**Sibship structure and risk of infectious mononucleosis: a population-based cohort study**

Klaus Rostgaard,1* Trine Rasmussen Nielsen,1,2 Jan Wohlfahrt,1 Henrik Ullum,3 Ole Pedersen,4 Christian Erikstrup,5 Lars Peter Nielsen6 and Henrik Hjalgrim1

1Department of Epidemiology Research and 2Department of Regulatory and Medical Affairs, Statens Serum Institut, Copenhagen, Denmark, 3Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark, 4Department of Clinical Immunology, Naestved Hospital, Naestved, Denmark, 5Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark and 6Department of Microbiologic Diagnostics and Virology, Statens Serum Institut, Copenhagen, Denmark

*Corresponding author. Department of Epidemiology Research, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark. E-mail: klp@ssi.dk

Accepted 19 May 2014

**Abstract**

**Background:** Present understanding of increased risk of Epstein-Barr virus (EBV)-related infectious mononucleosis among children of low birth order or small sibships is mainly based on old and indirect evidence. Societal changes and methodological limitations of previous studies call for new data.

**Methods:** We used data from the Danish Civil Registration System and the Danish National Hospital Discharge Register to study incidence rates of inpatient hospitalizations for infectious mononucleosis before the age of 20 years in a cohort of 2543225 Danes born between 1971 and 2008, taking individual sibship structure into account.

**Results:** A total of 12872 cases of infectious mononucleosis were observed during 35.3 million person-years of follow-up. Statistical modelling showed that increasing sibship size was associated with a reduced risk of infectious mononucleosis and that younger siblings conferred more protection from infectious mononucleosis than older siblings. In addition to this general association with younger and older siblings, children aged less than 4 years transiently increased their siblings' infectious mononucleosis risk. Our results were confirmed in an independent sample of blood donors followed up retrospectively for self-reported infectious mononucleosis.

**Conclusions:** Younger siblings, and to a lesser degree older siblings, seem to be important in the transmission of EBV within families. Apparently the dogma of low birth order in a sibship as being at the highest risk of infectious mononucleosis is no longer valid.

**Key words:** Epstein-Barr virus, infection pattern, delayed infection, birth order, Denmark
**Key Messages**

- Within sibships each additional sibling, especially younger siblings, is associated with lowered risk of infectious mononucleosis.
- The smaller the age difference between sibs, the larger the mutual reduction in risk of infectious mononucleosis.
- An additional sibling may be associated with up to 25% reduction in risk of infectious mononucleosis, despite the numerous alternative sources of EBV transmission.
- Similar results have been shown for risk of multiple sclerosis and EBV-related Hodgkin lymphoma.

**Introduction**

Infection with Epstein-Barr virus (EBV) is ubiquitous, and by adulthood 90% or higher of persons are infected, as illustrated by the prevalence among controls in numerous case-control studies.\(^1\)\(^-\)\(^4\) Primary infection with EBV in childhood is usually asymptomatic or accompanied only by mild symptoms. On the other hand, primary EBV infection in adolescence or adulthood is often accompanied by infectious mononucleosis (IM), the estimated proportion in restricted populations ranging from 25% to 77%\(^5\)\(^,\)^\(^6\). All this suggests that host factors, e.g. genetic constitution, in interaction with age-dependent environmental and behavioural factors, influence risk of IM. These host factors and age-dependent factors are very incompletely understood, e.g. the mechanism(s) of EBV transmission among pre-adolescent children are not well known\(^7\).

Studies have shown that sibship structure and interaction may influence the age at which individuals become exposed to infectious agents\(^8\)\(^,\)^\(^9\). For EBV, current understanding proposes that age at primary infection, and by implication the risk of IM, is higher for children of lower birth order and smaller sibships. This understanding has mainly been articulated in papers on cancer and autoimmune diseases, see e.g. Westergaard et al\(^10\)\(^,\)^\(^11\), Vianna et al\(^12\), Hernan et al\(^13\) and references in these. Remarkably, this interpretation seems to rest primarily on inferences made from studies of other infectious agents, such as hepatitis A, hepatitis B and polio virus\(^14\)\(^-\)\(^16\), whereas direct evidence is rather limited\(^17\). Moreover, in many Western societies marked secular changes have taken place during recent decades, which may have changed the role of family structure in infectious disease pressure in childhood. For instance, the proportion of children attending pre-school has vastly increased\(^18\), pre-school exposure replacing older sibs as an important source of infections\(^19\)\(^,\)^\(^20\). Accordingly, older findings for infections other than EBV may no longer be applicable to the epidemiology of clinical EBV infections in general and its presumed association with sibship structure in particular. Based on many epidemiological observations, we have hypothesized the existence of a set of common risk factors for IM, EBV-positive Hodgkin lymphoma and multiple sclerosis that include constitutionally determined ‘inadvertent’ immunological responses to EBV\(^21\). If that hypothesis is true, putative associations between sibship structure and IM risk will also be of relevance and an inspiration in aetiological studies of Hodgkin lymphoma and multiple sclerosis.

