

Hospital-diagnosed infections before age 20 and risk of a subsequent multiple sclerosis diagnosis

Yin Xu,¹ Kelsi A. Smith,^{2,3} Ayako Hiyoshi,¹ Fredrik Piehl,³ Tomas Olsson³ and Scott Montgomery^{1,2,4}

Abstract

The involvement of specific viral and bacterial infections as risk factors for multiple sclerosis has been studied extensively. However, whether this extends to infections in a broader sense is less clear and little is known about whether risk of a multiple sclerosis diagnosis is associated with other types and sites of infections, such as of the CNS. This study aims to assess if hospital-diagnosed infections by type and site before age 20 years are associated with risk of a subsequent multiple sclerosis diagnosis and whether this association is explained entirely by infectious mononucleosis, pneumonia, and CNS infections. Individuals born in Sweden between 1970-1994 were identified using the Swedish Total Population Register (n = 2,422,969). Multiple sclerosis diagnoses from age 20 years and hospital-diagnosed infections before age 20 years were identified using the Swedish National Patient Register. Risk of a multiple sclerosis diagnosis associated with various infections in adolescence (11-19 years) and earlier childhood (birth-10 years) was estimated using Cox regression, with adjustment for sex, parental socioeconomic position, and infection type. None of the infections by age 10 years were associated with risk of a multiple sclerosis diagnosis. Any infection in adolescence increased the risk of a multiple sclerosis diagnosis (hazard ratio 1.33, 95% confidence interval 1.21-1.46) and remained statistically significant after exclusion of infectious mononucleosis, pneumonia, and CNS infection (hazard ratio 1.17, 95% confidence interval 1.06-1.30). CNS infection in adolescence (excluding encephalomyelitis to avoid including acute disseminated encephalitis) increased the risk of a multiple sclerosis diagnosis (hazard ratio 1.85, 95% confidence interval 1.11-3.07). The increased risk of a multiple sclerosis diagnosis associated with viral infection in adolescence was largely explained by infectious mononucleosis. Bacterial infections in adolescence increased risk of a multiple sclerosis diagnosis, but the magnitude of risk reduced after excluding infectious mononucleosis, pneumonia and CNS infection (hazard ratio 1.31, 95% confidence interval 1.13-1.51). Respiratory infection in adolescence also increased risk of a multiple sclerosis diagnosis (hazard ratio 1.51, 95% confidence interval 1.30-1.75), but was not statistically significant after excluding infectious

mononucleosis and pneumonia. These findings suggest that a variety of serious infections in adolescence, including novel evidence for CNS infections, are risk factors for a subsequent multiple sclerosis diagnosis, further demonstrating adolescence is a critical period of susceptibility to environmental exposures that raise the risk of a multiple sclerosis diagnosis. Importantly, this increased risk cannot be entirely explained by infectious mononucleosis, pneumonia, or CNS infections.

Author affiliations:

1 Clinical Epidemiology and Biostatistics, School of Medical Sciences, Örebro University, Örebro, Sweden

2 Clinical Epidemiology Division, Karolinska Institutet, Stockholm, Sweden

3 Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

4 Department of Epidemiology and Public Health, University College London, London, UK

Correspondence to: Scott Montgomery

Clinical Epidemiology and Biostatistics, School of Medical Sciences, Campus USÖ, Örebro University, SE-703 62 Örebro, Sweden

E-mail: scott.montgomery@oru.se

Running head: Infection and multiple sclerosis

Keywords: infection; multiple sclerosis; adolescence; CNS

Abbreviations: BMI = body mass index; EBV = Epstein-Barr virus; HHV-6 = Human Herpesvirus 6; HLA = human leukocyte antigen; ICD = International Classification of Diseases; IgG = immunoglobulin G; LISA = Longitudinal Integrated Database for Health Insurance and Labour Market Studies

Introduction

Multiple sclerosis has both genetic and environmental risk factors, including human leukocyte antigen (HLA) type, infections, low vitamin D levels, higher body mass index (BMI), and smoking, sometimes with potent interactions with HLA risk genes.¹⁻⁴ The involvement of infections in increasing the risk of multiple sclerosis has been frequently studied, such as Epstein-Barr virus (EBV), Human Herpesvirus 6 (HHV-6), and *Chlamydia pneumoniae*. Two meta-analyses found that infectious mononucleosis in adolescents and young adults, a clinical manifestation of EBV infection, more than doubled the multiple sclerosis risk (relative risk ratios 2.30 and 2.17, respectively).^{5,6} Meta-analyses of case-control studies also found significantly higher prevalence of anti-EBV antibodies, including anti-viral capsid antigen (anti-VCA) immunoglobulin G (IgG) and anti-EBV nuclear antigen (anti-EBNA) IgG, in multiple sclerosis patients compared to controls without multiple sclerosis.^{7,8}

There is also extensive evidence supporting that HHV-6 is associated with increased multiple sclerosis risk.^{9,10} Most prior studies are cross-sectional serology studies from which it is difficult to determine when infection occurred and whether HHV-6 is the cause or the consequence of multiple sclerosis.⁹ However, a recent study that collected serum before multiple sclerosis onset suggested that HHV-6A in contrast with HHV-6B, determined serologically, was associated with an increased risk for future multiple sclerosis.¹¹ For multiple sclerosis risk associated with *Chlamydia pneumoniae*, a bacterial infection causing respiratory tract infections (including pneumonia), moderate evidence of increased multiple sclerosis risk has been found. A systematic review of six published papers before 2002 provided mixed evidence,¹² but a more recent meta-analysis of three case-control studies found increased anti-*Chlamydia pneumoniae* immunoglobulins in multiple sclerosis patients compared to controls without multiple sclerosis (odds ratio 1.35).¹³ The serology studies could not determine when the infection occurred and if it pre-dated multiple sclerosis onset. However, we recently conducted a study with prospectively recorded information, focusing on pneumonia using Swedish register data, which provided evidence of an association for pneumonia between ages 11 and 15 years with a raised risk of multiple sclerosis after age 20 years.¹⁴ Collectively, the evidence to date suggests that increased multiple sclerosis risk is associated with specific viruses, in particular EBV and perhaps HHV6A, but also other infections involving the airways.¹⁵ However, whether this extends to infections in a broader sense is less clear and little is known about whether there is an increased multiple sclerosis risk associated with other types and sites of infections, such as infections of the CNS.

Several theories have been proposed to explain the association between infections and increased multiple sclerosis risk. The molecular mimicry hypothesis suggests that infectious agents with homologous sequences or structures to a host's myelin antigens could trigger cross-activation of autoreactive T-cells to attack host tissue.¹⁵ Another suggested mechanism is that macrophages and natural killer cells activated by infectious agents elsewhere in the body, such as the lungs, can result in pro-inflammatory cytokine production and non-specific activation of pre-primed T-cells. This allows them to cross the blood-brain barrier, causing inflammation in the CNS, and inducing multiple sclerosis pathogenesis by triggering autoimmune responses against myelin.¹⁶ For example, autoreactive T-cells triggered by lung irritation are programmed to gain a migratory profile through bronchus-associated lymphoid tissue involvement, resulting in inflammation in the CNS.¹⁶

Adolescence and early adulthood may be important periods of susceptibility for the influence of infection on multiple sclerosis pathogenesis.^{1,6} The increased multiple sclerosis risk associated with EBV infection, in the form of infectious mononucleosis, in adolescence is approximately twice as high as infection in earlier childhood.^{1,6} Several other exposures are associated with increased multiple sclerosis risk when they occur in adolescence rather than earlier childhood: concussion,¹⁷ lower serum vitamin D level,¹⁸ higher BMI,¹⁹ and pneumonia.¹⁴ Migration studies also provide further evidence that adolescence may represent a window of susceptibility to environmental exposures. Individuals who migrate before age 15 years acquire the multiple sclerosis risk of their new country, whereas individuals who migrate at later ages retain the multiple sclerosis risk of their country of origin.³ Even though the period between ages 11 to 19 years potentially represents a time when some environmental exposures may initiate multiple sclerosis pathogenesis, overt clinical onset is often observed some years later, as multiple sclerosis may have a long prodrome of five to 10 years.²⁰ It should be noted that infections are often the consequences of multiple sclerosis.²¹ Thus, a potentially important additional advantage of studying exposures between ages 11 and 19 years is that they are less likely to be consequences of multiple sclerosis, which typically has its onset at later ages.

