower in areas with more socioeconomic deprivation.

The overall incidence rates for England, Scotland, Wales and Northern Ireland were 3.67, 3.19, 2.72, and 2.43 new cases per 100,000 person-years respectively. Adjusted incidence was lower in Wales compared with England. Incidence patterns across characteristics within devolved nations were generally similar to those in the UK overall, however, socioeconomic patterns were different in each devolved nation (Tables in data supplement).

Within England, adjusted incidence overlapped for most regions; London and the lowest adjusted relative incidence, and Yorkshire & The Humber and the North East had the highest (Table in data supplement).

Table 5-1 Incidence of microscopic colitis (United Kingdom 2000-2020)

Characteristic	Cases	Person-years	Incidence rate*	95% confidence interval	Unadjusted incidence rate ratio	95% confidence interval	Adjusted** incidence rate ratio	95% confidence interval
Age (years)								
<30	218	95239532	0.23	0.20 - 0.26	0.24	0.20 - 0.28	0.23	0.19 - 0.27
30-39	409	42359281	0.97	0.88 - 1.06	1.00		1.00	-
40-49	1019	43528518	2.34	2.20 - 2.49	2.42	2.16 - 2.72	2.37	2.11 - 2.65
50-59	1945	38578499	5.04	4.82 - 5.27	5.22	4.69 - 5.81	4.94	4.44 - 5.50
60-69	2736	30487048	8.97	8.64 - 9.32	9.29	8.38 - 10.31	8.68	7.82 - 9.63
70-79	2560	21934847	11.67	11.23 - 12.13	12.09	10.89 - 13.42	11.10	10.00 - 12.33
80&over	1339	14545642	9.21	8.73 - 9.71	9.53	8.53 - 10.65	8.23	7.36 - 9.19
Gender								
Male	2807	143935069	1.95	1.88 - 2.02	1.00	-	1.00	-
Female	7419	142738299	5.20	5.08 - 5.32	2.67	2.55 - 2.78	2.46	2.36 - 2.57
Calender period								
2000 - 2004	148	60862947	0.24	0.21 - 0.29	0.10	0.08 - 0.12	0.10	0.08 - 0.12
2005 - 2009	1740	70878250	2.45	2.34 - 2.57	1.00		1.00	-
2010 - 2014	3017	73314589	4.12	3.97 - 4.26	1.68	1.58 - 1.78	1.66	1.56 - 1.76
2015 - 2020	5321	81617583	6.52	6.35 - 6.70	2.66	2.52 - 2.80	2.58	2.44 - 2.72
Socioeconomic deprivation quintile (general practice-level)								
1-least deprived	2322	48806977	4.76	4.57 - 4.96	1.00	-	1.00	-
2	1956	50055578	3.91	3.74 - 4.08	0.82	0.77 - 0.87	0.83	0.78 - 0.88
3	2038	54283859	3.75	3.59 - 3.92	0.79	0.74 - 0.84	0.83	0.78 - 0.88
4	2019	65485221	3.08	2.95 - 3.22	0.65	0.61 - 0.69	0.73	0.68 - 0.77
5-most deprived	1891	68041733	2.78	2.66 - 2.91	0.58	0.55 - 0.62	0.71	0.67 - 0.75
Country								
England	9059	246652536	3.67	3.60 - 3.75	1.00	-	1.00	-
Scotland	646	20266082	3.19	2.95 - 3.44	0.87	0.80 - 0.94	0.96	0.87 - 1.07
Wales	379	13919497	2.72	2.46 - 3.01	0.74	0.67 - 0.82	0.80	0.71 - 0.91
Northern Ireland	142	5835254	2.43	2.06 - 2.87	0.66	0.56 - 0.78	0.84	0.71 - 1.00
Total population	10226	286673368	3.57	3.50 - 3.64				

*Incidence rate = newly diagnosed cases per 100,000 person-years contributed by the baseline population

**Incidence rate ratio adjusted for all varibles in the table and whether people's health records were from CPRD GOLD or CPRD Aurum

For the English sub-group population with HES-linked records, the incidence rate was 3.68 new cases per 100,000 person-years (i.e., 1 new case of microscopic colitis identified among every 27,174 people if they were each followed for 1 year) which was similar to the estimate for England overall (Table 5-2). Compared with the UK population and the overall English population, patterns of incidence were similar across age, gender, time and socioeconomic groups. Those with White ethnicity recorded had a higher incidence compared with all other ethnicity groups.

