Risk of inflammatory bowel disease following hospitalisation with infectious mononucleosis: nationwide cohort study from Denmark

Ebert, Anthony C.; Harper, Shania; Vestergaard, Marie V.; Mitchell, Wayne; Jess, Tine; Elmahdi, Rahma

Published in:
Nature Communications

DOI (link to publication from Publisher):
10.1038/s41467-024-52195-8

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

nature communications

Article

https://doi.org/10.1038/s41467-024-52195-8

Risk of inflammatory bowel disease following hospitalisation with infectious mononucleosis: nationwide cohort study from Denmark

Received: 9 April 2024

Accepted: 27 August 2024

Published online: 27 September 2024

Check for updates

Anthony C. Ebert[1,4], Shania Harper[2,4], Marie V. Vestergaard[1], Wayne Mitchell[2], Tine Jess MD[1,3] & Rahma Elmahdi[1,3]

Infectious mononucleosis (IM) is suspected to be associated with inflammatory bowel disease (IBD) development. Using a Danish nationwide cohort of people developing severe IM and their age-, sex-, and socioeconomic (SES) index-matched counterparts, we investigated the subsequent risk of IBD, Crohn's disease (CD), or ulcerative colitis (UC) development from 1977 to 2021. Among 39,684 severe IM patients we find a sex-, age-, and SES index-adjusted HR for IBD of 1.35 (95% CI: 1.22–1.49). This significantly increased risk was seen for both CD (HR: 1.56; 95% CI: 1.34–1.83) and to a lesser extent UC (HR: 1.23; 95% CI: 1.08–1.40) and remains following negative control matching with a cohort diagnosed with *Chlamydia trachomatis* infection (HR: 1.39; 95% CI: 1.01–1.91). Those with severe IM at 0–9 years had a particularly increased risk for CD (HR: 1.77; 95% CI: 1.26–2.49). Here we show an increased risk for IBD diagnosis following IM hospitalisation, indicating an association between severe EBV disease and later IBD development. Further exploration of the potential factors contributing to IBD susceptibility following EBV disease is warranted.

Inflammatory bowel disease (IBD), commonly diagnosed as either Crohn's disease (CD) or ulcerative colitis (UC), is a progressive immune-mediated inflammatory disorder (IMID) of the gastrointestinal (GI) tract, affecting >6.8 million individuals worldwide[1]. IBD is most prevalent in high-income countries whilst its incidence is most rapidly increasing in newly industrialised countries[2]. Approximately 0.5% of the Danish population had an IBD diagnosis in 2010 and by 2020, this rose to almost 0.75% as IBD prevalence increased by 63.3%[3], supporting the projection that 1% of individuals within high-income countries will have an IBD diagnosis by 2030[1]. The pathophysiology of IBD involves complex genetic, environmental, epithelial, microbial, and immune factors[4,5] that result in immune dysfunction leading to GI barrier dysregulation, and dysbiosis of the gut microbiota. Environmental exposures such as smoking, pro-inflammatory foods, and antibiotic use do not fully account for why IBD develops in genetically susceptible individuals[6,7]. Infectious agents, as either contributing to or protecting from the development of IBD, have largely been associated with disease in the context of GI-specific pathogens, including *Salmonella* and *Campylobacter* infection[8–10], or *Helicobacter pylori* infection in the induction of immune tolerance and reduced risk of IBD[6,11]. The association between systemic viral infections and IBD development has yet to be systematically explored.

Infectious mononucleosis (IM) is a clinical syndrome of fever, tonsillar pharyngitis, and lymphadenopathy, commonly (>90% of cases) due to acute infection with Epstein–Barr Virus (EBV)[12,13]. Although usually a benign, self-limiting condition, in approximately 5%

[1]PREDICT Center for Molecular Prediction of Inflammatory Bowel Disease, Department of Clinical Medicine, Aalborg University, Copenhagen, Denmark. [2]Department of Immunology and Inflammation, Imperial College London, London, UK. [3]Department of Gastroenterology and Hepatology, Aalborg University Hospital, Aalborg, Denmark. [4]These authors contributed equally: Anthony C. Ebert, Shania Harper. e-mail: rahma.elmahdi@sund.ku.dk

of acute infections, IM is associated with severe disease presenting with high fever, mesenteric adenitis, and extreme fatigue, which can require hospitalisation[14,15]. Associations between IM and several IMIDs including Sjögren's syndrome, rheumatoid arthritis, systemic lupus erythematosus[16–18], and multiple sclerosis (MS)[19,20] have previously been identified and the putative role of EBV in MS development has been shown most recently in a longitudinal analysis of a cohort of more than ten million American military personnel, undertaken by Bjornevik et al. [21].

Preliminary evidence from a study using a bacteriophage immunoprecipitation sequencing technique, like that used by Bjornevik et al., on a cohort of 119 military service personnel diagnosed with IBD and healthy matched controls, found that exposure to EBV was significantly associated with CD, supporting a biological association with EBV exposure and IBD development[22]. More recently, Loosen et al. [23], undertook a retrospective cohort study using data from a subset of German primary care physicians in the Disease Analyzer database to explore the association between general practitioner-reported IBD diagnosis following IM diagnosis. Although authors observed a significant association between IM and subsequent IBD, several limitations in study design, including short follow-up time, small cohort size, and lack of population-representative data (only 3% of German primary care practices were included in the analysis), make interpretation of these study findings challenging.

Therefore, to explore the association between severe IM and the development of IBD, we undertook a nationwide, population-representative cohort study using data from the Danish national health registries. The primary outcome was the diagnosis of IBD, including CD and UC following hospitalisation with IM. We additionally explored whether previous hospitalisation with IM impacted the severity of disease in those going on to develop IBD.

Results

We identified 40,244 individuals from the general Danish population between 1st January 1977 and 31st December 2021 with an IM hospitalisation (see Supplementary Table 1 for a complete list of diagnostic codes). Two-hundred eleven were excluded from the severe IM cohort based on a prevalent IBD diagnosis, with a further 349 excluded due to residence outside Denmark in the 10 years prior to IM hospitalisation (see Fig. 1 for flow diagram of study cohort inclusion). A total of 39,684 individuals were included in the severe IM cohort and were matched to 396,840 individuals without an IM diagnosis from the general Danish population, based on sex, calendar year of birth, and socioeconomic (SES) index of municipality of residence (SES index) at date of IM diagnosis (index date). These contributed over 8,010,000 person-years of follow-up time to the analyses. There was a comparable proportion of males to females within the severe IM cohort (52.6% male) and most individuals were diagnosed with IM at 15–19 years of age (40.6%). Only 4.6% of those with severe IM had a complication associated with their IM hospitalisation (including hepatosplenomegaly, neuropathy, and splenic rupture). Thirty-one and six-tenths per cent of total severe IM cases were diagnosed between 2010 and 2021 (see Table 1 for baseline characteristics of the cohort).

A total of 496 (incidence rate (IR): 0.67/1000 persons) and 3982 (IR: 0.50/1000 persons) IBD cases were diagnosed in the severe IM and matched cohort, respectively. The median time to IBD diagnosis in the severe IM cohort was 19.3 years compared with 22.1 years in the matched cohort (see Fig. 2 for the cumulative incidence of IBD, CD, and UC). The crude hazard ratio (HR) for IBD diagnosis following severe IM was 1.35 (95% CI: 1.23–1.48; see Table 2 for HR for IBD following severe IM). Following adjustment for sex, age, and calendar year of diagnosis, this increased risk of IBD remained (HR: 1.35; 95% CI: 1.22–1.49) and was particularly seen in CD (HR: 1.56; 95% CI: 1.34–1.83) compared with UC (HR: 1.23; 95% CI: 1.08–1.40). Clustered analysis with robust standard errors showed no difference in HR or confidence interval to three significant figures (see Supplementary Table 2).