We undertook the present study to investigate the significance of sibship structure on the age-specific risk of IM at age 0–19 years in a large population-based cohort in Denmark.

**Methods**

The nationwide Danish Civil Registration System (CRS) was implemented on 1 April 1968. All Danish citizens have since been assigned unique identification numbers (the CRS number), by which the CRS continuously monitors individual vital status, emigration status, identity of parents and residence\(^22\).

Using the CRS, we established a cohort of all Danish singletons born between 1 January 1971 and 31 December 2008. For each cohort member we identified sibs, defined as children born to the same mother as the individual. For all persons, information about dates of birth was obtained to allow the construction of variables such as the age difference between a proband and an exposing sib.

The Danish National Hospital Discharge Register was established in 1977 and has since recorded 99.9% of all discharges from Danish non-psychiatric hospitals\(^23\)\(^,\)^\(^24\). For each hospitalization, the register contains information on dates of admission and discharge, and discharge diagnoses classified according to the International Classification of Diseases, 8th revision (ICD-8) from 1977 to 1993, and the 10th revision (ICD-10) thereafter. In the Register we identified all inpatient hospitalizations of IM, classified as code 075* in ICD-8 and code B27* in ICD-10.

The study was approved by the Danish Data Protection Board, J.nr 2008-54-0472. According to Danish law,
informed consent is unnecessary for conducting a purely register-based study.

Statistical analysis

The cohort of singletons was followed for hospitalization for IM from 1 January 1977 or from date of birth, whichever occurred last, until the date of diagnosis of IM, death, emigration, age 20 or 31 December 2008, whichever came first. Exclusively in incidence calculations (Figures 1 and 2), the cohort was extended to include all persons in the CRS born since 1957 and alive in 1977.

Log-linear Poisson regression was used to model incidence rates and rate ratios (RR) and thereby assess the effects of sibship structure. The variables investigated included the number of sibs with a specific age and the number of sibs with a specific signed age interval to the proband. At the crudest, the latter amounted to counting the number of older and younger sibs of the proband. All these variables were treated as time-dependent variables, e.g. changing when new members were born into the sibship. All analyses were adjusted for birth cohort (calendar year of birth), sex-specific age (1-year groups) and sex-specific period (calendar year). Of importance, the identification problems normally relevant to the simultaneous inclusion of these three intimately related time scales for effect estimation are irrelevant for confounder control.

Based on present understanding of the epidemiology of EBV and consistent with the incidence data, risk factor analyses were carried out for children aged less than 7 years and for older children and adolescents separately as well as combined.

As a preliminary step, we analysed data restricted to children being followed up while they had at most one sibling, to find out which sibling variables and categorizations of these to use. The starting point of this exercise was a detailed table of rate ratios by the age of the child crossed with the age of the child’s sibling. This set-up is similar to that of age-period-cohort analyses, and similar techniques were used to obtain a parsimonious and interpretable data description. To limit the number of parameter estimates to present while, on the other hand, having no a priori hypothesis, we used the corrected Akaike criterion (AICC) and the Bayesian information criterion (BIC) with the number of outcome events as \( n \), both yielding a post hoc justification of our choices. The final set of predictors was number of sibs aged 0, 1, 2 and 3 years and the number of sibs with a specific age interval to the proband.