Table 5-2 Incidence of microscopic colitis (England HES-linked sub-population 2000-2020)

Characteristic	Cases	Person-years	Incidence rate*	95% confidence interval	Unadjusted incidence rate ratio	95% confidence interval	Adjusted** incidence rate ratio	95% confidence interval
Age (years)								
<30	162	70182191	0.23	0.20 - 0.27	0.25	0.20 - 0.30	0.23	0.19 - 0.28
30-39	293	31185388	0.94	0.84 - 1.05	1.00	-	1.00	-
40-49	704	31836554	2.21	2.05 - 2.38	2.35	2.05 - 2.70	2.19	1.91 - 2.51
50-59	1453	28188409	5.15	4.90 - 5.43	5.49	4.84 - 6.22	4.74	4.18 - 5.38
60-69	2098	22227892	9.44	9.04 - 9.85	10.05	8.89 - 11.35	8.13	7.19 - 9.19
70-79	1969	16069341	12.25	11.72 - 12.81	13.04	11.54 - 14.74	9.99	8.83 - 11.30
80&over	1061	10457341	10.15	9.55 - 10.78	10.80	9.49 - 12.29	7.61	6.68 - 8.66
Gender								
Male	2157	105868295	2.04	1.95 - 2.13	1.00	-	1	-
Female	5583	104278820	5.35	5.22 - 5.50	2.63	2.50 - 2.76	2.34	2.22 - 2.45
Calender period								
2000 - 2004	103	45735953	0.23	0.19 - 0.27	0.09	0.08 - 0.11	0.09	0.08 - 0.11
2005 - 2009	1249	50925196	2.45	2.32 - 2.59	1.00	-	1.00	-
2010 - 2014	2226	53160179	4.19	4.02 - 4.36	1.71	1.59 - 1.83	1.70	1.59 - 1.82
2015 - 2020	4162	60325787	6.90	6.69 - 7.11	2.81	2.64 - 3.00	2.81	2.63 - 2.99
Ethnicity* **								
White	7261	151630918	4.79	4.68 - 4.90	1.00	-	1.00	-
Mixed	24	2306641	1.04	0.70 - 1.55	0.22	0.15 - 0.32	0.56	0.37 - 0.83
Asian	111	11150991	1.00	0.83 - 1.20	0.21	0.17 - 0.25	0.38	0.31 - 0.45
Black	50	6603962	0.76	0.57 - 1.00	0.16	0.12 - 0.21	0.31	0.23 - 0.41
Other ethnicity	55	3466104	1.59	1.22 - 2.07	0.33	0.25 - 0.43	0.57	0.44 - 0.74
Unknown	239	34988499	0.68	0.60 - 0.78	0.14	0.13 - 0.16	0.26	0.23 - 0.29
Socioeconomic deprivation quintile (household-level)								
1-least deprived	2226	45635922	4.88	4.68 - 5.08	1.00	-	1.00	-
2	1864	42511554	4.38	4.19 - 4.59	0.90	0.85 - 0.96	0.98	0.92 - 1.04
3	1548	41288728	3.75	3.57 - 3.94	0.77	0.72 - 0.82	0.91	0.85 - 0.98
4	1170	41501509	2.82	2.66 - 2.99	0.58	0.54 - 0.62	0.83	0.77 - 0.90
5-most deprived	925	38886454	2.38	2.23 - 2.54	0.49	0.45 - 0.53	0.79	0.72 - 0.86
	7	322948	2.17	1.03 - 4.55	0.44	0.21 - 0.93	0.83	0.40 - 1.75
Socioeconomic deprivation quintile (general practice-level)								
1-least deprived	1771	35118332	5.04	4.81 - 5.28	1.00	-	1.00	-
2	1533	37571864	4.08	3.88 - 4.29	0.81	0.76 - 0.87	0.86	0.80 - 0.93
3	1601	40784609	3.93	3.74 - 4.12	0.78	0.73 - 0.83	0.92	0.86 - 0.99
4	1453	48093300	3.02	2.87 - 3.18	0.60	0.56 - 0.64	0.83	0.77 - 0.90
5-most deprived	1382	48579011	2.84	2.70 - 3.00	0.56	0.53 - 0.61	0.85	0.78 - 0.93
Urban-rural household location								
Urban	6305	183261153	3.44	3.36 - 3.53	1	-	1.00	-
Rural	1435	26885962	5.34	5.07 - 5.62	1.55	1.46 - 1.64	1.03	0.97 - 1.10
Region of England								
North East	430	7894621	5.45	4.96 - 5.99	1.00	-	1.00	-
North West	1289	35494560	3.63	3.44 - 3.84	0.67	0.60 - 0.74	0.67	0.60 - 0.74
Yorkshire & The Humber	449	8376977	5.36	4.89 - 5.88	0.98	0.86 - 1.12	1.00	0.88 - 1.15
East Midlands	222	5729497	3.87	3.40 - 4.42	0.71	0.60 - 0.84	0.79	0.67 - 0.93
West Midlands	1130	33085251	3.42	3.22 - 3.62	0.63	0.56 - 0.70	0.60	0.54 - 0.68
East of England	551	11521169	4.78	4.40 - 5.20	0.88	0.77 - 1.00	0.88	0.77 - 1.00
South West	1109	27192999	4.08	3.85 - 4.33	0.75	0.67 - 0.84	0.71	0.63 - 0.80
South Central	1184	26186719	4.52	4.27 - 4.79	0.83	0.74 - 0.93	0.79	0.70 - 0.89
London	639	35820609	1.78	1.65 - 1.93	0.33	0.29 - 0.37	0.49	0.43 - 0.56
South East Coast	737	18844713	3.91	3.64 - 4.20	0.72	0.64 - 0.81	0.65	0.57 - 0.73
Total population	7740	210147115	3.68	3.60 - 3.77				