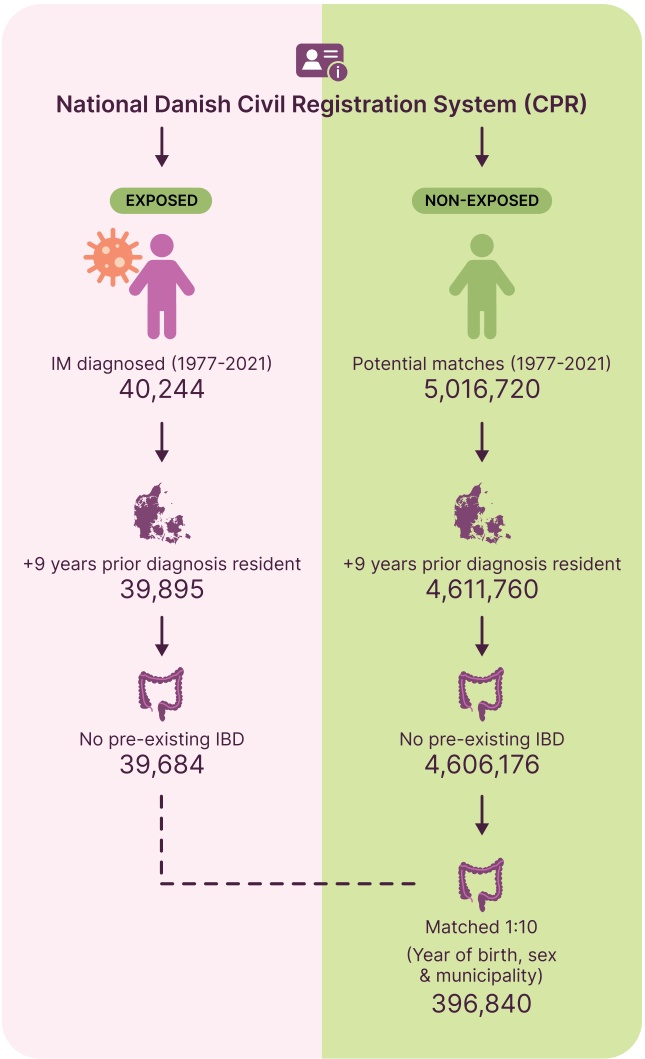

Fig. 1 | Cohort inclusion flow chart for severe IM and matched cohort. Exposed—severe IM cohort. Those with an ICD-10 coded, secondary care (hospital) diagnosis of IM between 1st January 1977 and 31st December 2021, derived from the National Patient Register (LPR), resident in Denmark for ≥9 years prior to diagnosis, derived from the National Danish Civil Registration System (CPR), and no record of pre-existing IBD, derived from the LPR. Linkage to non-exposed using the CPR. Non-exposed—matched cohort. Those without an ICD-10 coded hospital diagnosis of IM between 1st January 1977 and 31st December 2021, derived from the LPR, resident in Denmark for ≥9 years prior to diagnosis, derived from the CPR, and no record of pre-existing IBD, derived from the LPR. Randomly matched by sex (female or male), age (within the same birth year), and SES index for the municipality of residence at a 10:1 ratio to exposed. IM infectious mononucleosis, IBD inflammatory bowel disease.

This association was found in both females (HR: 1.36; 95% CI: 1.20–1.55) and males (HR: 1.34; 95% CI: 1.17–1.54) but only significantly associated with IBD in those diagnosed with IM at 15–19 years (HR: 1.41; 95% CI: 1.22–1.62), 20–24 years (HR: 1.41; 95% CI: 1.08–1.82), and 25–29 years (HR: 1.60; 95% CI: 1.09–2.33). A particularly increased risk of CD diagnosis following severe IM at 0–9 years was also observed (HR: 1.77; 95% CI: 1.26–2.49), which was not observed for UC or overall IBD (see Figs. 3 and 4 for HR for CD and UC development following severe IM, respectively). The cumulative incidence of IBD stratified by age group is presented in Fig. 5.

https://doi.org/10.1038/s41467-024-52195-8

Table 1 | Baseline characteristics of severe IM and matched cohort at the time of IM diagnosis

Characteristic	Severe I, (*N* = 39,684)	Matched cohort, (*N* = 396,840)
Age at IM diagnosis (years)		
0–9	6,234 (15.7%)	62,136 (15.7%)
10–14	6,183 (15.6%)	62,745 (15.8%)
15–19	16,102 (40.6%)	159,985 (40.3%)
20–24	5,386 (13.6%)	54,033 (13.6%)
25–29	2,056 (5.2%)	20,552 (5.2%)
≥30	3,723 (9.4%)	37,389 (9.4%)
Calendar year of birth		
1897–1939	324 (0.8%)	3,240 (0.8%)
1940–1959	2,263 (5.7%)	22,630 (5.7%)
1960–1979	14,848 (37.4%)	148,480 (37.4%)
1980–1999	16,560 (41.7%)	165,600 (41.7%)
2000–2020	5,689 (14.3%)	56,890 (14.3%)
Calendar year of IM diagnosis (index date)		
1977–1989	9,932 (25.0%)	99,320 (25.0%)
1990–1999	8,816 (22.2%)	88,160 (22.2%)
2000–2009	8,390 (21.1%)	83,900 (21.1%)
2010–2021	12,546 (31.6%)	125,460 (31.6%)
Sex		
Female	18,807 (47.4%)	188,070 (47.4%)
Male	20,877 (52.6%)	208,770 (52.6%)
SES index*		
Low	10,098 (41.2%)	100,980 (41.2%)
Medium	8,041 (32.8%)	80,410 (32.8%)
High	6,365 (26.0%)	63,650 (26.0%)
Duration of IM hospitalisation (days)		
0–7	34,684 (87.4%)	–
8–14	3,973 (10.0%)	–
>15	1,027 (2.6%)	–
Complications of IM diagnosis	878 (4.6%)	–

Complications of IM diagnosis include hepatosplenomegaly, neuropathy, or splenic rupture within 1 month of IM diagnosis. Severe IM cases include individuals hospitalised with IM between 1st January 1977 and 31st December 2021. The matched cohort includes individuals matched from the general Danish population on sex (male or female), date of birth (same birth year), and municipality of residence without IM hospitalisation.
IM infectious mononucleosis, *SES* socioeconomic status.
*SES index by tertile threshold: High = 0.00–0.48; medium = 0.48–0.58; low = 0.58–1.00 (see Supplementary Table 7 for calculation of SES by the municipality of residence).

We also observe an increased risk for IBD development in those diagnosed with severe IM over time (1977–1989 HR: 1.21; 95% CI: 1.03–1.43, 1990–1999 HR: 1.30; 95% CI: 1.10–1.54, 2000–2009 HR: 1.40; 95% CI: 1.15–1.70, 2010–2021 HR: 1.84; 95% CI: 1.45–2.35, $P = 0.049$) following adjustment for age at IM diagnosis and despite similar baseline numbers of IM hospitalisations in the two periods (1977–1989 = 9,930 vs 2010–2021 = 12,546). However, the total follow-up time for IBD development was considerably shorter for those diagnosed with IM between 2010–2021 (7.79 per 100,000 person-years) compared with 1977–1989 (39.59 per 100,00 person-years). Following restriction of calendar time of follow-up to 10 years after IM diagnosis for all calendar year of diagnosis categories, this calendar time effect is lost (10-year post-diagnosis HR 1977–1989: 1.62; 95% CI: 1.06–2.47, 1990–1999: 1.54; 95% CI: 1.17–2.03, 2000–2009: 1.60; 95% CI: 1.26–2.04, 2010–2021: 1.86; 95% CI: 1.46–2.38, $P = 0.97$; see Supplementary Table 3).