All analyses were performed using the SAS statistical software package (versions 9.2 and 9.3; SAS Institute, Cary, NC, USA); 95% confidence intervals (CIs) were based on Wald tests. Tests for effect modification were evaluated by inclusion of an interaction term, and all tests were likelihood-ratio based.
Results

The distribution at person-years of risk and the observed number of outcomes in the studied cohort of singletons are shown in Table 1. Overall, 12,872 persons were hospitalized for IM during 35.3 million person-years of follow-up contributed by 2,543,225 Danes. Age- and gender-specific incidence rates of IM are shown in Figure 1. IM displayed a bimodal age-specific incidence rate distribution in boys with peaks around the ages of 4 and 16 years, whereas for girls, the age-specific incidence rate distribution of mononucleosis was unimodal and peaked at 15 years. The 4-year peak for boys is statistically ‘real’, e.g. a trend test restricted to age 4–10 years yields a declining incidence rate of 10% per year (RR \(=0.90\)\), 95% confidence interval \(=0.88–0.92; P = 3 \times 10^{-17}\). The age- and sex-specific incidence rates of IM did not change materially over the study period, see Figure 2.

Table 2 shows the rate ratio of IM according to the number of sibs aged 0, 1, 2 and 3 years and the number of sibs with a specific age interval to the proband for two age categories of the proband. Having a very young sib conveyed a substantial increase in relative risk of IM, with risk estimates around 1.5. Otherwise, each extra sib conveyed a multiplicative protective effect which varied by sibling age interval. Hence, the protective effect was largest if the sib was around the same age as the proband, and negligible if the sib was at least 9 years older than the proband. A reduction of the detailed categorization of age interval to each sib to just counting the number of younger and older sibs was not acceptable statistically (0–6 years old \(P = 0.02\), 7–19 years old \(P = 0.0001\), 0–19 years old \(P = 0.00002\)); however, for completeness these estimates are also presented together with a birth order effect estimate in Table 2. Data were compatible with a common set of sibling parameters for the two categories of attained age, i.e. 0–6 and 7–19 years (\(P = 0.06\)), which therefore is the model used and referred to later. We found no interaction between calendar period (1977–86, 1987–96, 1997–2008) and sibship structure (\(P = 0.26\), data not shown), nor between sex and sibship structure (\(P = 0.77\), data not shown).

Discussion

The present analyses showed that family structure has retained importance to the epidemiology of IM and therefore EBV infection/transmission regardless of the changing roles of sibs and other children in modern affluent societies. An increasing proportion of children attend day care, which constitutes an important source of infection for other infectious diseases.19,20 This should imply a trend towards earlier infection with EBV and thus a decreased occurrence of IM in adolescents. Despite having a 31-year-long study period, the incidence pattern of hospitalized adolescent cases did not markedly change over time, as seen in other hospitalized or monospot-tested populations including adults.26–28

Under the assumption that the observed age-specific rate of hospitalized IM mirrors (by proportionality) the pattern for IM in general, see Visser et al. for a recent
example, our observations suggested that young siblings aged less than 4 years constitute an important source of primary EBV infections within families. Accordingly, the statistical analyses showed that exposure to a younger sibling in this age range was associated with a transiently increased risk of IM. Most likely, the exposure to younger siblings aged 0–3 years also carries an increased risk of asymptomatic EBV infection. This would explain why, in the long run, exposure to a younger sibling was associated with a decreased risk of IM. Specifically, by being infected by the younger sibling, the older sibling (proband) would no longer be at risk of IM in adolescence when it (IM) is most common. The long-term protective effect varied by the age difference between siblings in such a way that the farther apart from each other in age, the less likely they were to confer protection on each other. Presumably the closer in age, the more likely sibs are to be in contact with each other in a way conducive to EBV transmission.

Obviously contagion can go in both directions, i.e. from an older to a younger sibling, as also suggested by the analyses showing a reduced risk of IM by older siblings displaying a qualitatively similar sibling interval-dependent variation as the younger sibling effect. Interestingly, the estimates of the protective effect of sibs in different age intervals relative to the proband suggested that it is more common that a younger sib infects an older sib (giving rise to asymptomatic disease) than vice versa. This interpretation is consistent with that of two recent studies reporting that sibling contact reduced the risk of IM and increased the risk of EBV seropositivity, respectively. Moreover, it is also supported by the distribution of older and younger siblings in young persons tested for primary EBV infection at Statens Serum Institut (see Supplementary data available at IJE online).