*Incidence rate = newly diagnosed cases per 100,000 person-years contributed by the baseline population

**Incidence rate ratio adjusted for all varibles in the table and whether people's health records were from CPRD GOLD or CPRD Aurum

***The option 'Other ethnicity' can be selected on government ethnicity data collection forms. 'Unknown' is where no ethnicity information was coded in the person's general practice record or in their hospital record

HES = Hospital Episode Statistics (hospital inpatient and outpatient admissions)

Among young people under age 18 between 2000 and 2020, only 27 young people received a new diagnosis of microscopic colitis among 7.6 million followed-up across the UK, which was deemed too

few to calculate stable estimates of incidence by characteristics. Cases were distributed across all characteristics.

6 Prevalence of microscopic colitis in the United Kingdom (2020)

We identified 8,575 individuals with a diagnosis of microscopic colitis among a baseline population of 16,087,994, which represented a prevalence of 0.05%, equating to one in every 2,000 people with a diagnosis of microscopic colitis in the UK (Table 6-1).

Prevalence was considerably higher in females (0.08%) compared with males (0.03%) and increased with age for both groups. Prevalence decreased marginally with increasing socioeconomic deprivation in the UK and England but not in the devolved nations.

Age-standardised prevalence was marginally higher in England compared with the other devolved nations; when considered by gender, however, the higher prevalence was only among females (Table in data supplement). Within England, age-standardised prevalence varied modestly between regions, with London having the lowest prevalence and Yorkshire & The Humber having the highest.

Table 6-1 Prevalence of microscopic colitis per 100 population (United Kingdom 2020)

	Whole population		Females		Males	
Characteristic	**Prevalence**	**(95% confidence interval)**	**Prevalence**	**(95% confidence interval)**	**Prevalence**	**(95% confidence interval)**
Age at midyear (years)						
0-9	<0.01	(<0.01-<0.01)	<0.01	(<0.01-<0.01)	<0.01	(<0.01-<0.01)
10-19	<0.01	(<0.01-<0.01)	<0.01	(<0.01-<0.01)	<0.01	(<0.01-<0.01)
20-29	<0.01	(<0.01-<0.01)	0.01	(0.01-0.01)	<0.01	(<0.01-<0.01)
30-39	0.01	(0.01-0.01)	0.02	(0.01-0.02)	0.01	(0.00-0.01)
40-49	0.03	(0.03-0.03)	0.04	(0.04-0.05)	0.01	(0.01-0.02)
50-59	0.07	(0.06-0.07)	0.10	(0.09-0.11)	0.03	(0.03-0.04)
60-69	0.12	(0.12-0.13)	0.19	(0.18-0.20)	0.06	(0.05-0.06)
70-79	0.20	(0.19-0.21)	0.27	(0.26-0.29)	0.11	(0.11-0.12)
80&over	0.21	(0.20-0.22)	0.26	(0.25-0.27)	0.15	(0.13-0.16)
Gender						
Male	0.03	(0.03-0.03)				
Female	0.08	(0.08-0.08)				
Socioeconomic deprivation quintile (general practice-level)						
1-least deprived	0.07	(0.07-0.08)	0.11	(0.10-0.11)	0.04	(0.03-0.04)
2	0.06	(0.06-0.06)	0.09	(0.08-0.09)	0.03	(0.03-0.03)
3	0.06	(0.05-0.06)	0.08	(0.08-0.09)	0.03	(0.03-0.03)
4	0.04	(0.04-0.05)	0.07	(0.06-0.07)	0.02	(0.02-0.02)
5-most deprived	0.04	(0.04-0.04)	0.06	(0.06-0.07)	0.02	(0.02-0.02)
Country						
England	0.06	(0.05-0.06)	0.08	(0.08-0.08)	0.03	(0.03-0.03)
Scotland	0.05	(0.04-0.05)	0.07	(0.06-0.07)	0.02	(0.02-0.03)
Wales	0.04	(0.04-0.04)	0.06	(0.05-0.07)	0.02	(0.02-0.03)
Northern Ireland	0.04	(0.03-0.04)	0.05	(0.04-0.06)	0.03	(0.02-0.03)
Total population	0.05	(0.05-0.05)	0.08	(0.08-0.08)	0.03	(0.03-0.03)

… no cases available for prevalence calculation

Prevalence in the English sub-group population with HES-linked records was the same as in the overall English population (Table 6-2). Those with recorded White ethnicity had a higher prevalence compared with all other ethnicity groups and prevalence was higher in rural compared with urban areas.