We went on to assess the risk of severe IBD disease, in those diagnosed with IBD following IM diagnosis, compared with the matched cohort. We found that the risk of any IBD-related hospitalisation, corticosteroid use, surgery, use of immunomodulators, or use of biologics (see Supplementary Table 4 for diagnostic codes for IBD severity score) was not significantly increased following IM diagnosis (HR: 1.05; 95% CI: 0.9–1.19). This was the case in those diagnosed with both CD (HR: 1.14; 95% CI: 0.98–1.32) and UC (HR: 0.96; 95% CI: 0.84–1.11). With regards to the individual IBD severity outcomes, there was no significant association with hospitalisation (HR: 1.21; 95% CI: 0.94–1.55), corticosteroid use (HR: 0.84; 95% CI: 0.71–1.00), surgery (HR: 1.08; 95% CI: 0.9–1.31), immunomodulator use (HR: 1.16; 95% CI: 0.97–1.39), or biologics use (HR: 0.95; 95% CI: 0.79–1.13; see Supplementary Table 5 for risk of severe IBD following severe IM).

To assess the impact of unidentified bias in the primary analysis, we undertook negative control matching of the severe IM cohort with a cohort diagnosed with *Chlamydia trachomatis* (CT) infection. We identified a total of 10,877 individuals in the IM cohort suitable for negative-control matching to a total of 10,797 individuals who were not hospitalised with IM and had a CT diagnosis (see Supplementary Fig. 1 for flow chat of IM and CT negative-control cohort inclusion). Propensity score (PS) weighting was based on sex, age at diagnosis, date of diagnosis, and SES index (see Supplementary Fig. 2 for PS weight distributions). Baseline characteristics of the IM and CT total and weighted cohort at the time of diagnosis are presented in Supplementary Table 6. There were 99 (IR: 0.40/1,000 persons) and 67 (0.29/1,000 persons) IBD cases in the IM and CT cohorts, respectively. Following adjustment for sex, age, calendar year of diagnosis, and SES municipality of residence, the risk of IBD following IM diagnosis compared with CT diagnosis remained significantly increased (HR: 1.39; 95% CI: 1.01–1.91). This risk was seen in both CD (HR: 1.40; 95% CI: 0.84–2.33) and UC (HR: 1.39; 95% CI: 0.91–2.09). See Supplementary Table 7. However, this association was not significant by subtype, likely due to the fewer events observed (see Fig. 6 for the cumulative incidence of IBD, CD, and UC in severe IM compared to CT). Results from this negative control analysis indicate that the effect observed in our main analysis (i.e., a significant association with IBD development following severe IM), remains following comparison with a similar disease cohort.

Finally, we used the Danish nationwide Register of Laboratory Results for Research (RLRR)[24] to assess whether there was a difference in IBD diagnosis in confirmed EBV-positive IM cases that were hospitalised (severe IM), compared to confirmed EBV-positive IM cases who were not hospitalised by identifying positive test results for EBV capsid antigen IgM. We limited the analysis to samples taken at Køge, Vejle, and Esbjerg hospitals to ensure complete test coverage for a given population area and identified 6,169 individuals with a positive test. Of these, 854 were hospitalised with IM around the time of testing positive (+/− 31 days) compared with 5,315 with a positive test who were not hospitalised, giving a hospitalisation fraction of 0.14 of IM cases in the area. 26 of those hospitalised with IM following positive EBV test went on to develop IBD. Survival analysis showed an increased HR: 1.21 (95% CI: 0.42–3.51) for IBD diagnosis following hospitalisation with a positive test compared to not being hospitalised with a positive test, however, this was not found to be significant.

Discussion

In this population-representative, nationwide cohort study we follow almost 40,000 patients hospitalised with IM (severe IM) and over 396,000 sex-, age-, and SES index-matched individuals from the general Danish population for the development of IBD over 8 million person-years to identify a 35% increased risk of IBD development following severe IM. This is seen particularly for CD (56% increased risk) but also present for UC (23% increased risk). There is no difference in risk based on sex, age at IM diagnosis, or calendar year of IM diagnosis,

https://doi.org/10.1038/s41467-024-52195-8

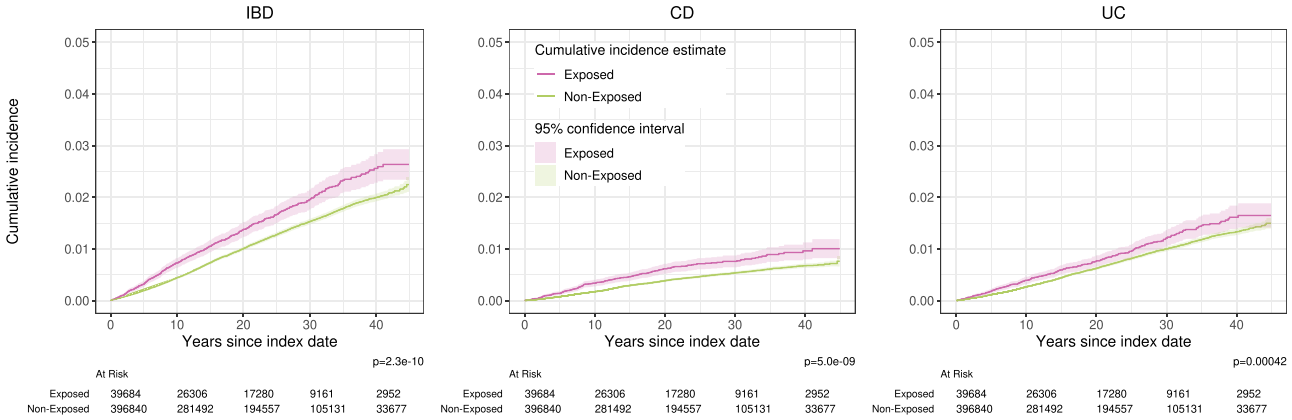

Fig. 2 | Cumulative incidence of IBD, CD, and UC diagnosis in severe IM compared with the matched cohort. Exposed—severe IM cohort. Those with an ICD-10 coded, secondary care (hospital) diagnosis of IM between 1st January 1977 and 31st December 2021. Non-exposed—matched cohort. Those without an ICD-10 coded hospital diagnosis of IM between 1st January 1977 and 31st December 2021 matched by sex, age, and SES index of the municipality to exposed. Index date—date of severe IM diagnosis in exposed or cohort inclusion in non-exposed, which is the start of follow-up. At risk—number of exposed and unexposed at risk of IBD development at 0, 10-, 20-, 30-, and 40-years post follow-up. Cumulative incidence *P*-value calculated using single-sided log-rank testing. IM infectious mononucleosis, IBD inflammatory bowel disease, CD Crohn's disease, UC ulcerative colitis.

however, we observe a particularly increased risk of CD (77%) in those developing severe IM between 0 years and 9 years. We did not identify a significant association between hospitalisation with IM and later development of severe IBD (including IBD-related hospitalisation, surgery, or treatment escalation). The association between the IBD and severe IM remains significantly increased with a similar effect size (39% increased risk) after negative control matching with a cohort diagnosed with *C. trachomatis* infection.

The main findings of this study indicate that hospitalisation with IM is associated with a significantly increased risk of IBD development. These findings are in keeping with those recently published by Loosen et al. who identify an increased risk of IBD following primary care diagnosis with IM in a subset of German practices at the same rate as that of this current study (HR: 1.35; 95% CI: 1.01–1.81)[23]. Although no other cohort studies to date have explored the association between IM and IBD development, the consistent findings between these two large cohorts indicate a possible underlying pathophysiological contribution of IM disease to IBD development.

Most IM cases are caused by the common and highly transmissible gamma herpes virus, EBV, with seroconversion occurring at the highest rates in early childhood and adolescence[13]. This high prevalence of seropositivity at an early period in life makes identifying putative associations between EBV and subsequent IBD development highly challenging, hence the focus of this study on severe IM or a subset of the severe clinical manifestation of EBV infection. To overcome the challenge of high EBV seroprevalence in detecting association with IBD, Nandy et al. [22] have attempted to apply unbiased, comprehensive serological viral screening to a small longitudinal cohort of 42 IBD patients with pre- and post-diagnostic serum samples. Authors show that EBV exposure levels are significantly associated with CD patients compared with healthy controls (CD = 78.2% vs healthy controls=53.3%, $P = 0.045$), suggesting a pathogenic role for EBV in either the onset or progression of CD. Furthermore, longitudinal analysis on the same cohort indicated this association may be putative as increased levels of EBV exposure were present as early as 5 years before CD diagnosis[22]. We also observed a particularly strong association with CD in those developing severe IM at 0–9 years, which was not present for UC or IBD overall. This may indicate that severe IM in this age group, although uncommon, may be related to a shared immunological response or susceptibility to failures in the programming of cellular immunity that also affect risk for CD. However, as the majority of people have symptom-free seroconversion when encountering EBV at this stage in life, we only identified 39 cases of severe IM going on to develop CD in this age group and the significance of this association remains unclear.