This study uses a large general population-based birth cohort in Sweden to assess risk of a multiple sclerosis diagnosis from age 20 years associated with hospital-diagnosed infections in adolescence (ages 11 to 19 years) and earlier childhood (between birth and age 10 years). We hypothesise that adolescence represents a period when environmental exposures are more likely to be causally associated with increased risk of a subsequent multiple sclerosis diagnosis, while exposures in earlier childhood are less likely to contribute to risk of a multiple sclerosis diagnosis. We investigate the differences in the risk of a multiple sclerosis diagnosis by

infection type and site, including CNS, gastrointestinal, genitourinary, respiratory, skin, bacterial, viral, and the other infections. Where specific infections are associated with risk of being diagnosed with multiple sclerosis (from previous studies, or those we identify here), we then excluded them to demonstrate the potential role of other types of infection.

Materials and Methods

Study population

The cohort comprises all individuals who were born in Sweden between 1970 and 1994 and reached age 20 years by the end of study (31st December 2014), identified by the Swedish Total Population Register. Dates for almost all births, deaths, immigration, and emigration are included in the Swedish Total Population Register since notification of these events is mandatory.²² The unique individual Swedish personal identification number issued to all residents at birth or immigration was used for data linkage with the Swedish National Patient Register (to identify both inpatient and outpatient diagnoses), the Population and Housing Census data (to identify parental occupation before 1990), the Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA, to identify parental occupation since 1990), and the Multi-generation Register (to identify parents of cohort members, so that in turn we could link with the Population and Housing Census data or LISA to identify parental occupation and thus produce a parental socioeconomic index). The Patient Register was established in 1964 and has complete national coverage since 1987, with hospital-based outpatient diagnoses being collected since 2001.²³ Primary care/general practitioner visits are not included in the Patient Register.²³ A validation study examining the quality of the Swedish Patient Register compared the multiple sclerosis patients identified from the Patient Register with those identified using information linked from six national registers in Sweden (including the Swedish Multiple Sclerosis Register, the Swedish Prescribed Drug Register, and the Micro-Data for Analysis of the Social Insurance System) in two sequential register-based case-definition algorithms (as the gold standard), and reported an acceptable level of accuracy for multiple sclerosis diagnoses (with a positive predictive value of 92.50%).²⁴ Every five years between 1960 and 1990, the Population and Housing Census collects data on all members of the Swedish population aged 16 years or older, including education, income, occupation, and employment.²⁵ LISA has collected similar information as the census annually since 1990.²⁵ All Swedish residents are automatically included in the registers used for this study and are thus eligible for selection. Ethical approval for this study was granted by the Swedish Ethical Review

Authority (reference number: 2019-04143). A total of 2,650,337 individuals were identified and 2,422,969 individuals were included in the analysis (1,244,970 men and 1,177,999 women; see Fig.1).

Infections

All hospital-diagnosed infections before age 20 years were identified from the Patient Register (see Supplemental Table 1 for the list of International Classification of Diseases (ICD) codes for infections included in the analyses). Infections (ever or never) were defined as having occurred if the ICD codes for infections listed in Supplemental Table 1 were found in either primary or secondary diagnoses. The number of diagnoses was not considered. Instead, infection was coded as ever or never (as some infections were rare or do not recur). We further categorized any infections before age 20 years according to type and site, including infectious mononucleosis, pneumonia, CNS, gastrointestinal, genitourinary, respiratory, skin, bacterial, viral, and the other infections (see Supplemental Table 1 for the list of ICD codes for infection by type and site). A similar procedure was followed for specific infections. To examine the age-defined periods of potential susceptibility for the association of infections with risk of a multiple sclerosis diagnosis, any infection diagnoses before age 20 years and infection by site and type were categorized as occurring in earlier childhood from birth to age 10 years, or in adolescence between ages 11 and 19 years, based on the age at infection.

Multiple sclerosis and potential confounders

A diagnosis of multiple sclerosis from age 20 years (from the 20th birthday and subsequently) was identified from the Patient Register using ICD codes (ICD-8 or -9 code 340, ICD-10 code G35). To ensure diagnostic accuracy, two multiple sclerosis diagnoses recorded at a minimum of six months apart were required, as initial progression of the disease is required to confirm diagnosis and a gap of at least six months make it more likely that relevant differential diagnoses can be made.²⁶ A study of health claims data in Canada has also found a high level of accuracy when using at least two multiple sclerosis diagnoses to identify multiple sclerosis cases (with sensitivity and positive predictive value above 0.90).²⁷ It is also unlikely that the algorithm used to identify multiple sclerosis diagnoses in our study would lead to double counting of a multiple sclerosis diagnosis during a single hospital admission since it would be highly unusual for patients to be transferred between hospitals or hospital departments over six months apart for the same initial admission (hospital admissions for multiple sclerosis tend to be for less than 30

days).²⁸ The first diagnosis was used to define date of first multiple sclerosis diagnosis. Individuals with only one multiple sclerosis diagnosis or with separate multiple sclerosis diagnoses recorded within six months were coded as “no multiple sclerosis” (n = 510), but coded as having “multiple sclerosis” in a separate sensitivity analysis (sensitivity analysis-6). The highest level of parental occupation nearest in time to the birth of the cohort member from census data and the LISA register was used to derive a three-category approximation of the European Socioeconomic Classification.²⁹

Statistical analysis

Cox proportional hazard regression was used to examine the association between having had any infections in adolescence or in earlier childhood with risk of a subsequent multiple sclerosis diagnosis. Follow-up was from age 20 years until the first diagnosis of multiple sclerosis, death, first emigration, or study end (31st December 2014), whichever occurred first. Cohort members who emigrated were censored on the first emigration date registered (individuals leaving Sweden) and could not re-enter the cohort to ensure there were no gaps in follow-up. The adjusted model included the measures of any infections between birth and age 10 years and between age 11 and 19 years, with adjustment for sex and parental socioeconomic position. Potential confounders were selected based on prior research and their availability in the registers.

In order to test whether risk of a multiple sclerosis diagnosis was associated with infections by type and site in adolescence or earlier childhood, two multivariable regression models were used. One single multivariable regression model with all types and sites of infection included together would lead to multicollinearity and over-adjustment given the high correlations (see Supplemental Table 2 for tetrachoric correlations) and some overlap among infections by type and site (for example, the same infection may be both respiratory and viral). Accordingly, one model included measures of CNS, genitourinary, respiratory, skin, gastrointestinal, and the category “other infections” in adolescence and earlier childhood, while another model included measures of bacterial, viral, gastrointestinal, and the category “other infections”. Both models were adjusted for sex and parental socioeconomic position.

In order to test whether the association between any hospital-diagnosed infections in adolescence or earlier childhood and risk of a subsequent multiple sclerosis diagnosis was not limited to infectious mononucleosis, pneumonia, and CNS infection, a three-step multivariable regression was used with each of these categories was excluded from the definition of infection. First, any infection excluding infectious mononucleosis was compared to no infection. In the

second step, any infection excluding infectious mononucleosis and pneumonia was compared to no infection. Finally, any infection excluding infectious mononucleosis, pneumonia, and CNS infection was compared to no infection. All models included measures of any infections in adolescence and earlier childhood, with adjustment for sex and parental socioeconomic position. The same three-step multivariable regression analysis was applied for bacterial and viral infection to test whether the association with risk of a subsequent multiple sclerosis diagnosis was independent of infectious mononucleosis, pneumonia, and CNS infection. In order to test whether the association between respiratory infection and risk of a subsequent multiple sclerosis diagnosis was independent of infectious mononucleosis and pneumonia, we also excluded infectious mononucleosis and pneumonia with adjustment for sex, parental socioeconomic position, and CNS infection.