Table 6-2 Prevalence of microscopic colitis per 100 population (England HES-linked sub-population 2020)

	Whole population		Females		Males	
Characteristic	**Prevalence**	**(95% confidence interval)**	**Prevalence**	**(95% confidence interval)**	**Prevalence**	**(95% confidence interval)**
Age at midyear (years)						
0-9	<0.01	(<0.01-<0.01)	<0.01	(<0.01-<0.01)	...	
10-19	<0.01	(<0.01-<0.01)	<0.01	(<0.01-<0.01)	<0.01	(<0.01-<0.01)
20-29	<0.01	(<0.01-<0.01)	0.01	(0.01-0.01)	<0.01	(<0.01-<0.01)
30-39	0.01	(0.01-0.01)	0.02	(0.01-0.02)	0.01	(0.01-0.01)
40-49	0.03	(0.03-0.03)	0.04	(0.04-0.05)	0.01	(0.01-0.02)
50-59	0.06	(0.06-0.07)	0.10	(0.09-0.11)	0.03	(0.03-0.04)
60-69	0.13	(0.12-0.14)	0.20	(0.18-0.21)	0.06	(0.06-0.07)
70-79	0.21	(0.20-0.22)	0.29	(0.28-0.31)	0.12	(0.11-0.13)
80&over	0.22	(0.21-0.24)	0.27	(0.25-0.29)	0.16	(0.14-0.18)
Gender						
Male	0.03	(0.03-0.03)				
Female	0.08	(0.08-0.08)				
Ethnicity*						
White	0.07	(0.07-0.07)	0.10	(0.10-0.11)	0.04	(0.04-0.04)
Mixed	0.01	(0.01-0.02)	0.01	(0.01-0.02)	0.01	(0.01-0.02)
Asian	0.01	(0.01-0.01)	0.01	(0.01-0.02)	0.01	(0.01-0.01)
Black	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.00	(0.00-0.01)
Other ethnicity	0.02	(0.01-0.02)	0.02	(0.02-0.03)	0.01	(0.01-0.02)
Unknown	0.01	(0.01-0.02)	0.03	(0.02-0.03)	0.01	(0.01-0.01)
Socioeconomic deprivation quintile (household-level)						
1-least deprived	0.08	(0.07-0.08)	0.11	(0.10-0.12)	0.04	(0.04-0.05)
2	0.07	(0.06-0.07)	0.10	(0.09-0.10)	0.03	(0.03-0.04)
3	0.05	(0.05-0.06)	0.08	(0.08-0.09)	0.03	(0.02-0.03)
4	0.04	(0.04-0.04)	0.06	(0.05-0.06)	0.02	(0.02-0.02)
5-most deprived	0.03	(0.03-0.03)	0.05	(0.04-0.05)	0.02	(0.02-0.02)
Socioeconomic deprivation quintile (general practice-level)						
1-least deprived	0.08	(0.07-0.08)	0.11	(0.11-0.12)	0.04	(0.04-0.05)
2	0.06	(0.06-0.06)	0.09	(0.08-0.09)	0.03	(0.03-0.03)
3	0.06	(0.06-0.06)	0.08	(0.08-0.09)	0.03	(0.03-0.04)
4	0.04	(0.04-0.04)	0.06	(0.06-0.07)	0.02	(0.02-0.02)
5-most deprived	0.04	(0.04-0.04)	0.06	(0.06-0.07)	0.02	(0.02-0.02)
Urban-rural household location						
Urban	0.05	(0.05-0.05)	0.07	(0.07-0.08)	0.03	(0.02-0.03)
Rural	0.09	(0.08-0.09)	0.13	(0.12-0.14)	0.05	(0.04-0.05)
Region of England						
North East	0.08	(0.08-0.09)	0.13	(0.12-0.15)	0.04	(0.03-0.05)
North West	0.05	(0.05-0.06)	0.08	(0.08-0.09)	0.03	(0.03-0.03)
Yorkshire & The Humber	0.09	(0.08-0.10)	0.12	(0.11-0.14)	0.05	(0.04-0.06)
East Midlands	0.06	(0.05-0.07)	0.08	(0.07-0.10)	0.03	(0.02-0.04)
West Midlands	0.05	(0.05-0.06)	0.08	(0.08-0.09)	0.03	(0.02-0.03)
East of England	0.07	(0.07-0.08)	0.11	(0.09-0.12)	0.04	(0.03-0.05)
South West	0.06	(0.06-0.07)	0.09	(0.09-0.10)	0.03	(0.03-0.04)
South Central	0.07	(0.06-0.07)	0.10	(0.09-0.11)	0.03	(0.03-0.04)
London	0.02	(0.02-0.03)	0.04	(0.03-0.04)	0.01	(0.01-0.02)
South East Coast	0.06	(0.06-0.07)	0.09	(0.08-0.09)	0.03	(0.03-0.04)
Total population	0.05	(0.05-0.06)	0.08	(0.08-0.08)	0.03	(0.03-0.03)

*The option 'Other ethnicity' can be selected on government ethnicity data collection forms. 'Unknown' is where no ethnicity information was coded in the person's general practice record or in their hospital record

HES = Hospital Episode Statistics

Among young people under age 18 in 2020, we identified fewer than 20 young people with a prevalent microscopic colitis diagnosis, which was deemed too few to calculate stable estimates of prevalence.