The vast majority of individuals are infected with EBV during childhood, resulting in a persistent and mostly latent infection. EBV infection most commonly manifests as IM when primary infection occurs in adolescence[15]. A Danish study assessing age-dependent variations both in IM attack rates and EBV seroconversion hazard rates found that 85% of the Danish population had seroconverted by age 30 years[14], with the peak incidence being in the 15–24 year age range. This is mirrored in the age distribution of the IM cases identified in the current study, where we also find the highest number of IBD cases in those diagnosed with IM in the 15–19-year age range for both CD and UC.

Interestingly, unlike Loosen et al., who identify this association between IM and later IBD development only for the CD subtype, we also find this increased risk is significantly associated with UC diagnosis. The presence of EBV-induced gene 3 and EBV-encoded small RNA1, indicative of the activity of EBV-infected cells, has previously been associated with both CD and UC, particularly during periods of inflammatory disease flare[25,26]. Additionally, recent exploration of gene regulatory regions and transcription factors using genome-wide association study catalogues for multiple autoimmune diseases identify particularly strong associations with EBV nuclear antigen 2 (EBNA2), an EBV-derived transcription factor, and UC ($p < 10^{-8}$). Supporting the potential role of genetic mechanisms dependent on EBNA2 in the induction of disease in individuals susceptible to UC[27].

Although linked to several cancers, including Hodgkin's Lymphoma, a high proportion of Burkitt's lymphomas, several nasopharyngeal carcinomas[28], and lymphoproliferative disorders associated with transplantation and immunosuppression[29], it is only in recent years that EBV and IM have been implicated in the development of complex, immune-mediated diseases[30–32]. Numerous mechanisms have been hypothesised for the potential contribution of EBV in the development of these diseases. Most recently, clonally expanded B-lymphocytes have been demonstrated to be cross-reactive for the central nervous system protein glial cell adhesion molecule and EBNA1, exposing their high-affinity molecular mimicry in MS[33,34]. Similarly, cross-reactive molecular mimicry may be at play in the immune tissue of the GI mucosa or against GI commensals in IBD[35,36]. Alternatively, under the immune exhaustion hypothesis, EBV infection in the acute phase could be exhausting the immune system's resources, potentially increasing vulnerability to IBD in genetically susceptible individuals as

https://doi.org/10.1038/s41467-024-52195-8

Table 2 | Crude incidence, crude HR, and adjusted HR for IBD, CD, and UC diagnosis following severe IM compared with the matched cohort, stratified by sex, age at IM diagnosis and year of IM diagnosis

	Severe IM, (*N* = 39,684)				Matched cohort, (*N* = 396,840)				Crude HR	Crude HR 95% CI	Crude *P*-value	Adjusted HR	Adjusted* HR 95% CI	Adjusted* *P*-value
	Total follow-up, (100,000 person-years)	IBD events, (*N*)	Total severe IM, (*N*)	IBD IR, (per 1,000 persons)	Total follow-up, (100,000 person-years)	IBD events, (*N*)	Total matched cohort, (*N*)	IBD IR, (per 1000 persons)						
IBD	7.42	496	39,684	0.67	80.1	3,982	396,840	0.5	1.35	1.23–1.48	2.6×10^{-10}	1.35	1.22–1.49	6.0×10^{-9}
CD	7.42	203	39,684	0.27	80.1	1,417	396,840	0.18	1.55	1.33–1.79	6.6×10^{-9}	1.56	1.34–1.83	2.5×10^{-8}
UC	7.42	293	39,684	0.40	80.1	2,565	396,840	0.32	1.24	1.10–1.40	4.3×10^{-4}	1.23	1.08–1.40	0.002
Sex														
Female	3.38	260	18,807	0.77	36.8	2,084	188,070	0.57	1.36	1.20–1.55	0.846	1.36	1.20–1.55	0.859
Male	4.03	236	20,877	0.58	43.3	1,898	208,770	0.44	1.34	1.17–1.53		1.34	1.17–1.54	
Age at IM diagnosis														
0–9	1.26	70	6,234	0.55	13.4	621	62,136	0.46	1.20	0.94–1.53	0.685	1.20	0.93–1.53	0.746
10–14	1.30	86	6,183	0.66	14.2	760	138,272	0.53	1.24	0.99–1.55		1.24	0.99–1.55	
15–19	2.95	214	16,102	0.72	32.3	1,562	159,043	0.52	1.40	1.22–1.62		1.41	1.22–1.62	
20–24	0.96	64	5,386	0.67	10.4	499	54,033	0.38	1.41	1.08–1.83		1.41	1.08–1.82	
25–29	0.38	31	2,056	0.82	4.0	205	20,552	0.40	1.59	1.09–2.33		1.60	1.09–2.33	
≥30	0.57	31	3,723	0.55	5.8	222	37,389	0.38	1.43	0.98–2.08		1.43	0.98–2.08	
Calendar year of IM diagnosis														
1977–1989	3.29	156	39,684	0.47	36.3	1,434	99,320	0.40	1.21	1.03–1.43	0.047	1.21	1.03–1.43	0.049
1990–1999	2.10	149	39,684	0.71	22.8	1,256	88,160	0.55	1.30	1.09–1.54		1.30	1.10–1.54	
2000–2009	1.33	114	39,684	0.86	14.0	865	83,900	0.62	1.39	1.15–1.69		1.40	1.15–1.70	
2010–2021	0.69	77	39,684	1.11	7.1	427	125,460	0.60	1.84	1.45–2.35		1.84	1.45–2.35	

Severe IM cases include individuals hospitalised with IM between 1st January 1977 and 31st December 2021. The matched cohort includes individuals matched from the general Danish population on sex (male or female), date of birth (same birth year), and SES index for the municipality of residence, without IM hospitalisation. HR was calculated using Cox proportional hazards regression modelling with a maximum likelihood estimate. *P*-values were calculated using a single-sided *Z*-test based on standard error.
IM infectious mononucleosis, *IR* incidence rate, *HR* hazard ratio, *95% CI* 95% confidence interval.
*Adjusted for sex, age, and SES by municipality of residence as appropriate.

https://doi.org/10.1038/s41467-024-52195-8

Subgroup		Exposed Events/N	Exposed IR	Non-Exposed Events/N	Non-Exposed IR	Adjusted HR (95% CI)	p value
Sex	Female	122/18807	0.36	824/188070	0.22	1.61 (1.33,1.95)	0.54
	Male	81/20877	0.20	593/208770	0.14	1.47 (1.16,1.85)	
Age at IM diagnosis	0–9	39/6234	0.31	232/62136	0.17	1.77 (1.26,2.49)	0.71
	10–14	34/6183	0.26	289/62745	0.20	1.28 (0.90,1.83)	
	15–19	82/16102	0.28	613/159985	0.19	1.46 (1.16,1.84)	
	20–24	27/5386	0.28	163/54033	0.16	1.81 (1.20,2.71)	
	25–29	10/2056	0.27	58/20552	0.15	1.82 (0.93,3.55)	
	30+	11/3723	0.19	62/37389	0.11	1.82 (0.96,3.45)	
Year of IM diagnosis	1977–1989	54/9932	0.16	450/99320	0.12	1.33 (1.01,1.77)	0.21
	1990–1999	62/8816	0.30	458/88160	0.20	1.48 (1.14,1.93)	
	2000–2009	48/8390	0.36	324/83900	0.23	1.57 (1.16,2.12)	
	2010–2021	39/12546	0.56	185/125460	0.26	2.16 (1.53,3.04)	