A total of seven sensitivity analyses were performed. In the first sensitivity analysis, the analysis was stratified by sex. In the second sensitivity analysis, a first multiple sclerosis diagnosis from age 25 years was used to further exclude the possibility of reverse causation given the relatively long prodromal period in multiple sclerosis. In the third sensitivity analysis, we investigated the risk of a multiple sclerosis diagnosis after excluding encephalomyelitis from any infections and CNS infection as this category includes acute disseminated encephalitis, which can be difficult to differentiate from multiple sclerosis onset. We were unable to identify specifically acute disseminated encephalitis (which requires ICD code precision of four digits, which was not available in our data). Thus, encephalomyelitis diagnoses (ICD-8 or -9 code 323, ICD-10 code G04) were excluded as they include acute disseminated encephalitis. In the fourth sensitivity analysis, we excluded individuals who were born after 1987 to examine the influence of the incomplete national coverage of the Patient Register until 1987 on the association between any infections and risk of a multiple sclerosis diagnosis. In the fifth sensitivity analysis, we investigated whether the association between any infections and risk of a multiple sclerosis diagnosis was confounded by BMI in a subset sample with BMI information available (31.17% of the population included here). BMI was only available for those included in the Swedish Military Conscription Register and was largely limited to males. In the sixth sensitivity analysis, patients with a single multiple sclerosis diagnosis ($n = 510$) were additionally coded as having “multiple sclerosis”. In the final sensitivity analysis, the first diagnosis of any demyelinating disease of the CNS, in addition to multiple sclerosis, including acute disseminated encephalomyelitis, other demyelinating disease of CNS, other acute disseminated demyelination, and optic neuritis (ICD-8 or -9 code: 323, 341; ICD-10 code: G04, H46, G36, G37),³⁰ were used to get closer to the possible onset date of multiple sclerosis. In this analysis,

those with the first demyelinating diseases of the CNS diagnosed before age 20 years were excluded from the analysis.

No evidence of violation of the proportional hazards assumption was found using Schoenfeld residuals. All analyses were conducted using Stata statistical software, version 16 (StataCorp, College Station, TX).

Data availability

All registers used here are available upon completion and approval of an application for data release from the Swedish National Board of Health and Welfare, Statistics Sweden and the Swedish Military Conscription Register.

Results

Sample characteristics

Some 4,022 individuals (0.17%) were diagnosed with multiple sclerosis from age 20 years, at a mean age of 30.11 years (median age 29.70). The median follow-up time from age 20 years until the first diagnosis of multiple sclerosis was 9.70 years (interquartile range 8.04 years, range 0.02-24.88 years). Table 1 shows characteristics of the birth cohort. A total number of 462,157 (19.07%) and 338,352 (13.96%) individuals had hospital-diagnosed infections between birth and age 10 years and between age 11 and 19 years, respectively. Compared with males, females were more likely to have a multiple sclerosis diagnosis (hazard ratio [HR] 2.52, 95% confidence interval [CI] 2.35-2.70). Parental socioeconomic position was not statistically significantly associated with risk of being diagnosed with multiple sclerosis.

Any infection and risk of a multiple sclerosis diagnosis

Compared with no infection from birth to age 10 years, any infection from birth to age 10 years was not statistically significantly associated with increased risk of a subsequent multiple sclerosis diagnosis in both unadjusted (HR 0.96, 95% CI 0.88-1.04) and adjusted models (HR 0.99, 95%CI 0.90-1.08). Whereas, individuals with any infection in adolescence (between ages 11 and 19 years) were at greater risk of being diagnosed with multiple sclerosis in both unadjusted (HR 1.39, 95%CI 1.26-1.52) and adjusted models (HR 1.33, 95%CI 1.21-1.46), compared with individuals with no infection in adolescence. The increased risk of a subsequent multiple sclerosis diagnosis associated with any infection in adolescence remained statistically

significant after the exclusion of infectious mononucleosis, pneumonia, and CNS infection, HR 1.17, 95%CI 1.06-1.30 (see Fig. 2).

Risk of a multiple sclerosis diagnosis by type and site of infection

Table 2 shows the infections by type and site. For the risk of a multiple sclerosis diagnosis associated with infection by type and site, infectious mononucleosis in adolescence between ages 11 and 19 years increased the risk of being diagnosed with multiple sclerosis with adjustment for pneumonia, sex, and parental socioeconomic position (HR 2.64, 95%CI 2.06-3.38). CNS, genitourinary, respiratory, bacterial, and viral infection in adolescence also increased the risk of being diagnosed with multiple sclerosis in the adjusted model, with HRs ranging from 1.28 to 2.80 (see Table 2). Further adjustment for infectious mononucleosis and pneumonia did not eliminate the increased risk of a multiple sclerosis diagnosis associated with CNS infection in adolescence, HR 2.80, 95%CI 1.90-4.12.

The increased risk of being diagnosed with multiple sclerosis associated with respiratory infection in adolescence was attenuated and became statistically non-significant when respiratory infection without infectious mononucleosis and pneumonia diagnoses in adolescence was further compared with no infection with adjustment for sex, parental socioeconomic position, and CNS infection in adolescence, HR 1.22, 95%CI 0.99-1.49. Similarly, compared with no viral infection, viral infection excluding infectious mononucleosis in adolescence did not statistically significantly increase the risk of being diagnosed with a subsequent multiple sclerosis in both unadjusted, HR 1.18, 95%CI 0.96-1.44, and adjusted models, HR 1.14, 95%CI 0.93-1.40 (see Fig. 3). The increased risk of being diagnosed with multiple sclerosis associated with bacterial infection in adolescence remained statistically significant when individuals with bacterial infection but without CNS infection, infectious mononucleosis, and pneumonia diagnoses was compared with those without bacterial infection in both unadjusted, HR 1.46, 95%CI 1.26-1.69, and adjusted models, HR 1.31, 95%CI 1.13-1.51 (see Fig. 4).

Sensitivity analyses

Sensitivity analysis-1 found similar HR associated with any infections or infection by site and type when stratified by sex (data not shown). Restricting multiple sclerosis diagnoses to the first from age 25 years in sensitivity analysis-2 showed that genitourinary infection in adolescence may be a consequence of prodromal multiple sclerosis since the association was

attenuated and not statistically significant in the adjusted model (HR 1.30, 95%CI 0.84-2.02). However, the increased risk of being diagnosed with multiple sclerosis associated with CNS, respiratory, bacterial, and viral infection in adolescence remained statistically significant, although decreased in magnitude, with HRs ranging from 1.24 to 2.16 (see Table 3). Sensitivity analysis-3 found that individuals with CNS infection excluding encephalomyelitis in adolescence were still at greater risk of being diagnosed with multiple sclerosis in the adjusted model, HR 1.85, 95%CI 1.11-3.07. Sensitivity analysis-3 also revealed that there were no substantial differences in HR between analyses including and excluding encephalomyelitis from the definition of any infections (see Supplemental Fig. 1). Sensitivity analysis-4 found no substantial differences in HR between analyses excluding and including individuals who were born after 1987 (data not shown). Sensitivity analysis-5 of a subset population with measured BMI available found that individuals with any infections in adolescence remained at statistically significantly greater risk of being diagnosed with multiple sclerosis after adjusting for BMI, sex, any infections by age 10 years, and parental socioeconomic position (HR 1.40, 95%CI 1.14-1.72). Sensitivity analysis-6 found no substantial differences in HRs between the analysis defining patients with a single multiple sclerosis diagnosis as “no multiple sclerosis” and the analysis defining patients with a single multiple sclerosis diagnosis as having “multiple sclerosis” (data not shown). Sensitivity analysis-7 found no substantial differences in HR for risk of a multiple sclerosis diagnosis associated with any infection in adolescence between the analysis defining onset using the first diagnosis of multiple sclerosis and the analysis defining onset using the first diagnosis of any demyelinating disease if prior to the first multiple sclerosis diagnosis (HR 1.39, 95%CI 1.30-1.50 and HR 1.34, 95%CI 1.23-1.47 for analysis defining onset using the first demyelinating diseases of the CNS diagnosed from age 20 and 25 years, respectively).