7 Appendices

Appendix A Code lists for defining cases

Microscopic colitis CPRD GOLD	Description	CPRD medcode	Read code
	Microscopic colitis Collagenous colitis Lymphocytic colitis	30678 39119 35424	J436.00 J436000 J436100

Microscopic colitis CPRD Aurum	Description	CPRD medcode	SNOMED CODE
	CC - Collagenous colitis Collagenous colitis Lymphocytic colitis Lymphocytic-plasmacytic colitis Microscopic colitis	2806491000006113 32563019 52549017 3006581000006119 353418019	19311003 19311003 31437008 31437008 235753003

Appendix B Numbers of people in case and baseline populations

Table 7-1 Baseline population and people with new diagnoses of microscopic colitis in the United Kingdom (Incidence analyses)

	Whole population				Females				Males			
Characteristic	**Incident microscopic colitis**		**Baseline population**		**Incident microscopic colitis**		**Baseline population**		**Incident microscopic colitis**		**Baseline population**	
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
Age (years)												
<30	218	2.1	14,421,918	37.6	168	2.3	7,502,136	38.5	50	1.8	6,919,782	36.8
30-39	409	4.0	6,439,278	16.8	295	4.0	3,204,496	16.4	114	4.1	3,234,782	17.2
40-49	1,019	10.0	4,672,977	12.2	774	10.4	2,190,912	11.2	245	8.7	2,482,065	13.2
50-59	1,945	19.0	4,059,458	10.6	1,482	20.0	1,953,338	10.0	463	16.5	2,106,120	11.2
60-69	2,736	26.8	3,189,193	8.3	2,008	27.1	1,542,538	7.9	728	25.9	1,646,655	8.8
70-79	2,560	25.0	2,612,518	6.8	1,792	24.2	1,323,099	6.8	768	27.4	1,289,419	6.9
80&over	1,339	13.1	2,928,248	7.6	900	12.1	1,790,853	9.2	439	15.6	1,137,395	6.0
Gender												
Male	2,807	27.4	18,816,218	49.1								
Female	7,419	72.6	19,507,372	50.9								
Socioeconomic deprivation quintile (general practice-level)												
1-least deprived	2,322	22.7	6,283,878	16.4	1,695	22.8	3,224,439	16.5	627	22.3	3,059,439	16.3
2	1,956	19.1	6,764,882	17.7	1,437	19.4	3,461,152	17.7	519	18.5	3,303,730	17.6
3	2,038	19.9	7,158,849	18.7	1,433	19.3	3,663,673	18.8	605	21.6	3,495,176	18.6
4	2,019	19.7	9,118,082	23.8	1,495	20.2	4,636,402	23.8	524	18.7	4,481,680	23.8
5-most deprived	1,891	18.5	8,997,899	23.5	1,359	18.3	4,521,706	23.2	532	19.0	4,476,193	23.8
Country												
England	9,059	88.6	33,687,424	87.9	6,572	88.6	17,156,307	87.9	2,487	88.6	16,531,117	87.9
Scotland	646	6.3	2,422,620	6.3	481	6.5	1,226,261	6.3	165	5.9	1,196,359	6.4
Wales	379	3.7	1,663,966	4.3	271	3.7	844,264	4.3	108	3.8	819,702	4.4
Northern Ireland	142	1.4	549,580	1.4	95	1.3	280,540	1.4	47	1.7	269,040	1.4
Total population	10,226	100.0	38,323,590	100.0	7,419	100.0	19,507,372	100.0	2,807	100.0	18,816,218	100.0

Table 7-2 Total midyear* baseline population and people with diagnosed microscopic colitis** in the United Kingdom (Prevalence analyses)