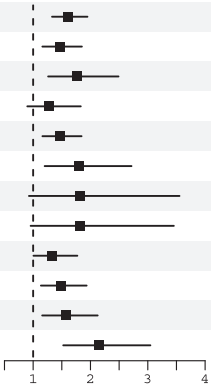

Fig. 3 | Crude incidence rates and adjusted HR for CD diagnosis in severe IM compared with the matched cohort stratified by sex, age at IM diagnosis and year of IM diagnosis. Exposed—severe IM cohort. Those with an ICD-10 coded, secondary care (hospital) diagnosis of IM between 1st January 1977 and 31st December 2021. Non-exposed—matched cohort. Those without an ICD-10 coded hospital diagnosis of IM between 1st January 1977 and 31st December 2021 matched by sex, age, and SES index of the municipality to exposed. Events—number of CD diagnoses over total follow-up time. *N*—total number of exposed and unexposed follow-ups for events. HR was calculated using Cox proportional hazards regression modelling with a maximum likelihood estimate. HR adjusted for sex, age, and SES by municipality of residence as appropriate. *P*-values were calculated using single-sided *Z*-tests based on standard error. Error bars represent 95% CI for HR estimate. Black squares at the centre of error bars denote the HR estimate. IM infectious mononucleosis, CD Crohn's disease, IR incidence rate, HR hazard ratio, 95% CI 95% confidence interval.

Subgroup		Exposed Events/N	Exposed IR	Non-Exposed Events/N	Non-Exposed IR	Adjusted HR (95% CI)	p value
Sex	Female	138/18807	0.41	1260/188070	0.34	1.20 (1.01,1.43)	0.59
	Male	155/20877	0.38	1305/208770	0.30	1.28 (1.09,1.52)	
Age at IM diagnosis	0–9	31/6234	0.25	389/62136	0.29	0.85 (0.59,1.22)	0.26
	10–14	52/6183	0.40	471/62745	0.33	1.22 (0.91,1.62)	
	15–19	132/16102	0.45	1062/159985	0.33	1.37 (1.14,1.64)	
	20–24	37/5386	0.39	336/54033	0.32	1.21 (0.86,1.70)	
	25–29	21/2056	0.56	147/20552	0.37	1.51 (0.95,2.38)	
	30+	20/3723	0.35	160/37389	0.28	1.28 (0.80,2.04)	
Year of IM diagnosis	1977–1989	102/9932	0.31	984/99320	0.27	1.16 (0.94,1.42)	0.44
	1990–1999	87/8816	0.41	798/88160	0.35	1.20 (0.96,1.49)	
	2000–2009	66/8390	0.50	541/83900	0.39	1.29 (1.00,1.67)	
	2010–2021	38/12546	0.55	242/125460	0.34	1.61 (1.14,2.26)	

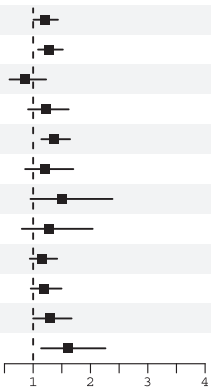

Fig. 4 | Crude incidence rates and adjusted HR for UC diagnosis in severe IM compared with the matched cohort stratified by sex, age at IM diagnosis, and year of IM diagnosis. Exposed—severe IM cohort. Those with an ICD-10 coded, secondary care (hospital) diagnosis of IM between 1st January 1977 and 31st December 2021. Non-exposed—matched cohort. Those without an ICD-10 coded hospital diagnosis of IM between 1st January 1977 and 31st December 2021 matched by sex, age, and SES index of the municipality to exposed. Events—number of UC diagnoses over total follow-up time. *N*—total number of exposed and unexposed follow-ups for events. HR was calculated using Cox proportional hazards regression modelling with a maximum likelihood estimate. HR adjusted for sex, age, and SES by municipality of residence as appropriate. *P*-values were calculated using single-sided *Z*-tests based on standard error. Error bars represent 95% CI for HR estimate. Black squares at the centre of error bars denote the HR estimate. IM infectious mononucleosis, UC ulcerative colitis, IR incidence rate, HR hazard ratio, 95% CI 95% confidence interval.

is also seen in MS with EBV-specific T-lymphocytes eventually becoming exhausted through cytokine polyfunctionality[37]. The EMBOLD study, a Phase 2 study of allogenic T-cell immunotherapy for MS was however recently terminated when the primary end-point of change in confirmed disability improvement among patients with non-active progressive MS was not met after 12 months of treatment[38] highlighting the importance of mechanistic elucidation for the development of potential drug targets in complex IMIDs.

IM is characterised by an aggressive peripheral $CD8^+$ T-cell lymphocytosis, that gives the disease its name. It is likely that it is the extent of the expansion of these activated T-cells that contributes to the clinical features of IM, and the magnitude of this response is the differentiating factor in those developing severe IM and being hospitalised. Individuals expressing certain human leucocyte antigen (HLA) alleles, including HLA-A2 and HLA-B8 are more likely to express high levels of circulating $CD8^+$ T-cells against EBV lytic proteins[39]. We know that the HLA region plays a key role in IBD disease pathophysiology, with this genetic region being where the most significant risk variants and the largest effect sizes for IBD susceptibility are located[40]. It is plausible, therefore, that HLA class II gene region alleles with HLA traits A and B implicated in IBD development might also be playing a role in the adaptive immune response to EBV and severe IM development.

A particular strength of this current study lies in the use of highly reliable population-representative, nationwide health registry data ensuring that all diagnoses of both IM and IBD captured are reliable. Danish registries allow for complete follow-up due to complete data on emigration and death, with no loss to follow-up when moving within the country. Due to nationwide registration used for state reimbursement, there is also a low risk of misclassification of outcomes. Our sensitivity analysis evaluating whether the risk of any EBV positive test