If the hypothesis of a set of common risk factors for IM, EBV-positive Hodgkin lymphoma and multiple sclerosis is true, then the results of the present study are also of relevance for the study of Hodgkin lymphoma and multiple sclerosis. The common use in epidemiological studies of birth order as a proxy for childhood infections seems to be inferior to counting the number of older as well as the number of younger sibs, mutually adjusted. We also note that large recent studies of sibship structure and risk of multiple sclerosis that have made the

### Table 2. Rate ratio (RR) of infectious mononucleosis by sibship characteristics.

<table>
<thead>
<tr>
<th>Sibship</th>
<th>Age 0-6 years RR (95% CI)</th>
<th>Age 7-19 years RR (95% CI)</th>
<th>Age 0-19 years RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per sib aged 0-3 years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 0</td>
<td>1.04 (0.82–1.32)</td>
<td>1.14 (0.95–1.37)</td>
<td>1.20 (1.07–1.34)</td>
</tr>
<tr>
<td>Age 1</td>
<td>1.45 (1.17–1.81)</td>
<td>1.44 (1.24–1.67)</td>
<td>1.57 (1.42–1.73)</td>
</tr>
<tr>
<td>Age 2</td>
<td>1.21 (0.98–1.49)</td>
<td>1.55 (1.36–1.77)</td>
<td>1.44 (1.31–1.59)</td>
</tr>
<tr>
<td>Age 3</td>
<td>1.05 (0.87–1.26)</td>
<td>1.22 (1.07–1.39)</td>
<td>1.16 (1.05–1.28)</td>
</tr>
<tr>
<td>Per sib by age interval</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 9 years younger</td>
<td>NA</td>
<td>0.81 (0.75–0.87)</td>
<td>0.81 (0.76–0.87)</td>
</tr>
<tr>
<td>6-9 years younger</td>
<td>1.19 (0.69–2.03)</td>
<td>0.85 (0.80–0.90)</td>
<td>0.85 (0.80–0.90)</td>
</tr>
<tr>
<td>3-5 years younger</td>
<td>1.01 (0.81–1.27)</td>
<td>0.81 (0.77–0.85)</td>
<td>0.82 (0.79–0.86)</td>
</tr>
<tr>
<td>0-2 years younger</td>
<td>0.80 (0.66–0.95)</td>
<td>0.76 (0.72–0.80)</td>
<td>0.75 (0.72–0.79)</td>
</tr>
<tr>
<td>0-2 years older</td>
<td>0.81 (0.74–0.90)</td>
<td>0.89 (0.85–0.94)</td>
<td>0.87 (0.84–0.91)</td>
</tr>
<tr>
<td>3-5 years older</td>
<td>0.94 (0.87–1.02)</td>
<td>0.94 (0.90–0.98)</td>
<td>0.94 (0.90–0.97)</td>
</tr>
<tr>
<td>6-8 years older</td>
<td>0.90 (0.81–1.00)</td>
<td>0.97 (0.92–1.02)</td>
<td>0.96 (0.91–1.00)</td>
</tr>
<tr>
<td>At least 9 years older</td>
<td>1.02 (0.95–1.10)</td>
<td>0.96 (0.93–1.00)</td>
<td>0.98 (0.94–1.01)</td>
</tr>
<tr>
<td>Summary effects of sibs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of younger sibs</td>
<td>0.88 (0.74–1.04)</td>
<td>0.80 (0.78–0.83)</td>
<td>0.80 (0.78–0.83)</td>
</tr>
<tr>
<td>Number of older sibs</td>
<td>0.93 (0.89–0.98)</td>
<td>0.95 (0.93–0.97)</td>
<td>0.95 (0.93–0.96)</td>
</tr>
<tr>
<td>Birth order</td>
<td>0.93 (0.89–0.97)</td>
<td>1.00 (0.98–1.03)</td>
<td>0.99 (0.97–1.01)</td>
</tr>
</tbody>
</table>

All estimates are mutually adjusted and adjusted for age, gender, calendar period and birth cohort. NA, not applicable.

*a* Obtained by replacing the ‘age interval to proband’ variables in the model.