	Whole population				Females				Males			
Characteristic	**microscopic colitis diagnosis**		**Midyear population**		**microscopic colitis diagnosis**		**Midyear population**		**microscopic colitis diagnosis**		**Midyear population**	
	Number	*%*	Number	*%*	Number	*%*	Number	*%*	Number	*%*	Number	*%*
Age at midyear (years)												
0-9	<5		1713807	*10.7*	<5		835987	*10.4*	<5		877820	*10.9*
10-19	16	*0.2*	1853712	*11.5*	10	*0.2*	908023	*11.3*	6	*0.3*	945689	*11.8*
20-29	106	*1.2*	2208415	*13.7*	77	*1.2*	1113497	*13.8*	29	*1.3*	1094918	*13.6*
30-39	272	*3.2*	2360920	*14.7*	197	*3.1*	1163884	*14.5*	75	*3.4*	1197036	*14.9*
40-49	606	*7.1*	2125316	*13.2*	456	*7.2*	1029171	*12.8*	150	*6.7*	1096145	*13.6*
50-59	1419	*16.5*	2169646	*13.5*	1067	*16.8*	1063799	*13.2*	352	*15.8*	1105847	*13.7*
60-69	2026	*23.6*	1633317	*10.2*	1548	*24.4*	817484	*10.2*	478	*21.4*	815833	*10.1*
70-79	2492	*29.1*	1258967	*7.8*	1810	*28.5*	660266	*8.2*	682	*30.6*	598701	*7.4*
80&over	1634	*19.1*	763894	*4.7*	1176	*18.5*	452849	*5.6*	458	*20.5*	311045	*3.9*
Gender												
Male	2231	*26.0*	8043034	*50.0*								
Female	6344	*74.0*	8044960	*50.0*								
Socioeconomic deprivation quintile (general practice-level)												
1-least deprived	2039	*23.8*	2785112	*17.3*	1517	*23.9*	1409337	*17.5*	522	*23.4*	1375775	*17.1*
2	1695	*19.8*	2925918	*18.2*	1270	*20.0*	1468387	*18.3*	425	*19.0*	1457531	*18.1*
3	1690	*19.7*	3010200	*18.7*	1230	*19.4*	1514386	*18.8*	460	*20.6*	1495814	*18.6*
4	1637	*19.1*	3717660	*23.1*	1220	*19.2*	1846660	*23.0*	417	*18.7*	1871000	*23.3*
5-most deprived	1514	*17.7*	3649104	*22.7*	1107	*17.4*	1806190	*22.5*	407	*18.2*	1842914	*22.9*
Country												
England	7507	*87.5*	13587624	*84.5*	5559	*87.6*	6786898	*84.4*	1948	*87.3*	6800726	*84.6*
Scotland	587	*6.8*	1284409	*8.0*	440	*6.9*	647373	*8.0*	147	*6.6*	637036	*7.9*
Wales	352	*4.1*	873122	*5.4*	259	*4.1*	439572	*5.5*	93	*4.2*	433550	*5.4*
Northern Ireland	129	*1.5*	342839	*2.1*	86	*1.4*	171117	*2.1*	43	*1.9*	171722	*2.1*
Total population	8575	*100.0*	16087994	*100.0*	6344	*100.0*	8044960	*100.0*	2231	*100.0*	8043034	*100.0*

*midyear=In cohort on July 1 2020 **diagnosis in general practice record before or on July 1 2020

Appendix C Population pyramids of age and gender distribution

The following population pyramids compare the CPRD baseline population (2020 mid-year population of 16,087,944) with the Office of National Statistics 2019 mid-year population counts for the UK (66,796,807), England (13,587,624), Scotland (1,284,409), Wales (873,122), and Northern Ireland (342,839). *Note that scales on graphs are different; x-axis label is persons rather than millions.*