https://doi.org/10.1038/s41467-024-52195-8

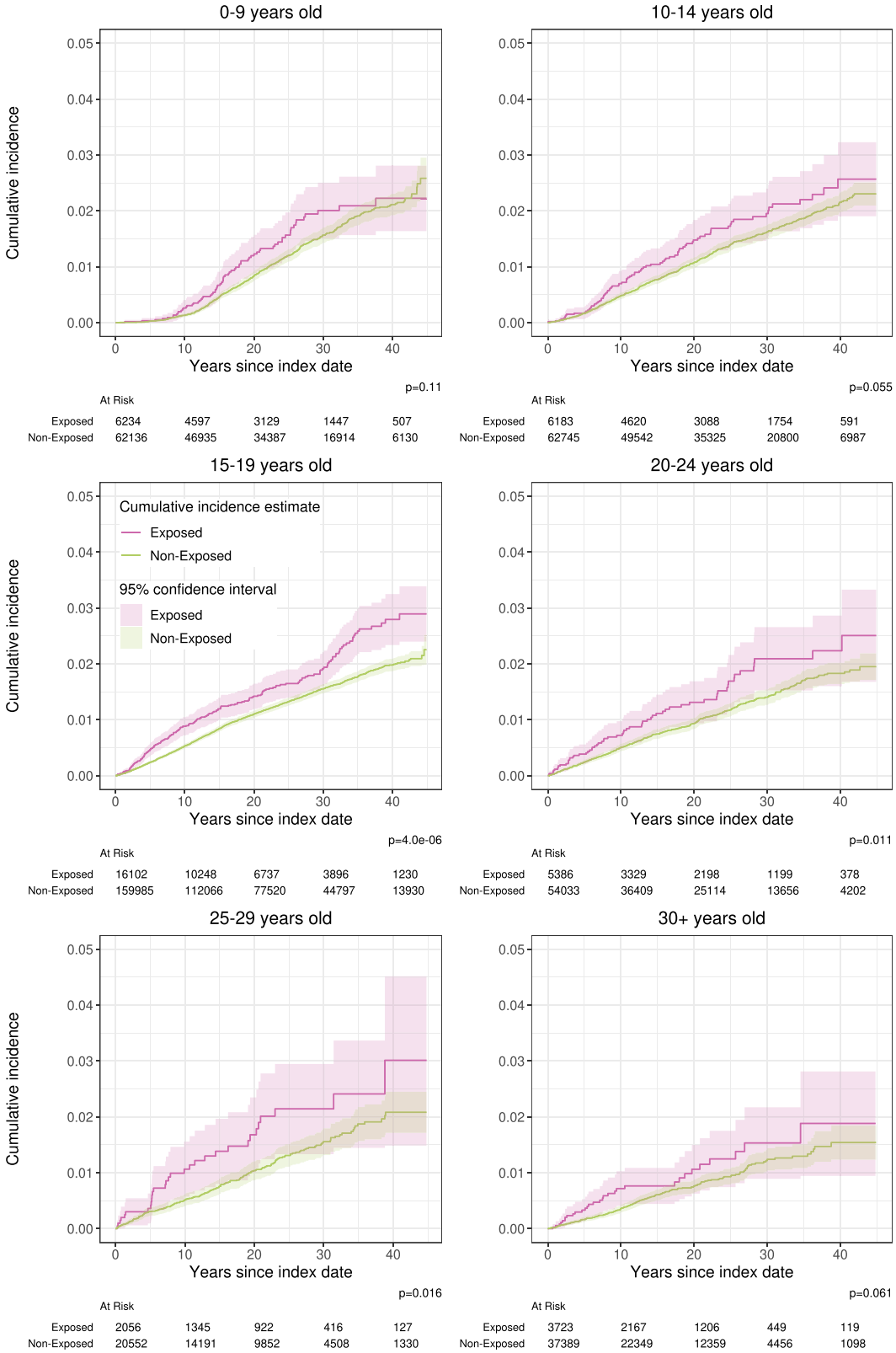

Fig. 5 | Cumulative incidence of IBD in severe IM compared with the matched cohort, stratified by age category (0–9, 10–14, 15–19, 20–24, 25–29 and ≥30-years-old). Exposed—severe IM cohort. Those with an ICD-10 coded, secondary care (hospital) diagnosis of IM between 1st January 1977 and 31st December 2021. Non-exposed—matched cohort. Those without an ICD-10 coded hospital diagnosis of IM between 1st January 1977 and 31st December 2021 matched by sex, age, and SES index of the municipality to exposed. Index date—date of severe IM diagnosis in exposed or cohort inclusion in non-exposed, which is the start of follow-up. At risk—number of exposed and unexposed at risk of IBD development at 0, 10-, 20-, 30-, and 40-years post follow-up. *P*-values were calculated using single-sided log-rank testing. IM infectious mononucleosis, IBD inflammatory bowel disease.

https://doi.org/10.1038/s41467-024-52195-8

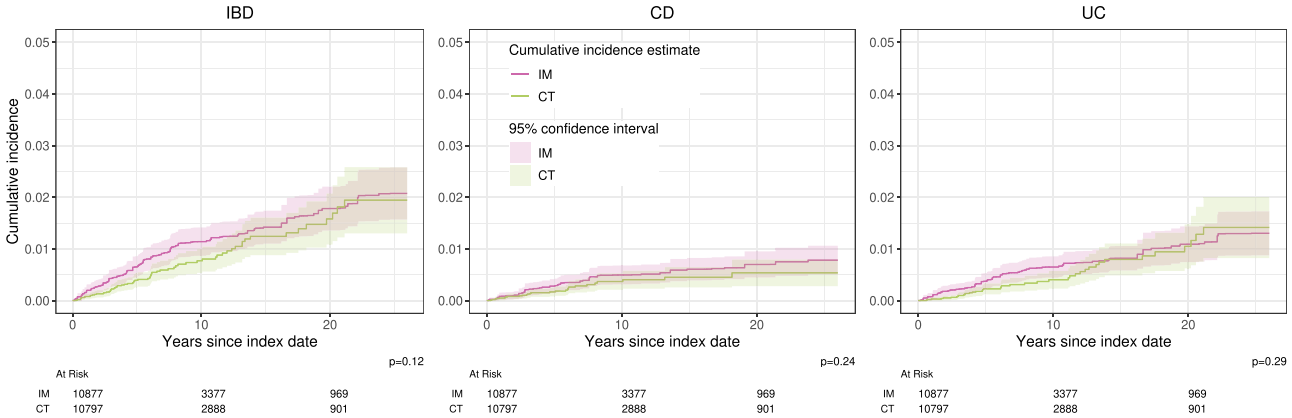

Fig. 6 | Cumulative incidence of IBD, CD, and UC diagnosis in severe IM compared with the negative-control *C. trachomatis* cohort. Those with an ICD-10 coded, secondary care (hospital) diagnosis of IM between 1st January 1977 and 31st December 2021. CT—cohort with *C. trachomatis* infection. Those without an ICD-10 coded hospital diagnosis of IM between 1st January 1996 and 31st December 2021, matched by sex, age, and SES index of the municipality to IM cohort. Index date—date of severe IM or CT diagnosis, which is the start of follow-up. At risk—number of IM and CT cohorts at risk of IBD development at 0, 10-, and 20-years post-follow-up. *P*-values were calculated using single-sided log-rank testing. IM infectious mononucleosis, IBD inflammatory bowel disease, CD Crohn's disease, UC ulcerative colitis, CT *C. trachomatis* infection, IM severe IM cohort.

was associated with IBD development did not show a significant difference in hazards of IBD for those hospitalised with IM and the hospitalisation fraction identified in this current study (0.1) is similar to that found in previous Danish data (0.07), indicating the generalisability and replicability of our findings[14]. Additionally, using these data, we were able to identify IM patients from 1977 to 2021, providing 44 years of follow-up, which is reflected in the size of the cohorts included in our analyses. This longer duration of follow-up is particularly important in the exploration of associations between early life events and later, chronic disease development, as is the case for the current study. Of particular importance in this study is our use of severe IM (i.e., hospitalisation with IM) as an exposure, which has not been previously done and allows for the use of a consistent and biologically plausible marker of innate immune dysfunction following EBV infection in the development of IBD. Another important strength of this work is the use of Danish IBD diagnoses which have previously been validated and found to be close to complete (95%) and highly valid (90–97%) when using two diagnoses for ascertainment, as is the case for the current study[41,42]. Finally, the use of a CT cohort for negative-control matching in our sensitivity analysis further strengthens the validity of the association between IM hospitalisation and the development of IBD identified in the current study.

Whilst we use high-quality data with reliable diagnoses of both IM and IBD, there are some limitations to this work. The observational design has within it an inherent risk of bias due to the inability to control for unknown or unmeasured confounders such as health-seeking behaviours in the cohort. Although this may have impacted our findings of a difference in the risk of later IBD diagnosis in the severe IM group (hospitalised with IM) compared with the matched cohort (not hospitalised with IM), it is unlikely to fully explain the effect observed in our results considering the large effect size and strong significance detected, as is reflected in the results of our negative control matching sensitivity analysis, where IBD diagnosis remains significantly increased in those hospitalised with IM compared to those with a CT diagnosis. Additionally, our null finding for severe IBD in severe IM compared with the matched cohort is not only an indication of similar disease severity following IM hospitalisation but also innately indicative of similar health-seeking behaviours around the time of and following IBD diagnosis in the two groups. Finally, our analysis restricting follow-up to 10 years following IM diagnosis across the calendar year of diagnosis illustrates a consistently significant increased risk for IBD following IM from 1977 to 2021, indicating that our data is not subject to period effects and supports the validity of the findings observed in this work. A final limitation of this work is the lack of data on smoking habits which may be a potential confounder as recent data has indicated an interaction between current smoking status and IM risk[43]. However, if this were systemically biasing findings, we would not expect to see a significant risk of UC diagnosis following severe IM, due to the protective nature of smoking in UC disease development and course[44].

In conclusion, we identify an association between severe IM and subsequent development of IBD, particularly CD. Further exploration of the potential role of EBV infection in IBD pathophysiology is warranted.