Discussion

Any hospital-treated infections in adolescence rather than in earlier childhood increased the risk of a multiple sclerosis diagnosis from age 20 years, though the effect size was small (HR 1.33) as not all types of infection are associated with a subsequent multiple sclerosis diagnosis (possibly other unmeasured characteristics of infections, such as severity/duration or repeated exposure, may also be important). This association remained statistically significant among individuals with infection but without infectious mononucleosis, pneumonia, and CNS

infection (HR 1.17). Further evidence of an association between CNS infection and a subsequent raised risk of a multiple sclerosis diagnosis was observed (HR 1.85) and this may be a conservative estimate as we had to exclude encephalomyelitis diagnoses. Individuals with bacterial infections but without infectious mononucleosis, pneumonia, and CNS infection remained at greater risk of a multiple sclerosis diagnosis. However, unlike CNS, respiratory, viral, and bacterial infections, the association with genitourinary infection in adolescence was attenuated notably when those with multiple sclerosis-onset by age 25 years were excluded. This might suggest that early-onset multiple sclerosis, or prodromal disease activity increases genitourinary infection risk, although such infections tend to occur later in the multiple sclerosis disease course. The threshold for exposure to infections before age 20 years, might potentially underestimate genitourinary and related infections associated with first sexual activity, although the median age of first sexual intercourse in Sweden has been estimated at 17 years.³¹

Recently, we published research on a similar topic, but exclusively examined if pneumonia before age 20 years represented an increased risk for a subsequent multiple sclerosis diagnosis using a Swedish register-based case-control study (n=55,588).¹⁴ The earlier study considered IM as a potential confounding factor, while using genitourinary infection as a control condition to identify potential reverse causation. The current cohort study has greater statistical power to investigate rarer types of infection due to a substantially larger number of individuals and had the aim of assessing if a more comprehensive *variety* of types of infection (by type, site and whether viral or bacterial) are associated with a raised risk of a subsequent MS diagnosis independently of each other. The previous study found that pneumonia in adolescence was associated with an increased risk of a multiple sclerosis diagnosis,¹⁴ so we used this information to inform the design of the current study, with an analysis excluding people who had pneumonia (along with other types of infection previously associated with raised multiple sclerosis risk) to provide further evidence that associations with multiple sclerosis are not explained through confounding by the excluded infections.

Associations of infections with multiple sclerosis

Infections are often the consequence of multiple sclerosis,²¹ but to tackle this problem we considered only infections between before age 20 years and performed a sensitivity analysis limiting age of first multiple sclerosis diagnosis to over age 25 years, so that there is a delay of at least five years between exposure and multiple sclerosis diagnosis. This approach tackles the possibility of reverse causation such that health-related behavioural changes, mental health

burden, and physical manifestations of preclinical multiple sclerosis during the prodromal period.²⁰ In addition, although we have not identified infections at the molecular level, we have been able to effectively capture all prospective infections of sufficient severity to result in a hospital diagnosis up to age 19 years and preceding a multiple sclerosis diagnosis.

The association we found between infections in adolescence and increased risk of a subsequent multiple sclerosis diagnosis was not only explained by infections previously linked with multiple sclerosis risk: infectious mononucleosis and pneumonia, nor due to CNS infection, which although not having been previously investigated as a risk factor for multiple sclerosis (to our knowledge), could plausibly affect later multiple sclerosis risk as indicated by this study. Thus, other types and sites of infection in adolescence were also risk factors increasing the likelihood of a subsequent multiple sclerosis diagnosis. Bacterial infections other than pneumonia could be of interest as individuals with a bacterial infection but without infectious mononucleosis, pneumonia, or CNS infection were at greater risk of being diagnosed with multiple sclerosis, though the magnitude of the HR reduced after excluding these diagnoses. Studies of bacterial DNA in cerebrospinal fluid of multiple sclerosis patients have had mixed results, with some studies unable to detect bacteria such as Mycobacteria, spirochetes, Campylobacter, Bartonella, and Streptococcus,³² but others have been able to identify bacteria causing pneumonia such as *Chlamydia pneumoniae*.¹³ Possible explanations for the variability of results could be differences in screening systems targeting antigens, small sample sizes, or cross-sectional study designs.³³ Viral infections excluding infectious mononucleosis were not found to be a risk factor for a multiple sclerosis diagnosis, as the observed increased risk was reduced and statistically nonsignificant after exclusion of patients with infectious mononucleosis. Indeed, studies of associations between multiple sclerosis and other virus infections, including Human Endogenous Retroviruses (HERV), measles virus, and rubella virus, have produced mixed evidence,^{9,33} possibly due to the clearance of viral nucleic acids by the time of multiple sclerosis diagnosis, leading to difficulty in establishing a virus as a cause of multiple sclerosis.³⁴ Our combined results suggest that multiple infectious pathogens rather than a single pathogen contribute to risk of a multiple sclerosis diagnosis.³⁴ However, the current study did not consider whether multiple infectious pathogens act independently or interact with each other (an additive or multiplicative effect).

Our results also suggest that adolescence rather than earlier childhood seems to be a period of susceptibility to infections exposures relevant to multiple sclerosis diagnosis. Such an age-specific pattern is consistent with prior studies on other environmental risk factors, including concussion,¹⁷ vitamin D,¹⁸ and higher BMI.¹⁹ Other periods may be relevant for some exposures,

as high level of sun exposures in childhood was inversely associated with multiple sclerosis risk, whereas smoking in adulthood is associated with increased multiple sclerosis risk.³

Possible mechanisms involved in multiple sclerosis susceptibility

Infections in adolescence as factors increasing the risk of a multiple sclerosis diagnosis could be partially explained by their possible involvement in damage to the CNS triggering autoimmune processes pertinent for multiple sclerosis pathogenesis. The magnitude of the association between CNS infection in adolescence and risk of a subsequent multiple sclerosis diagnosis was greater than respiratory, bacterial, and viral infections, but similar to infectious mononucleosis. Reverse causation is unlikely as CNS infection in adolescence was consistently associated with increased risk of a multiple sclerosis diagnosis when we restricted to multiple sclerosis diagnoses from age 25 years. The magnitude of the estimate is likely to be conservative as we had to exclude all diagnoses associated with ICD-8 or -9 code 323 and ICD-10 code G04, to ensure exclusion of acute disseminated encephalitis which can be difficult to differentiate from multiple sclerosis onset. Inflammation in the CNS caused by autoreactive T-cells due to possible molecular mimicry between the infectious pathogens and an individual's myelin antigens may be another possible mechanism. For example, three EBV proteins, Epstein-Barr nuclear antigen-1 (EBNA-1), the tegument protein BRRF2, and the small capsid protein (BFRF3), share epitopes with myelin basic proteins (e.g., septin-9).^{35,36} Antibodies to the EBNA-1, BRRF2, and BFRF3 cross-reacted with myelin antigens which causes recurrent autoimmune damage in the CNS.^{35,36} Antibodies to EBNA-1 can also cross-react with the chloride-channel protein Anoctamin 2 (ANO2) which increases multiple sclerosis risk.³⁷ Indeed, the presence of anti-EBNA-1 IgG and anti-VCA IgG antibodies has been shown to be significantly higher among multiple sclerosis patients compared with controls,⁷ and increased CD4+ and CD8+ T cell responses to EBNA-1 in multiple sclerosis patients compared with controls have been identified.³⁵ It is also possible that autoreactive T-cells triggered by infectious pathogens could be programmed to gain a migratory profile, cross the blood-brain barrier, and cause inflammation in the CNS.¹⁶ For example, pneumonia caused by *Streptococcus pneumoniae* can induce interleukin-17 (IL-17) production and T-helper type 17 (Th17) responses,³⁸ which can infiltrate the CNS,³⁹ consistent with the association of pneumonia in adolescence with subsequent multiple sclerosis risk that we reported previously.¹⁴

Strengths and limitations

Strengths of the present study include the use of a large general population-based birth cohort to identify somewhat infrequent events with high statistical power. Unlike prior cross-sectional serology studies, infection in adolescence and earlier childhood were measured prospectively. We also included a sensitivity analysis using multiple sclerosis diagnoses from age 25 years, which further reduced the likelihood of reverse causation between infection in adolescence and risk of a subsequent multiple sclerosis diagnosis. The range of infections included in the current study was not limited to previously identified risk factors for multiple sclerosis (such as infectious mononucleosis and pneumonia),¹⁴ which help us to identify further potential infectious risk factors for a multiple sclerosis diagnosis. Finally, we examined infection at two theoretically relevant developmental stages (adolescence and earlier childhood) to assess potential periods of susceptibility for risk of a multiple sclerosis diagnosis associated with infection.