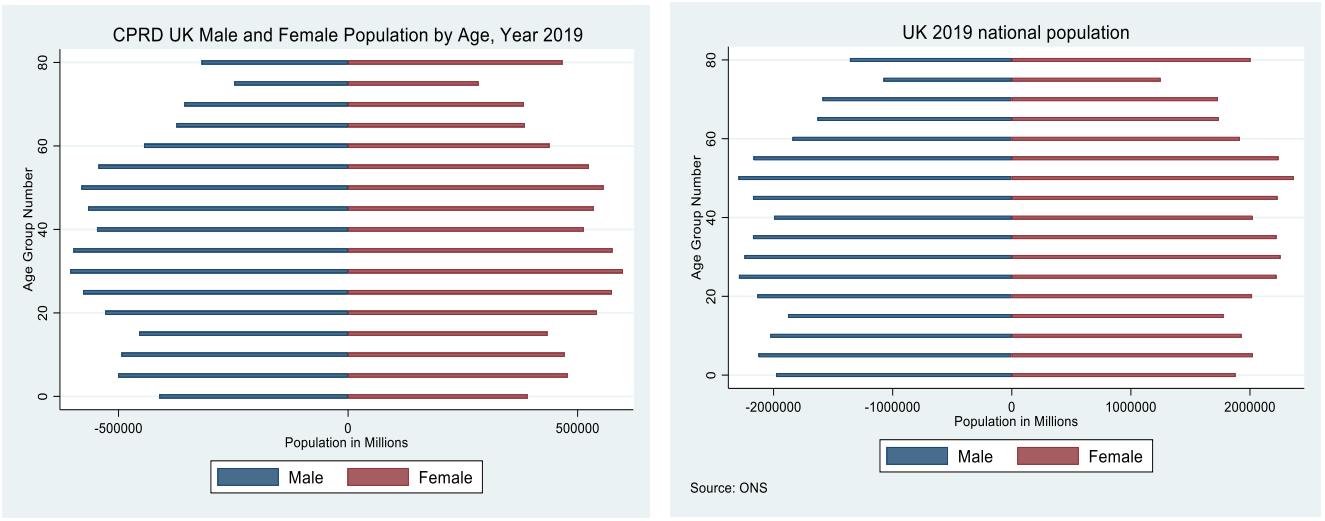


Figure 7-1 Population pyramids: United Kingdom CPRD population (left) and ONS population (right)

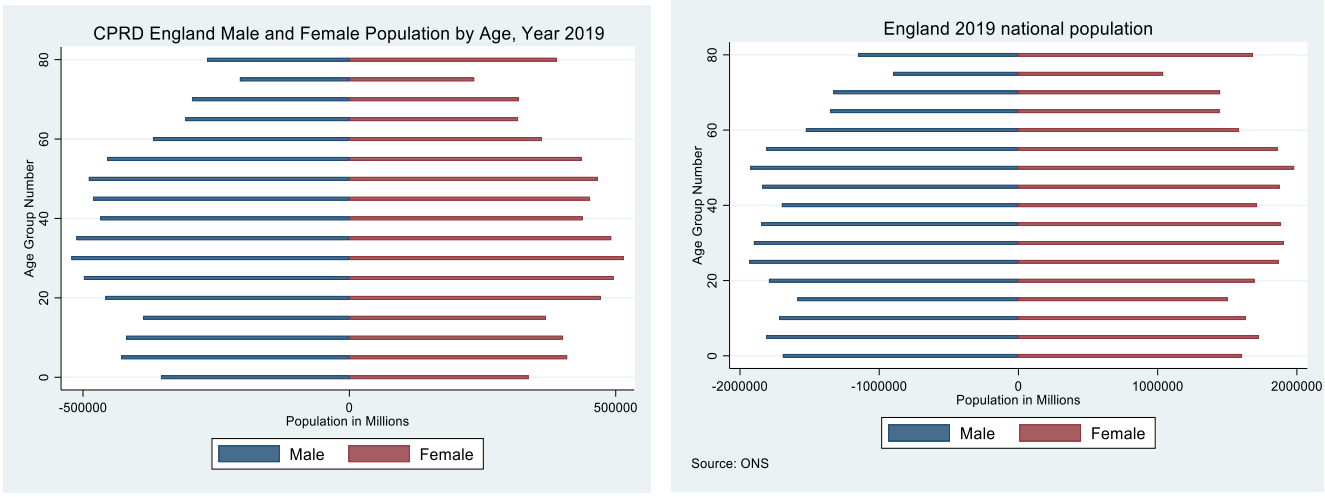
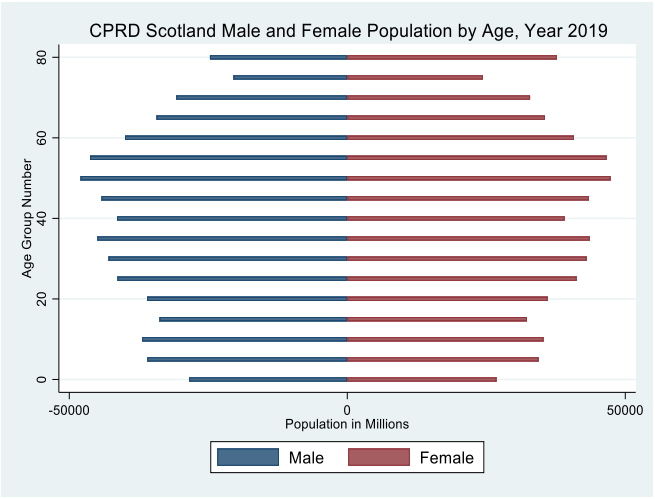


Figure 7-2 Population pyramids: England CPRD population (left) and ONS population (right)