Methods

Data sources

We used the Danish National Civil Register (CPR) to identify all individuals living in Denmark between 1st January 1977 and 31st December 2021 and linked these data at an individual level, using a unique personal identification number assigned to all Danish residents, to the Danish National Patient Register (LPR), the Danish National Prescription Register (DPR) and the RLRR. Through the CPR, information such as date of birth, sex, municipality of residence, and date of immigration or emigration is available. SES indices for the municipality of residence were retrieved from the Ministry of Interior and Health (see Supplementary Table 7). All hospital admissions with diagnosis codes have been recorded in the LPR since 1977. The DPR contains individual-level information, including Anatomical Therapeutic Chemical classification code, dose, and pack size for all prescriptions redeemed at outpatient pharmacies in Denmark since 1994. As a CPR number is required for all diagnoses recorded in the LPR and prescriptions redeemed in the DPR, data from these registries are not only representative of the entire resident Danish population but are additionally complete, i.e., these data comprise all national health system hospital diagnoses or publicly reimbursed prescription in Denmark. The RLRR[24] includes results from biochemistry, microbiology, clinical immunology, and haematology hospital laboratories. Tests ordered by general practitioners in Denmark are also available in the register. The RLRR became available nationwide for clinical biochemistry and haematology in 2015 but was limited to selected hospital laboratories for microbiology and clinical immunology results. According to Danish law, studies that are based on registry data alone are not required to obtain permission from the regional ethics committees, which is

confirmed by The Central Denmark Region Committees on Health Research Ethics (legislation: 1–10–72-148-19).

Study population

We identified individuals hospitalised with IM from 1st January 1977 to 31st December 2021 to create the severe IM cohort. We matched these index cases at a ratio of 1:10 to any individual in the general Danish population without a record of hospitalisation with IM in the LPR based on year of birth, sex (male or female), and municipality of residence at IM diagnosis. The index case and their matched counterparts must have been residents of Denmark for at least 10 years prior to IM diagnosis to ensure incident disease cases are identified. The exposure, severe IM, was defined as any hospital diagnosis with IM (ICD-8 codes 07500, 07501, 07508, and 07509 or ICD-10 codes B27.0, B27.1, B27.8, and B27.9) in the LPR. We further classified severe IM exposure by the presence of ICD-10 coded IM complications including hepato-splenomegaly (R16), neuropathy (G63.8, G63.0, or G57), or splenic rupture (D73.5) using the LPR.

The primary outcome for this study was diagnosis with IBD, including CD or UC (ICD-8 codes 56308–09 and ICD-10 code K50 for CD; ICD-8 codes 56319 and 56904, and ICD-10 code K51 for UC). We restricted the outcome of IBD to those with two or more recorded inpatient or out-patient hospital contacts within 2 years, with the last recorded subtype taken as the classification for CD or UC to ensure the validity of IBD diagnoses.

We also assessed whether a history of severe IM is associated with severe IBD within IBD patients using a composite outcome variable of IBD-related hospitalisation, surgery, or administration of immunomodulators, biological therapies, or systemic steroids in the LPR or DPR. Since these measures are based on ICD-10 codes, this analysis is restricted to patients diagnosed with IBD after 1994.

Main analysis

Statistical analysis. We undertook a matched cohort study with a risk-set sampling design and used Cox proportional regression modelling to calculate the HR and 95% CI of IBD diagnosis in those with severe IM. We randomly matched IM cases by sex (male or female), year of birth (within 1 year), and socio-economic status for the municipality of residence (SES index; see Supplementary Table 8) to those without an IM diagnosis at a ratio of 1:10 from the general Danish population, using the CPR, and followed the cohort from the index date to the date of second diagnosis with IBD. Individuals were censored from analysis at the point of death, emigration, or end of the study period (31st December 2021); whichever came first. Analysis was also undertaken for the risk of IBD diagnosis in severe IM compared with the matched cohort for CD and UC separately. We stratified analysis by sex, age at IM diagnosis (0–9, 10–14, 15–19, 20–24, 25–29, and ≥30 years), and calendar year of IM diagnosis (1977–1989, 1990–1999, 2000–2009, and 2010–2021). Cumulative incidence of IBD, CD, and UC in severe IM compared to the matched cohort was also undertaken with a log-rank test used to calculate *P*-values. Finally, the association between severe IM and risk of severe IBD development was assessed also using Cox regression modelling by following those diagnosed with IBD for the first event of a composite variable for severe IBD disease in the severe IM group compared with the matched cohort. All analyses were adjusted for sex, age, and calendar year of diagnosis. To confirm our results are not sensitive to modelling assumptions, we further undertake Cox regression analysis where IM cases and their matched counterparts are clustered and effect sizes are reported with robust standard errors.

Sensitivity analyses

Negative control matching with Chlamydia. To assess the impact of unmeasured and systemic confounding, such as positive health-seeking behaviours, on the association between severe IM and IBD, we undertook negative-control matching with CT, which presents with a similar age distribution to that of IM. The calendar year for inclusion in this analysis started on 1st January 1996, when SES indexes for municipalities in Denmark were first recorded. All individuals with a CT diagnosis from this period who were resident in Denmark at least 9 years prior to their diagnosis and had no pre-existing IBD were included in the CT cohort. Individuals with a prior diagnosis of IM were also excluded from the CT cohort. To balance the covariates in the IM and CT groups, we fitted PS models for the two cohorts. PS was calculated using logistic regression for the probability of covariate (sex, age at diagnosis, year of diagnosis, and SES index)[45]. We subsequently implemented the PS using standardized mortality ratio weights. We used weighted Cox proportional hazards regression models[46] to estimate the risk of IBD diagnosis in IM compared to the CT cohort.

Risk of hospitalisation with positive EBV serology. We used the RLRR to identify tests for EBV capsid antigen IgM using the Nomenclature for Properties and Units code NPU12738. The database includes 41,648 test results for EBV capsid antigen IgM, of which 84% were analysed at three hospital laboratories (53% were analysed at Køge (from 2014), 24% at Vejle (from 2016), and 7% at Esbjerg Hospital (from 2017)). We limited analysis to samples taken at these three hospital laboratories to ensure a complete coverage and positive tests from patients with an existing IBD diagnosis were excluded. A survival analysis was then undertaken to assess the risk of IBD diagnosis in individuals with a positive EBV test and hospitalised with IM within 31 days of the test, compared with those without hospitalisation. Patients were followed from test result date to date of second IBD diagnosis or censoring event (end of follow-up time (31st December 2021), death, or emigration). Age at testing and sex were included as covariates in the model. The analysis was conducted for CD, UC, and IBD separately. All analyses were conducted in R (v4.3.0).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The study was based on data from the Danish National Health registers (https://sundhedsdatastyrelsen.dk). The registered data are protected by the Danish Act on Processing of Personal Data and are accessed through application to and approval from the Danish Data Protection Agency and the Danish Health Data Authority.

Code availability

The code is available on request from the corresponding author.

References

1. Alatab, S. et al. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol. Hepatol.* **5**, 17–30 (2020).
2. Ng, S. C. et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* **390**, 2769–2778 (2017).
3. Larsen, L. et al. Has the incidence of inflammatory bowel disease peaked? Evidence from the population-based NorDIBD cohort 1978–2020. *Am. J. Gastroenterol.* **118**, 501–510 (2023).
4. Chang, J. T. Pathophysiology of inflammatory bowel diseases. *N. Engl. J. Med.* **383**, 2652–2664.
5. de Souza, H. S. & Fiocchi, C. Immunopathogenesis of IBD: current state of the art. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 13–27 (2015). 2015 131.
6. Piovani, D. et al. Environmental risk factors for inflammatory bowel diseases: an umbrella review of meta-analyses. *Gastroenterology* **157**, 647–659.e4 (2019).