This study also has some potential limitations. The number of infections in adolescence and earlier childhood may be underestimated as infections diagnosed and treated in primary care/general practice could not be included and hospital-based outpatient register data were not available until 2001, but we have identified the more serious infections diagnosed and treated in hospital. Further, the Swedish National Patient Register did not achieve complete national coverage for inpatient registrations until 1987. Thus, some infection diagnoses may be missing, especially for these who were born before 1987, but our sensitivity analysis found no substantial differences in HRs between analyses excluding and including individuals who were born after 1987. The ICD codes used in our data had three-digit precision, which made it impossible to specifically identify certain diagnoses (we had to exclude all encephalomyelitis for the sensitivity analysis, to ensure we excluded acute disseminated encephalitis). As serious infections during adolescence are not common, and because of the age structure of the cohort, we were unable to examine the risk of a multiple sclerosis diagnosis from 30 years associated with infections by type and site due to relatively short follow-up into adulthood and insufficient statistical power. We were also unable to examine the risk of a multiple sclerosis diagnosis associated with specific types of CNS infection due to small numbers of patients with multiple sclerosis and earlier CNS infections (for example, only 555 patients have encephalitis diagnoses in adolescence and only one of them have multiple sclerosis diagnosis).

We did not have information on the first symptomatic onset of multiple sclerosis, so we could not assess multiple sclerosis risk earlier than the first multiple sclerosis diagnosis and excluded individuals with a diagnosis of multiple sclerosis before age 20 years. However, when the first demyelinating disease of the CNS of any type was used to defined possible onset of multiple

sclerosis, the associations were little changed. Only 56 patients with a multiple sclerosis diagnosis from age 20 had a demyelinating disease of the CNS before age 20 years. When these individuals were excluded from the analysis, the HR for risk of a multiple sclerosis diagnosis associated with any infection in adolescence reduced only slightly in magnitude (HR 1.30 95%CI 1.18-1.43). There were also no substantial differences in HR for risk of a multiple sclerosis diagnosis associated with any infection in adolescence between the analysis using the first multiple sclerosis diagnosis from age 20 years and the analysis using the first diagnosis of any demyelinating disease of the CNS from age 20 years to identify possible onset of multiple sclerosis. The measure of multiple sclerosis onset by the first demyelinating disease of the CNS may not identify the true onset of multiple sclerosis pathogenesis, as it may be initially asymptomatic or with non-specific symptoms, indicating the difficulty of identifying onset with absolute certainty in the majority of patients. The sensitivity analysis for multiple sclerosis diagnosis from age 25 years found consistent associations for all of the infections (except for infections of the urogenital tract), which reduces the likelihood of reverse causation. We could not control for some potential confounding factors as data were not available for everyone in this cohort, including prescribed medication, smoking, BMI, HLA type, and vitamin D. We examined infection before age 20 years and most people tend to start smoking more heavily at a later age, so smoking is unlikely to be a major confounding factor. BMI before age 20 was only available for those included in the Swedish Military Conscription Register (31.17% of the population were included here and mainly men). Sensitivity analysis of this subset population revealed that infections in adolescence remained a statistically significant factor for increased risk of a subsequent multiple sclerosis diagnosis after adjusting for BMI, sex, infections in earlier childhood, and parental socioeconomic position. Thus, the association between infection in adolescence and increased risk of a multiple sclerosis diagnosis appears robust and independent of BMI. It remains possible that other characteristics or exposures (such as prescribed medication, vitamin D, or HLA type) may potentially increase both the risk of infections and a multiple sclerosis diagnosis, and this may vary by infection type, although we might expect to see associations with infections before age 11 years if other factors that persist across the life-course explain the association with infections in adolescence.

Conclusion

In conclusion, this study provides further evidence that adolescence, but not earlier childhood, is a period of raised susceptibility to exposures associated with subsequent increased risk of

being diagnosed with multiple sclerosis. This increased risk is associated with various infections, but not all types and sites. Importantly, these associations cannot be entirely explained by previously described infectious risk factors, infectious mononucleosis or pneumonia, nor by the raised risk associated with CNS infections described here.

Funding

This study was supported by grants from the UK Economic and Social Research Council (ESRC) to the International Centre for Life Course Studies (ES/R008930/1) and Nyckelfonden.

Competing interests

Fredrik Piehl has received research grants from Genzyme and Merck KGaA, and fees for serving as Chair of DMC in clinical trials with Parexel. Tomas Olsson has received unrestricted multiple sclerosis research grants and/or honoraria for advisory boards/lectures from Biogen, Novartis, Sanofi, AstraZeneca and Merck. Scott Montgomery has received multiple sclerosis research grants and/or honoraria for advisory boards/lectures from Roche, Novartis, AstraZeneca, Teva and IQVIA.

References

1. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann Neurol*. 2007a; 61: 288-299.
2. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Ann Neurol*. 2007b; 61: 504-513.
3. Ramagopalan SV, Dobson R, Meier UC, Giovannoni G. Multiple sclerosis: risk factors, prodromes, and potential causal pathways. *Lancet Neurol*. 2010; 9: 727-739.
4. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol*. 2017; 13: 25-36.
5. Thacker EL, Mirzaei F, Ascherio A. Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis. *Ann Neurol*. 2006; 59: 499-503.
6. Handel AE, Williamson AJ, Disanto G, Handunnetthi L, Giovannoni G, Ramagopalan SV. An updated meta-analysis of risk of multiple sclerosis following infectious mononucleosis. *PloS one*. 2010; 5: e12496.
7. Santiago O, Gutierrez J, Sorlozano A, de Dios Luna J, Villegas E, Fernandez O. Relation between Epstein-Barr virus and multiple sclerosis: analytic study of scientific production. *Eur J Clin Microbiol Infect Dis*. 2010; 29: 857-866.
8. Almohmeed YH, Avenell A, Aucott L, Vickers MA. Systematic review and meta-analysis of the sero-epidemiological association between Epstein Barr virus and multiple sclerosis. *PloS one*. 2013; 8: e61110.
9. Virtanen JO, Jacobson S. Viruses and multiple sclerosis. *CNS Neurol Disord Drug Targets*. 2012; 11: 528-544.
10. Libbey JE, Cusick MF, Fujinami RS. Role of pathogens in multiple sclerosis. *Int Rev Immunol*. 2014; 33: 266-283.
11. Engdahl E, Gustafsson R, Huang J, et al. Increased serological response against human herpesvirus 6A is associated with risk for multiple sclerosis. *Front Immunol*. 2019; 10: 2715.
12. Swanborg RH, Whittum-Hudson JA, Hudson AP. Infectious agents and multiple sclerosis—are *Chlamydia pneumoniae* and human herpes virus 6 involved?. *J Neuroimmunol*. 2003; 136: 1-8.
13. Bagos PG, Nikolopoulos G, Ioannidis A. *Chlamydia pneumoniae* infection and the risk of multiple sclerosis: a meta-analysis. *Mult Scler J*. 2006; 12: 397-411.

14. Smith KA, Hiyoshi A, Burkill S, et al. Hospital diagnosed pneumonia before age 20 years and multiple sclerosis risk. *BMJ Neurol Open*. 2020; 2: e000044.
15. Fujinami RS, von Herrath MG, Christen U, Whitton JL. Molecular mimicry, bystander activation, or viral persistence: infections and autoimmune disease. *Clin Microbiol Rev*. 2006; 19: 80–94.
16. Odoardi F, Sie C, Streyl K, et al. T cells become licensed in the lung to enter the central nervous system. *Nature*. 2012; 488: 675-679.
17. Montgomery S, Hiyoshi A, Burkill S, Alfredsson L, Bahmanyar S, Olsson T. Concussion in adolescence and risk of multiple sclerosis. *Ann Neurol*. 2017; 82: 554-561.
18. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA*. 2006; 296: 2832-2838.
19. Xu Y, Hiyoshi A, Brand JS, et al. Higher body mass index at ages 16 to 20 years is associated with increased risk of a multiple sclerosis diagnosis in subsequent adulthood among men. *Mult Scler J*. 2020; 27: 147-150.
20. Tremlett H, Marrie RA. The multiple sclerosis prodrome: emerging evidence, challenges, and opportunities. *Mult Scler J*. 2020; 27: 6-12.
21. Montgomery S, Hillert J, Bahmanyar S. Hospital admission due to infections in multiple sclerosis patients. *Eur J Neurol*. 2013; 20: 1153-1160.
22. Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol*. 2016; 31: 125-136.
23. Ludvigsson JF, Andersson E, Ekblom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 2011; 11: 450.
24. Murley C, Friberg E, Hillert J, Alexanderson K, Yang F. Validation of multiple sclerosis diagnoses in the Swedish National Patient Register. *Eur J Epidemiol*. 2019; 34: 1161-1169.
25. Ludvigsson JF, Svedberg P, Olén O, Bruze G, Neovius M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *Eur J Epidemiol*. 2019; 34: 423-437.
26. Miller DH, Weinshenker BG, Filippi M, et al. Differential diagnosis of suspected multiple sclerosis: a consensus approach. *Mult Scler J*. 2008; 14: 1157-1174.
27. Marrie RA, Fisk JD, Stadnyk KJ, et al. The incidence and prevalence of multiple sclerosis in Nova Scotia, Canada. *Can J Neurol Sci*. 2013; 40: 824-831.
28. Marrie RA, Elliott L, Marriott J, et al. Dramatically changing rates and reasons for hospitalization in multiple sclerosis. *Neurol*. 2014; 83: 929-937.

