

Supplementary Table 1: ATC, ICD, SNOMED, and topography codes used for identification of study populations*

Journal Pre-proof

	TOPOGRAPHY	SNOMED	ICD9 [¶]	ICD10	ATC
Antibiotics					J01
Crohn's disease			555	K50	
Celiac disease	T64x, T65x (excluding T652, T65200, T653, T65300, T654, T656, T65600, T65620, T6592, T6593)	M58, M5800, M58000, M58001, M58005, M58006, M58007, M58000X, D6218, D62180, D6218X, D6218Y, D62188			
Inflammatory bowel disease			556, 555	K50, K51, K52.3	
Microscopic colitis	T67x, T68x	M40600x, M47170x			
Non-steroidal anti-inflammatory drugs					M01A
Proton pump inhibitors					A02BC
Serotonin Reuptake Inhibitors					N06AB
Statins					C10AA
Steroids (budesonide)					A07EA06
Ulcerative colitis			556	K51	

*Abbreviation: Anatomical therapeutic chemical classification (ATC), International classification of disease (ICD), Systematized Nomenclature of Medicine (SNOMED). ¶ ICD9 started in 1987. Since the first patients with microscopic colitis was included in 1990, all study participants had ≥3 years of exposure in the Patient Register prior to identification of comorbidities.

Supplementary Table 2: Definition of types of gastrointestinal infection using ICD coding

Type of gastroenteritis	Infectious agent	ICD-7	ICD-8	ICD-9	ICD-10
Infectious diseases of the GI tract	Any	040-049	000-009	001-009	A00-A09
Bacterial infections	Typhoid/Paratyphoid	040-041	001-002	002	A01
	<i>Salmonella</i>	042	003	003	A02
	<i>Shigella</i>	045	004	004	A03
	<i>Escherichia coli</i>	-	008,0	008A	A04.0-A04.4
	<i>Campylobacter</i>	-	-	008E	A04.5
	<i>Yersinia Enterocolitica</i>	-	-	008E	A04.6
	<i>Clostridium difficile</i>	-	-	008E	A04.7
	Other bacterial infection	-	008,2-008,3	008F	A04 (except A04.0-A04.7)
	Bacterial food poisoning	049	005	005	A05
Parasitic infections	Amoeba	046	006	006	A06
	Giardia	047,02	007,10	007B	A07.1
	Cryptosporidium	-	-	-	A07.2
	Isospora	-	-	-	A07.3
	Other protozoic GI infections	047 (except 047,02)	007 (except 007,.10)	007 (except than 007B)	A07 (except A07.1-A07.3)
Viral infections	Rotavirus	-	-	008L	A08.0
	Norovirus	-	-	-	A08.1
	Adenovirus	-	-	008H	A08.2
	Other viral GI infection	-	008,8-008,9	008M (except 008L, 008H)	A08 (except A08.0-A08.2)
Infectious disease, not otherwise defined	Not defined if viral or bacterial	048, 571, 764, 785,6	008 (except 008,0; 008,2; 008,3), 009	008, 008W, 009	A09

Supplementary Table 3: Definition of immune-mediated conditions using ICD coding and histopathology

Disease	ICD7	ICD8	ICD9	ICD10
<i>Diabetes</i>	260	250	250	E10-14
<i>Psoriasis</i>	706,09	696 (969,30 excluded)	696A-C + E-W	L40
<i>SLE</i>	456,20	734,1	710A	M32.1; M32.8; M32.9
<i>Rheumatoid arthritis</i>	722,00 + 722,01	712,3; 714,93	714	M05; M06; M08; M09; M12.3
<i>Thyroiditis</i>	254	245	245A-X	E06
<i>Hyperthyroidism</i>	252.00, 252.01, 252.02, 254.00;	242	242	E05
<i>Sarcoidosis</i>	138,00 to 138,04	135	135	D86
<i>Primary biliary cirrhosis</i>			571G	K74.3
<i>ANCA vasculitis and other vasculitis</i>		446	446	M31
<i>Spondylopathy</i>				M45.9
<i>Autoimmune hepatitis</i>		573,0x (chronic hepatitis) 571,9x	573D + 573E	K75.4; K75.5; K75.9 (for exclusion criteria, if AIH had been the outcome we would have used only K75.4)
<i>Primary sclerosing cholangitis</i>				K83.0A
<i>Celiac disease</i>		Defined according to histopathology		