*b* Obtained by replacing all the sibship variables in the model.
distinction between older and younger siblings have found associations very similar in size and direction to ours, not necessarily appreciating the finding. Similarly, a stronger protective role for younger siblings than for older siblings has been found in EBV-related Hodgkin lymphoma in younger adults, whereas neither of these factors was associated with risk of EBV-unrelated Hodgkin lymphoma.33

Little is known about risk factors for IM and hence potential confounders of the studied relations. It has been hypothesized that the individual’s response to EBV is influenced by a number of factors, including age at infection, infectious load, efficiency of the immune system in controlling the virus infection, and the genetic composition of the individual.34 Socio-economic affluence in childhood has previously been implicated as a risk factor for IM in adolescence.17 To some extent this association has been believed to reflect the presumed association with small sibship size and isolated upbringing and thus delayed exposure to EBV. Presumably sibship structure is intimately linked with age at infection and infectious load in such a way as to make these mediators between sibship structure and IM rather than confounders to be adjusted for. Hypothetically, genes linked to sibship structure (e.g. regulating fecundity) could be in linkage disequilibrium with genes involved in, for example, handling EBV infection. However, very little is known about the genetics of IM at the time of writing.7,21 The efficiency of the immune system in controlling the virus infection will be determined partly by genetics, partly by the timing, type and size of various stimuli.35 Again we consider the latter factors mostly mediators of the association between sibship structure and IM.

The Danish population is ethnically very homogeneous; Denmark has one of the most equal income distributions in the world;24 and hospital treatment is free in Denmark.24 All this mitigates hospitalized IM cases being disproportionately recruited from well-connected affluent social classes with low thresholds for seeking medical attention. All in all, we see little potential for confounding of our results by any of these risk factors.

Our study has a number of strengths and limitations to consider. Among the strengths are the population-based setting and the size of the studied cohort. Also, secular changes in patterns and severity of infection could be taken into account by adjusting for possible calendar period and birth cohort effects. The age distribution of the hospitalized IM cases conformed to expectations regarding the full population of all incident IM cases. We can think of only two realistic limitations of the present study. First, many parents would agree that the threshold for reacting to disease symptoms in their offspring is increased with increasing birth order.37,38 If so, this would translate into later-born siblings being at lower risk of being hospitalized in case of getting IM than earlier-born. What we have found in general is the opposite. This means that our finding of younger siblings as being more protective than older siblings cannot be an artefact of this hypothesized birth order-dependent hospitalization threshold effect. Post hoc modelling yielded the same conclusion about the absence of a birth order effect in the hypothesized direction (data not shown). By the same token, although we did not validate the diagnoses of IM, diagnostic misclassification is unlikely to influence our findings materially. The fraction of misclassified IM cases in the Danish National Patient Register is likely to be substantial, but no specific study on this topic exists. Second, if we apply our estimates of the effect of the 0–3-year-old siblings as direct sources of hospitalized IM, to IM in general, the result may be exaggerated. According to Aaby and coworkers (see e.g. Nielsen39), a case of infectious disease is more likely to be severe when the source of the infection is inside the household than outside the household, due to prolonged and close contact with the source of contagion, leading to exposure to a larger dose of the infectious agent. So a severe, i.e. hospitalized, case of IM would be more likely to originate from within-family contagion than a mild case of IM. However, a comparison with the association between occurrence and age at self-reported IM and sibship structure in a cohort of 18 906 Danish blood donors40 showed remarkably similar magnitude and direction of associations between the two cohorts (see Supplementary data, available at IJE online). Thus, our results are unlikely to be an artefact of using exclusively hospitalized IM as outcome.

Conclusion

In conclusion, family structure has retained importance to the epidemiology of IM regardless of the changing roles of sibs and other children in modern affluent societies. We believe that the presented model captures the typical modes of infection in an environment that is affluent yet with exposure to many other children very early in childhood, e.g. in kindergartens. We are therefore confident that the most important finding of this study, that increasing sibship size, especially by younger sibs, reduces the risk of IM in adolescence, is true of hospitalized IM as well as IM in general, in modern affluent societies. Seemingly, the dogma of the firstborn in a sibship as being at the highest risk of IM is no longer valid.

Supplementary Data

Supplementary data are available at IJE online.
**Funding**

This work was supported by the Danish Cancer Society [grant number DP08155] and the Danish Council for Independent Research [grant number 09-072164].

Conflict of interest: None.

**References**