29. Harrison E, Rose D. The European socio-economic classification (ESeC) user guide. Colchester: Institute for Social and Economic Research, University of Essex; 2006
30. Marrie RA, Yu N, Blanchard J, Leung S, Elliott L. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba. *Neurol.* 2010; 74: 465-471.
31. Hansen BT, Kjær SK, Arnheim-Dahlström L, et al. Age at first intercourse, number of partners and sexually transmitted infection prevalence among Danish, Norwegian and Swedish women: estimates and trends from nationally representative cross-sectional surveys of more than 100 000 women. *Acta Obstet Gynecol Scand.* 2020; 99: 175-185.
32. Lindsey JW, Patel S. PCR for bacterial 16S ribosomal DNA in multiple sclerosis cerebrospinal fluid. *Mult Scler J.* 2008; 14: 147-152.
33. Marrodan M, Alessandro L, Farez MF, Correale J. The role of infections in multiple sclerosis. *Mult Scler J.* 2019; 25: 891-901.
34. von Herrath MG, Fujinami RS, Whitton JL. Microorganisms and autoimmunity: making the barren field fertile?. *Nat Rev Microbiol.* 2003; 1: 151-157.
35. Cepok S, Zhou D, Srivastava R, et al. Identification of Epstein-Barr virus proteins as putative targets of the immune response in multiple sclerosis. *J Clin Invest.* 2005; 115: 1352-1360.
36. Lindsey JW. Antibodies to the Epstein-Barr virus proteins BFRF3 and BRRF2 cross-react with human proteins. *J Neuroimmunol.* 2017; 310: 131-134.
37. Tengvall K, Huang J, Hellström C, et al. Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. *Proc Natl Acad Sci.* 2019; 116: 16955-16960.
38. Marqués JM, Rial A, Muñoz N, et al. Protection against *Streptococcus pneumoniae* serotype 1 acute infection shows a signature of Th17-and IFN- γ -mediated immunity. *Immunobiol.* 2012; 217: 420-429.
39. Tzartos JS, Friese MA, Craner MJ, et al. Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. *Am J Pathol.* 2008; 172: 146-155.

Figure legends

Figure 1 Flow diagram of participant selection.

Temporary duplicates are due to the assignment of a previously deceased person's personal identification number to some immigrants to Sweden.

Figure 2 Hazard ratio and 95% confidence interval for risk of a multiple sclerosis diagnosis associated with any infections stratified by age period and models.

Abbreviations: IM = infectious mononucleosis; CI = confidence interval.

Hazard ratios are from the three-step multivariable regression with infectious mononucleosis (step 1), pneumonia (step 2), and CNS infection (step 3) was excluded from the definition of infection. All models included measures of any infections from birth to age 10 years and 11 to 19 years, with adjustment for sex and parental socioeconomic position. Having no infection was the reference group.

Figure 3 Hazard ratio and 95% confidence interval for risk of a multiple sclerosis diagnosis associated with viral infection stratified by age period and models.

Abbreviations: IM = infectious mononucleosis; CI = confidence interval.

Hazard ratios are from the three-step multivariable regression with exclusion of: infectious mononucleosis (step 1), pneumonia (step 2), and CNS infection (step 3). All models included measures of viral infection from birth to age 10 years and 11 to 19 years, with adjustment for sex and parental socioeconomic position. Having no viral infection was the reference group.

Figure 4 Hazard ratio and 95% confidence interval for risk of a multiple sclerosis diagnosis associated with bacterial infection stratified by age period and models.

Abbreviations: IM = infectious mononucleosis; CI = confidence interval.

Hazard ratios are from the three-step multivariable regression with exclusion of: infectious mononucleosis (step 1), pneumonia (step 2), and CNS infection (step 3). All models included measures of bacterial infection from birth to age 10 years and 11 to 19 years, with adjustment for sex and parental socioeconomic position. Having no bacterial infection was the reference group.

Table 1 Characteristics of the birth cohort stratified by multiple sclerosis and hazard ratios for risk of a multiple sclerosis diagnosis

	No multiple sclerosis, <i>n</i> (%) (<i>n</i> = 2 418 947)	With multiple sclerosis, <i>n</i> (%) (<i>n</i> = 4022)	Unadjusted HR ^a (95%CI)
Any infections, birth to age 10 years			
No	1 957 383 (80.92)	3429 (85.26)	1 (reference)
Yes	461 564 (19.08)	593 (14.74)	0.96 (0.88, 1.04)
Any infections, age 11-19 years			
No	2 081 114 (86.03)	3503 (87.10)	1 (reference)
Yes	337 833 (13.97)	519 (12.90)	1.39 (1.26, 1.52)
Infectious mononucleosis, birth to age 10 years			
No	2 414 580 (99.82)	4010 (99.70)	1 (reference)
Yes	4367 (0.18)	12 (0.30)	1.72 (0.98, 3.03)
Infectious mononucleosis, age 11-19 years			
No	2 400 384 (99.23)	3958 (98.41)	1 (reference)
Yes	18 563 (0.77)	64 (1.59)	2.80 (2.19, 3.59)
Pneumonia, birth to age 10 years			
No	2 360 373 (97.58)	3956 (98.36)	1 (reference)
Yes	58 574 (2.42)	66 (1.64)	0.86 (0.67, 1.10)
Pneumonia, age 11-19 years			
No	2 403 185 (99.35)	3995 (99.33)	1 (reference)
Yes	15 762 (0.65)	27 (0.67)	1.45 (0.99, 2.11)
Sex			
Male	1 243 773 (51.42)	1197 (29.76)	1 (reference)
Female	1 175 174 (48.58)	2825 (70.24)	2.52 (2.35, 2.70)
Parental socioeconomic position			
High	838 893 (34.68)	1277 (31.75)	1 (reference)
Intermediate	410 593 (16.97)	663 (16.48)	1.06 (0.96, 1.16)
Low	1 169 461 (48.35)	2082 (51.77)	1.01 (0.94, 1.08)
Follow-up time (years), median (interquartile range)	9.70 (8.04)	11.38 (13.34)	-

HR = hazard ratio; CI = confidence interval.

^aUnadjusted survival analysis where the association between each variable in the table (excluding the median follow-up time) and risk of a multiple sclerosis diagnosis was examined separately.