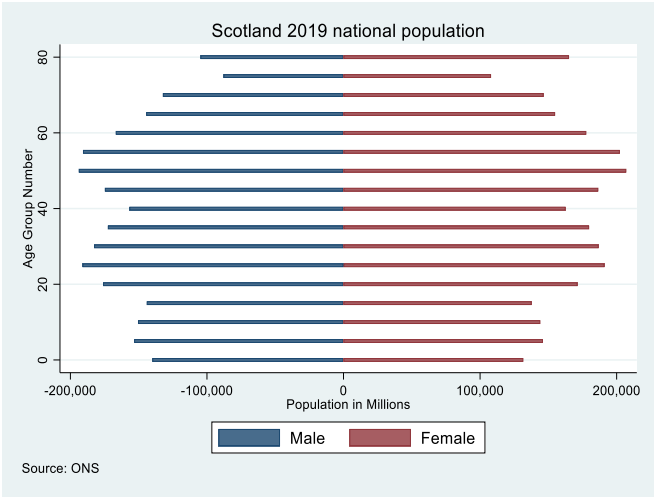
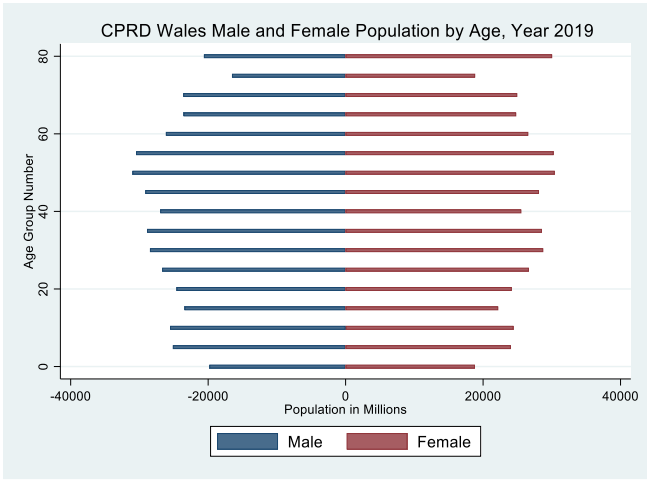


Figure 7-3 Population pyramids: Scotland CPRD population (left) and ONS population (right)



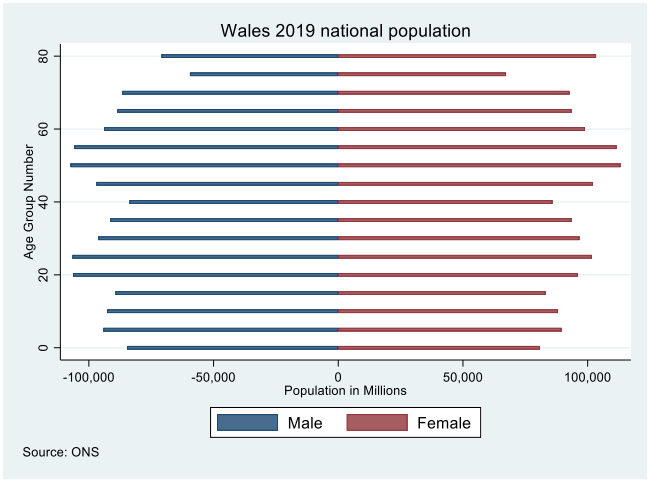
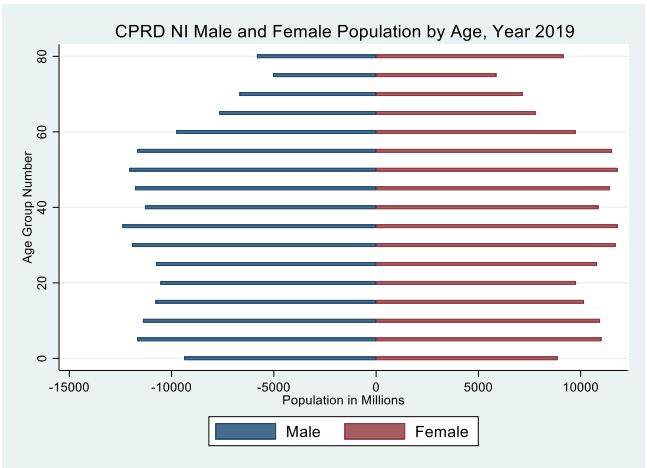


Figure 7-4 Population pyramids: Wales CPRD population (left) and ONS population (right)



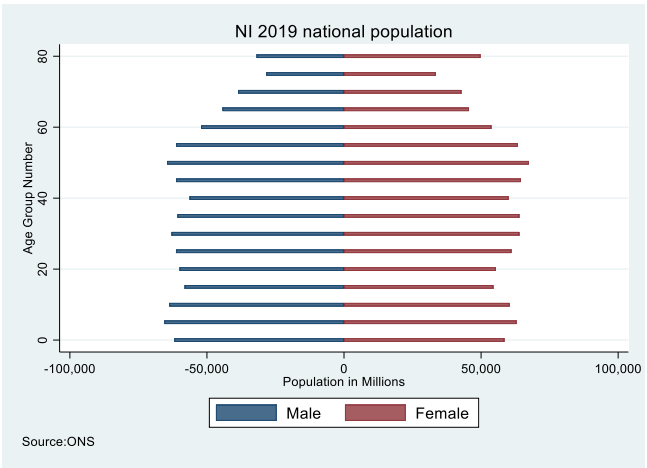


Figure 7-5 Population pyramids: Northern Ireland CPRD population (left) and ONS population (right)