7. Pazmandi, J. et al. Early-onset inflammatory bowel disease as a model disease to identify key regulators of immune homeostasis mechanisms. *Immunol. Rev.* **287**, 162–185 (2019).
8. Mann, E. A. & Saeed, S. A. Gastrointestinal infection as a trigger for inflammatory bowel disease. *Curr. Opin. Gastroenterol.* **28**, 24–29 (2012).
9. Hutchinson, E. Risk of IBD increases after *Salmonella* or *Campylobacter gastroenteritis*. *Nat. Rev. Gastroenterol. Hepatol.* **6**, 561–561 (2009). 610 2009.
10. Jess, T. et al. Enteric *Salmonella* or *Campylobacter* infections and the risk of inflammatory bowel disease. *Gut* **60**, 318–324 (2011).
11. Lidar, M. et al. Infectious serologies and autoantibodies in inflammatory bowel disease: insinuations at a true pathogenic role. *Ann. N Y Acad. Sci.* **1173**, 640–648 (2009).
12. James, R. R. Infectious mononucleosis: epidemiology and pathophysiology. *Emerg. Med. N.* **23**, 6–10 (2001).
13. Gares, V. et al. The role of the early social environment on Epstein Barr virus infection: a prospective observational design using the millennium cohort study. *Epidemiol. Infect.* **145**, 3405–3412 (2017).
14. Rostgaard, K. et al. Primary Epstein–Barr virus infection with and without infectious mononucleosis. *PLoS One* **14**, e0226436 (2019).
15. Kuri, A. et al. Epidemiology of Epstein–Barr virus infection and infectious mononucleosis in the United Kingdom. *BMC Public Health* **20**, 1–9 (2020).
16. Ball, R. J. et al. Systematic review and meta-analysis of the sero-epidemiological association between Epstein–Barr virus and rheumatoid arthritis. *Arthritis Res. Ther.* **17**, 274 (2015).
17. Houen G., Trier N. H. Epstein–Barr virus and systemic autoimmune diseases. *Front. Immunol.* **11**, 587380 (2021).
18. Chen, X. et al. Epstein–Barr virus and human herpesvirus 6 infection in patients with systemic lupus erythematosus. *Virol. J.* **20**, 1–11 (2023).
19. Soldan, S. S. & Lieberman, P. M. Epstein–Barr virus and multiple sclerosis. *Nat. Rev. Microbiol.* **21**, 51–64 (2023).
20. Bjornevik, K. et al. Epstein–Barr virus as a leading cause of multiple sclerosis: mechanisms and implications. *Nat. Rev. Neurol.* **19**, :160–:171 (2023). 193 2023.
21. Bjornevik, K. et al. Longitudinal analysis reveals high prevalence of Epstein–Barr virus associated with multiple sclerosis. *Science (80-)* **375**, 296–301 (2022).
22. Nandy, A. et al. 463: EPSTEIN BARR VIRUS (EBV) EXPOSURE PRECEDES CROHN'S DISEASE DEVELOPMENT.*Gastroenterology* **162**, S-105 (2022).
23. Loosen, S. H. et al. Infectious mononucleosis is associated with an increased incidence of Crohn's disease: results from a cohort study of 31 862 outpatients in Germany. *Eur. J. Gastroenterol. Hepatol.* **35**, 255–260 (2023).
24. Arendt, J. F. H. et al. Existing data sources in clinical epidemiology: laboratory information system databases in Denmark. *Clin. Epidemiol.* **12**, 469 (2020).
25. Gehlert, T., Devergne, O. & Niedobitek, G. Epstein–Barr virus (EBV) infection and expression of the interleukin-12 family member EBV-induced gene 3 (EBI3) in chronic inflammatory bowel disease. *J. Med. Virol.* **73**, 432–438 (2004).
26. Yanai, H. et al. Epstein–Barr virus infection of the colon with inflammatory bowel disease. *Am. J. Gastroenterol.* **94**, 1582–1586 (1999).
27. Harley, J. B. et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018). 505 2018.
28. Wong, Y. et al. Estimating the global burden of Epstein–Barr virus-related cancers. *J. Cancer Res. Clin. Oncol.* **148**, 31 (2022).
29. Gequelin, L. C. F. et al. Epstein–Barr virus: general factors, virus-related diseases and measurement of viral load after transplant. *Rev. Bras. Hematol. Hemoter.* **33**, 383 (2011).
30. Loosen, S. H. et al. Infectious mononucleosis is associated with an increased incidence of multiple sclerosis: results from a cohort study of 32,116 outpatients in Germany. *Front. Immunol.* **13**, 937583 (2022).
31. Loosen, S. H. et al. Infectious mononucleosis is associated with an increased incidence of NAFLD. *Eur. J. Clin. Invest.* **53**, e13911 (2023).
32. Roderburg, C. et al. The association between infectious mononucleosis and cancer: a cohort study of 24,190 outpatients in Germany. Cancers **14**, 5837 (2022).
33. Lanz, T. V. et al. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature* **603**, 321–327 (2022). 6037900 2022.
34. Garabatos, N. & Santamaria, P. Gut microbial antigenic mimicry in autoimmunity. *Front Immunol.* **13**, 873607 (2022).
35. English, J., Patrick, S. & Stewart, L. D. The potential role of molecular mimicry by the anaerobic microbiota in the aetiology of autoimmune disease. *Anaerobe* **80**, 102721 (2023).
36. Verma, A. et al. Elucidating potential molecular signatures through host-microbe interactions for reactive arthritis and inflammatory bowel disease using combinatorial approach. *Sci. Rep.* **10**, 1–16 (2020). 101 2020.
37. Pender, M. P. et al. Defective T-cell control of Epstein–Barr virus infection in multiplesclerosis. *Clin. Transl. Immunol.* **6**, e126 (2017).
38. Atara Biotherapeutics. Phase 1/2 Study to Evaluate the Safety and Efficacy of ATA188 in Subjects With Progressive Multiple Sclerosis. ClinicalTrials.gov identifier NCT03283826. Last Updated 23 Feb 2024. Accessed 17 July 2024. clinicaltrials.gov.
39. Callan, M. F. C. The immune response to Epstein–Barr virus. *Microbes Infect.* **6**, 937–945 (2004).
40. Horton, R. et al. Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**, 889–899 (2004).
41. Fonager, K. et al. Assessment of the diagnoses of Crohn's disease and ulcerative colitis in a Danish hospital information system. *Scand. J. Gastroenterol.* **31**, 154–159 (1996).
42. Jacobsen, H. A., Jess, T. & Larsen, L. Validity of inflammatory bowel disease diagnoses in the Danish National Patient Registry: a population-based study from the North Denmark region. *Clin. Epidemiol.* **14**, 1099–1109 (2022).
43. Hedström, A. K. et al. Smoking and Epstein–Barr virus infection in multiple sclerosis development. *Sci. Rep.* **10**, 10960 (2020).
44. Ho, S.-M. et al. Challenges in IBD research: environmental triggers. *Inflamm. Bowel Dis.* **25**, S13–S23 (2019).
45. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
46. Desai, R. J. & Franklin, J. M. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. *BMJ* **367**, l5657 (2019).

Acknowledgements

With thanks to Heidi Søgaard Christensen, Manasi Agrawal, and Rasmus Froberg Brøndum for their contribution to the earlier work undertaken exploring the EBV testing using the Danish nationwide RLRR and Jesus Vicente Torresano Lominchar and Sabrina Hjort Andersen for contribution in the conception and creation of manuscript tables and figures. T.J. received funding from the Danish National Research Foundation (DNRF148) and the Novo Nordisk Foundation (NNF21OC0068631). R.E. received funding from the Lundbeck Foundation (R403-2022-1531).

Author contributions

R.E., S.H., and A.E. developed the study concept and design. A.E., S.H., and M.V.V. undertook all analyses. All authors were responsible for data interpretation. R.E. and S.H. undertook the writing of the first draft of the manuscript. A.E., W.M., M.V.V., and T.J. were responsible for the critical revision of important intellectual content. T.J. and R.E. obtained funding for the work undertaken.

Competing interests
The authors have no financial, professional, or personal conflicts to declare.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52195-8.

Correspondence and requests for materials should be addressed to Rahma Elmahdi.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.