Table 2 Infections by type and site and hazard ratios for risk of a multiple sclerosis diagnosis

	No multiple sclerosis <i>n</i> (% yes) (<i>n</i> = 2 418 947)	With multiple sclerosis, <i>n</i> (% yes) (<i>n</i> = 4022)	Unadjusted HR ^a (95%CI)	Adjusted HR ^b (95%CI)	Adjusted HR ^c (95%CI)
CNS, birth to age 10 years	11 760 (0.49)	26 (0.65)	1.36 (0.93, 2.00)	1.47 (1.00, 2.16)	
CNS, age 11–19 years	5 492 (0.23)	26 (0.65)	2.85 (1.94, 4.20)	2.80 (1.90, 4.12)	
Gastrointestinal, birth to age 10 years	127 406 (5.27)	198 (4.92)	1.05 (0.91, 1.22)	1.08 (0.93, 1.24)	1.07 (0.93, 1.24)
Gastrointestinal, age 11–19 years	84 406 (3.49)	156 (3.88)	1.11 (0.95, 1.30)	1.10 (0.94, 1.29)	1.09 (0.93, 1.28)
Genitourinary, birth to age 10 years	20 220 (0.84)	28 (0.70)	1.15 (0.79, 1.66)	0.93 (0.64, 1.36)	
Genitourinary, age 11–19 years	16 322 (0.67)	33 (0.82)	2.03 (1.44, 2.86)	1.46 (1.03, 2.06)	
Respiratory, birth to age 10 years	229 060 (9.47)	260 (6.46)	0.88 (0.78, 1.00)	0.93 (0.82, 1.06)	
Respiratory, age 11–19 years	114 286 (4.72)	190 (4.72)	1.59 (1.37, 1.84)	1.51 (1.30, 1.75)	
Skin, birth to age 10 years	33 135 (1.37)	35 (0.87)	0.72 (0.52, 1.00)	0.74 (0.53, 1.03)	
Skin, age 11–19 years	43 024 (1.78)	34 (0.85)	1.07 (0.76, 1.50)	1.00 (0.71, 1.40)	
Bacterial, birth to age 10 years	174 728 (7.22)	199 (4.95)	0.94 (0.82, 1.09)		0.95 (0.82, 1.10)
Bacterial, age 11–19 years	137 209 (5.67)	216 (5.37)	1.52 (1.32, 1.74)		1.28 (1.11, 1.47)
Viral, birth to age 10 years	224 239 (9.27)	251 (6.24)	0.89 (0.79, 1.01)		0.95 (0.84, 1.09)
Viral, age 11–19 years	115 293 (4.77)	162 (4.03)	1.53 (1.31, 1.79)		1.38 (1.17, 1.62)
Other, birth to age 10 years	11 795 (0.49)	4 (0.10)	0.46 (0.17, 1.22)	0.47 (0.18, 1.25)	0.47 (0.17, 1.24)
Other, age 11–19 years	30 871 (1.28)	26 (0.65)	1.43 (0.97, 2.10)	1.13 (0.76, 1.66)	1.04 (0.71, 1.54)

HR = hazard ratio; CI = confidence interval.

^aUnadjusted survival analysis where the association between each variable in the table and risk of a multiple sclerosis diagnosis was examined separately. For each infection by type and site, no such specific infection was the reference group.

^bThis model included measures of CNS, gastrointestinal, genitourinary, respiratory, skin, and the category 'other infections' at ages 0–10 years and 11–19 years, with adjustment for sex and parental socioeconomic position. For each infection by type and site, no such specific infection was the reference group.

^cThis model included measures of gastrointestinal, bacterial, viral, and the category 'other infections' at ages 0–10 and 11–19 years, with adjustment for sex and parental socioeconomic position. For each infection by type and site, no such specific infection was the reference group.

Table 3 Sensitivity analysis-2 with multiple sclerosis diagnosis from age 25 years

	No multiple sclerosis, <i>n</i> (%) (<i>n</i> = 1 832 801)	With multiple sclerosis, <i>n</i> (%) (<i>n</i> = 3221)	Unadjusted HR ^a (95%CI)	Adjusted HR ^b (95%CI)	Adjusted HR ^c (95%CI)
CNS, age 11–19 years					
No	1 828 767 (99.78)	3205 (99.50)	1 (reference)	1 (reference)	
Yes	4034 (0.22)	16 (0.50)	2.18 (1.33, 3.56)	2.16 (1.32, 3.53)	
Genitourinary, age 11–19 years					
No	1 823 095 (99.47)	3201 (99.38)	1 (reference)	1 (reference)	
Yes	9706 (0.53)	20 (0.62)	1.74 (1.12, 2.71)	1.30 (0.84, 2.02)	
Respiratory, age 11–19 years					
No	1 765 645 (96.34)	3087 (95.84)	1 (reference)	1 (reference)	
Yes	67 156 (3.66)	134 (4.16)	1.54 (1.30, 1.83)	1.49 (1.25, 1.77)	
Bacterial, age 11–19 years					
No	1 751 221 (95.55)	3069 (95.28)	1 (reference)		1 (reference)
Yes	81 580 (4.45)	152 (4.72)	1.47 (1.25, 1.73)		1.24 (1.04, 1.47)
Viral, age 11–19 years					
No	1 771 591 (96.66)	3098 (96.18)	1 (reference)		1 (reference)
Yes	61 210 (3.34)	123 (3.82)	1.66 (1.39, 1.99)		1.53 (1.27, 1.85)

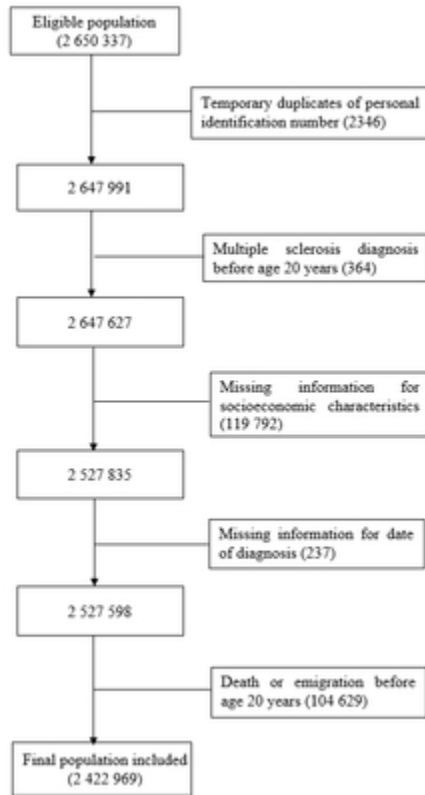
HR = hazard ratio; CI = confidence interval.

Only infections significantly associated with increased risk of a multiple sclerosis diagnosis from age 20 years (in Table 2) were reported here.

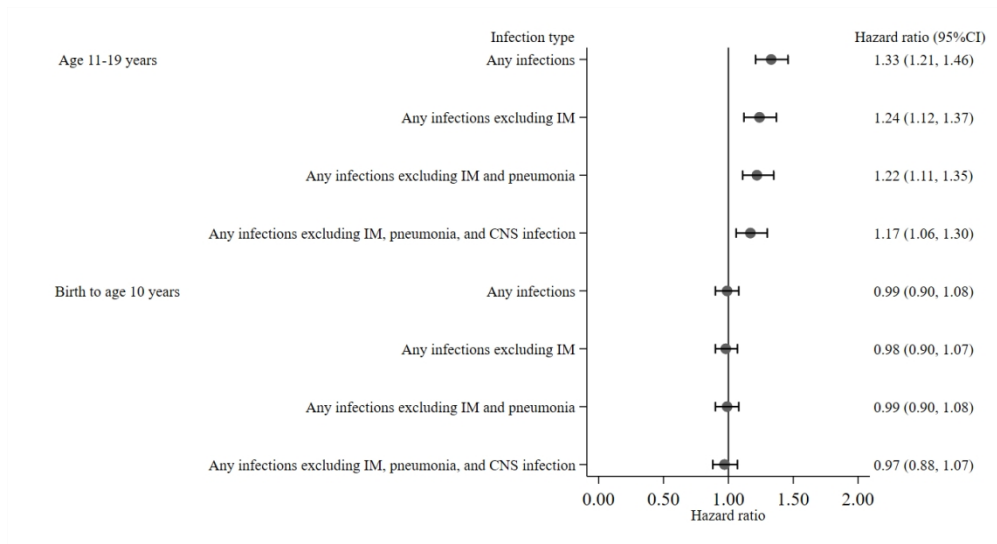
^aUnadjusted survival analysis where the association between each variable in the table and risk of a multiple sclerosis diagnosis was examined separately. For each infection by type and site, no such specific infection was the reference group.

^bThis model included measures of CNS, gastrointestinal, genitourinary, respiratory, skin, and the category 'other infections' at ages 0–10 years and 11–19 years, with adjustment for sex and parental socioeconomic position.