Supplementary Table 4: Characteristics of cases and their unaffected siblings*

	Population controls N=13,699	Microscopic colitis N=7091	Collagenous colitis N=2216	Lymphocytic colitis N=4875
Sex, female	6882 [50.2]	5115 [72.1]	1705 [76.9]	3410 [70.0]
Age at start of follow up				
Mean [SD] years	54.2 [14.6]	54.1 [15.2]	56.9 [13.1]	52.8 [15.9]
Median [IQR] years	57.0 [46.0-65.0]	57.0 [45.0-65.0]	59.0 [50.0-66.0]	56.0 [42.0-65.0]
< 50 years	4370 [31.9]	2317 [32.7]	553 [25.0]	1764 [36.2]
≥ 50 years	9329 [68.1]	4774 [67.3]	1663 [75.1]	3111 [63.8]
Country of birth				
Nordic	13515 [98.7]	7005 [98.8]	2205 [99.5]	4800 [98.5]
Other	184 [1.3]	86 [1.2]	11 [0.5]	75 [1.5]
Years of diagnosis				
1990 - 2006	3890 [28.4]	1951 [27.5]	629 [28.4]	1322 [27.1]
2007 - 2011	4740 [34.6]	2425 [34.2]	751 [33.9]	1674 [34.3]
2012 - 2016	5069 [37.0]	2715 [38.3]	836 [37.7]	1879 [38.5]
History of gastroenteritis	538 [3.9]	543 [7.7]	158 [7.1]	385 [7.9]
Time from gastroenteritis to index date				
Mean [SD] years	17.1 [12.0]	15.4 [12.0]	15.25 [12.3]	15.43 [11.9]
Median [IQR] years	14.8 [6.3-26.4]	12.3 [4.7-25.4]	11.42 [5.4-23.1]	12.47 [4.6-25.9]
0.5 < 1 years	7 [0.1]	24 [0.3]	8 [0.4]	16 [0.3]
1 < 3 years	48 [0.4]	63 [0.9]	17 [0.8]	46 [0.9]
3 < 5 years	57 [0.4]	51 [0.7]	14 [0.6]	37 [0.8]
> = 5 years	426 [3.1]	405 [5.7]	119 [5.4]	286 [5.9]
Education[¶]				
≤ 9 years	3371 [24.6]	1333 [18.8]	495 [22.3]	838 [17.2]
10-12 years	5956 [43.5]	3125 [44.1]	1024 [46.2]	2101 [43.1]
≥ 13 years	4065 [29.7]	2513 [35.4]	674 [30.4]	1839 [37.7]
Missing	307 [2.2]	120 [1.7]	23 [1.0]	97 [2.0]

*Abbreviations: Standard deviation (SD) and Interquartile range (IQR). ¶ Categories based on compulsory school, high school, and college.

Supplementary Table 5: Association between gastrointestinal infection and risk of MC according to histologic subtypes*

Exposures	Controls No exposed	CC No. exposed	CC OR, 95% CI	LC No exposed	LC OR, 95% CI	P _{heterogeneity}
Any	1920	358	3.12 [2.70-3.59]	672	2.42 [2.19-2.68]	0.005
Bacteria	502	94	2.84 [2.17-3.70]	201	2.85 [2.38-3.41]	0.992
<i>Salmonella sp</i>	52	5	1.43 [0.47-3.62]	13	1.66 [0.80-3.18]	0.813
<i>Escherichia sp</i>	8	2	8.66 [0.81-189.22]	4	3.01 [0.77-10.60]	0.450
<i>Clostridium difficile</i>	142	47	4.07 [2.70-6.12]	79	4.56 [3.33-6.25]	0.664
Parasites	34	6	3.28 [1.10-9.13]	12	2.14 [0.98-4.35]	0.509
Virus	167	41	3.23 [2.10-4.91]	49	2.19 [1.50-3.13]	0.172
Norovirus	40	8	2.66 [1.04-6.36]	14	2.96 [1.48-5.73]	0.849
Others	1327	272	3.49 [2.95-4.13]	486	2.43 [2.16-2.73]	<0.001

*Abbreviations: collagenous colitis (CC), confidence interval (CI), lymphocytic colitis (LC), microscopic colitis (MC), number (No.), odds ratio (OR), Models were adjusted for age, sex, county, calendar year, education, and comorbidities according to categories listed in Table 1.

Supplementary Table 6: Association between gastrointestinal infection and risk of MC according to histologic subtypes using unaffected sibling as comparators*

Exposures	Controls No exposed	CC No. exposed	CC OR, 95% CI	LC No. exposed	LC OR, 95% CI	P_{heterogeneity}
Any	538	158	1.79 [1.38-2.33]	385	1.81 [1.53-2.15]	0.997

Bacteria	138	39	2.29 [1.38-3.82]	102	1.63 [1.20-2.22]	0.904
<i>Salmonella sp</i>	15	2	1.00 [0.16-6.19]	7	1.36 [0.50-3.74]	0.854
<i>Escherichia sp</i>	5	0	-	2	0.81 [0.14-4.78]	
<i>Clostridium difficile</i>	35	15	4.57 [1.73-12.11]	38	2.22 [1.29-3.82]	0.887
Parasites	8	5	1.94 [0.47-8.03]	8	4.73 [1.09-0.61]	0.862
Virus	43	13	2.03 [0.81-5.03]	31	1.77 [0.98-3.21]	0.960
Norovirus	11	2	1.92 [0.29-12.92]	7	2.55 [0.73-8.92]	0.929
Others	379	121	1.73 [1.28-2.33]	279	1.97 [1.61-2.42]	0.960

*Abbreviations: collagenous colitis (CC), confidence interval (CI), lymphocytic colitis (LC), microscopic colitis (MC), number (No.), odds ratio (OR), Models were adjusted for age, sex, county, calendar year, education, and comorbidities according to categories listed in Table 1.