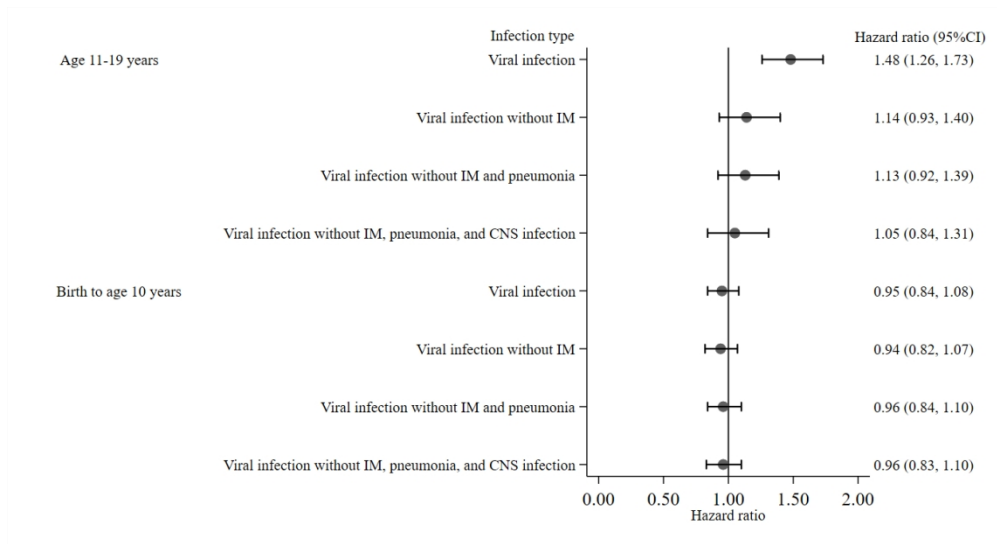
^cThis model included measures of gastrointestinal, bacterial, viral, and the category 'other infections' at ages 0–10 and 11–19 years, with adjustment for sex and parental socioeconomic position.



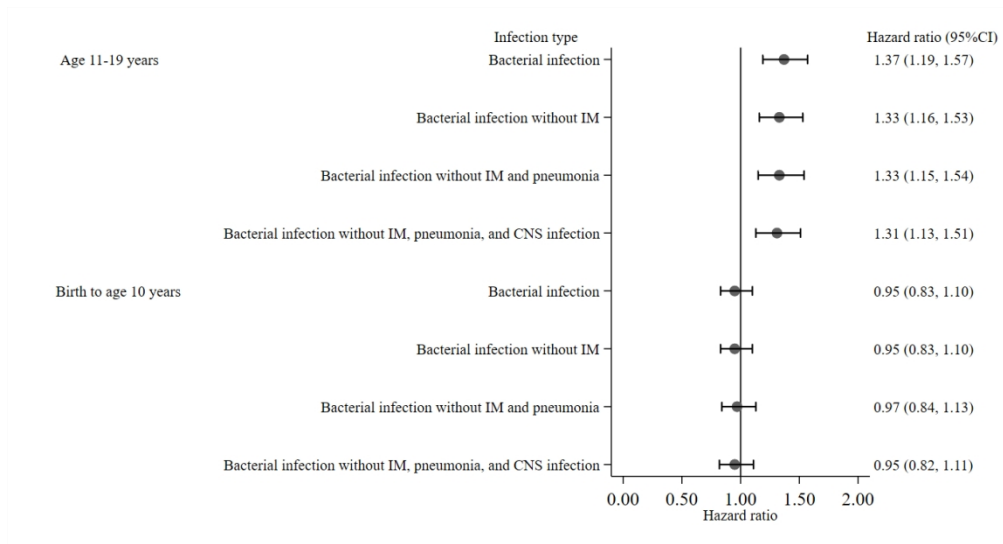
9x17mm (600 x 600 DPI)



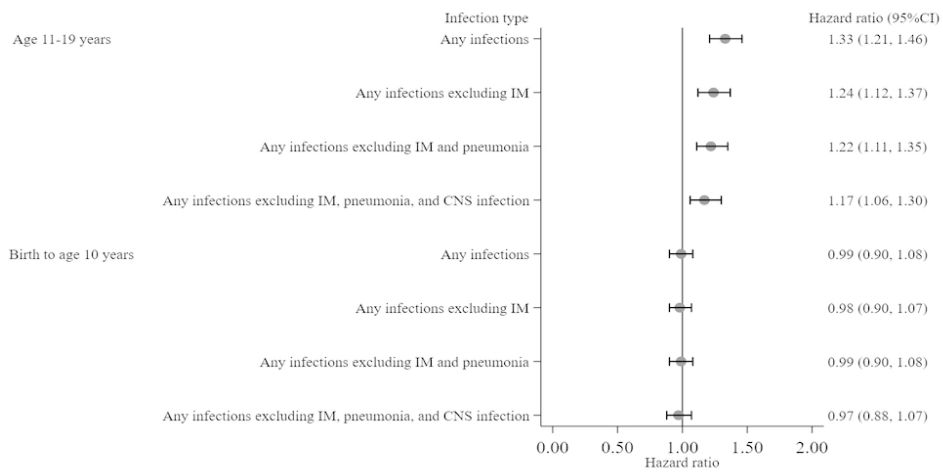
370x199mm (96 x 96 DPI)



370x199mm (96 x 96 DPI)



370x199mm (96 x 96 DPI)



Hazard ratio and 95% confidence interval for risk of a multiple sclerosis diagnosis associated with any infections stratified by age period and models.

361x180mm (72 x 72 DPI)

STROBE statement: Reporting guidelines checklist for cohort, case-control and cross-sectional studies

SECTION	ITEM NUMBER	CHECKLIST ITEM	REPORTED ON PAGE NUMBER:
TITLE AND ABSTRACT			
	1a	Indicate the study's design with a commonly used term in the title or the abstract	2
	1b	Provide in the abstract an informative and balanced summary of what was done and what was found	2
INTRODUCTION			
Background and objectives	2	Explain the scientific background and rationale for the investigation being reported	4-5
	3	State specific objectives, including any pre-specified hypotheses	6
METHODS			
Study design	4	Present key elements of study design early in the paper	6-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-7
Participants	6a	Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	6-7
	6b	Cohort study—For matched studies, give matching criteria and number of exposed and unexposed Case-control study—For matched studies, give matching criteria and the number of controls per case Variables	Non-applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-8
Data sources/measurements	8*	For each variable of interest, give sources of data and details of methods of assessment	7-8

SECTION	ITEM NUMBER	CHECKLIST ITEM	REPORTED ON PAGE NUMBER:
		(measurement). Describe comparability of assessment methods if there is more than one group.	
Bias	9	Describe any efforts to address potential sources of bias.	8-10
Study size	10	Explain how the study size was arrived at	6-7, figure1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.	7-10
Statistical methods	12a	Describe all statistical methods, including those used to control for confounding	8-10
	12b	Describe any methods used to examine subgroups and interactions	9
	12c	Explain how missing data were addressed	6-7, figure1
	12d	Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	8
	12e	Describe any sensitivity analyses	9-10
RESULTS			
Participants	13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6-7, figure1
	13b	Give reasons for non-participation at each stage	Non-applicable
	13c	Consider use of a flow diagram	7, figure 1
Descriptive Data	14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10
	14b	Indicate number of participants with missing data for each variable of interest	6-7, figure1
	14c	Cohort study—Summarise follow-up time (eg, average and total amount)	10
Outcome Data	15*	Cohort study—Report numbers of outcome events or summary measures over time Case-control study—Report numbers in each exposure category, or summary measures of exposure Cross-sectional study—Report numbers of outcome events or summary measures	10

SECTION	ITEM NUMBER	CHECKLIST ITEM	REPORTED ON PAGE NUMBER:
Main Results	16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g. 95% confidence interval). Make clear which confounders were adjusted for and why they were included	10-11
	16b	Report category boundaries when continuous variables were categorized	Non-applicable
	16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Non-applicable
	16d	Report results of any adjustments for multiple comparisons	Non-applicable
Other Analyses	17a	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses	11
	17b	If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken	Non-applicable
	17c	If detailed results are available elsewhere, state how they can be accessed	Non-applicable
DISCUSSION			
Key Results	18	Summarise key results with reference to study objectives	12-13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16-17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-17
Generalisability	21	Discuss the generalisability (external validity) of the study results Other information	15
FUNDING			
	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Xu *et al.* show that a variety of serious infections in adolescence – not limited to infectious mononucleosis, pneumonia and CNS infections – are associated with increased risk of a subsequent multiple sclerosis diagnosis.