A Genetic Basis for Infectious Mononucleosis: Evidence From a Family Study of Hospitalized Cases in Denmark

Klaus Rostgaard, Jan Wohlfahrt, and Henrik Hjalgrim

Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark

(See the Editorial Commentary by Balfour on pages 1690–1.)

Background. Circumstantial evidence from genome-wide association and family studies of various Epstein-Barr virus–associated diseases suggests a substantial genetic component in infectious mononucleosis (IM) etiology. However, familial aggregation of IM has scarcely been studied.

Methods. We used data from the Danish Civil Registration System and the Danish National Hospital Discharge Register to study rate ratios of IM in a cohort of 2,823,583 Danish children born between 1971 and 2011. Specifically, we investigated the risk of IM in twins and in first-, second-, and third-degree relatives of patients with IM. In the analyses, IM was defined as a diagnosis of IM in a hospital contact. Effects of contagion between family members were dealt with by excluding follow-up time the first year after the occurrence of IM in a relative.

Results. A total of 16,870 cases of IM were observed during 40.4 million person-years of follow-up from 1977 to 2011. The rate ratios and the associated 95% confidence intervals were 9.3 (3.0–29) in same-sex twins, 3.0 (2.6–3.5) in siblings, 1.9 (1.6–2.2) in children, 1.4 (1.3–1.6) in second-degree relatives, and 1.0 (0.9–1.2) in third-degree relatives of IM patients. The rate ratios were very similar for IM in children (aged 0–6 years) and older children/adolescents (aged 7–19 years).

Conclusions. We found evidence of familial aggregation of IM that warrants genome-wide association studies on IM disease etiology, especially to examine commonalities with causal pathways in other Epstein-Barr virus–related diseases.

Keywords. family study; children; cohort; contagion; genetics.
shown to cluster with each other within families [14, 15], consistent with a mutual set of risk factors for the 2 conditions. We hypothesize that these common risk factors include constitutionally determined “inadvertent” immunologic responses to EBV, which, in addition to HL and MS, may also lead to IM. This interpretation of the association with HL and MS suggests that there should be a substantial genetic component and familial clustering in IM [16].

One recommendation from a National Institutes of Health meeting on EBV vaccine research was the “prevention of infectious mononucleosis and EBV-associated cancers, facilitated by identification of disease-predictive surrogate markers.” [6]. Therefore, understanding the etiology of IM, including evidence of genetic predisposition, could be important to efforts to prevent a host of EBV-related diseases—for example, through design of vaccines and biomarker-based screening for early-stage EBV-associated malignancies. Whereas there is an abundance of studies on genetic influences on other EBV-related diseases and traits [17–22], the number and scope of studies of familial or genetic exposures on the risk of IM itself is small [23–26]. To examine whether IM clusters in families beyond contagion, we therefore assessed the relative risks of IM after IM diagnosis in a twin or a first-, second-, or third-degree relative in the cohort of all Danes born in the period 1971–2011.

**MATERIAL AND METHODS**

The nationwide Danish Civil Registration System (CRS) was implemented on 1 April 1968. All Danish citizens have since been assigned a unique identification number (the CRS number), by which the CRS continuously monitors individual vital status, emigration status, identity of parents, and residence (see Odumade et al and references therein [4]). To avoid inflation of the estimated rate ratios (RRs) associated with having an affected relative, we therefore excluded follow-up during the first year after IM in a relative in all analyses.

Log-linear Poisson regression was used to model incidence rates and RRs and thereby assess the associations between risk of IM and having an affected relative (ie, a relative having had IM). The variables investigated included having an affected relative of a given type and years since the most recent occurrence of IM in such type of relative examined as a log-linear trend. All these variables were time-dependent, that is, potentially changing when new relatives were born or became affected. All RR s were adjusted for birth cohort (calendar year of birth), sex-specific age (1-year groups), sex-specific period (calendar year), and having a relative of the type considered in the exposure (eg, the RR when having an affected older sibling was adjusted for “having an older sibling”).

Associations were assessed for all ages (0–19 years) and by age group (0–6 and 7–19 years). All analyses were performed using SAS statistical software version 9.3 (SAS Institute, Cary, North Carolina). Two-sided P values and 95% confidence intervals (CIs) were based on Wald tests. Tests for effect modification

---

**Statistical Analysis**

The cohort was followed for IM from 1 January 1977 or date of birth, whichever occurred last, until the date of IM, death, emigration, age 20, or 31 December 2011, whichever occurred first. Patients remain highly contagious for several months after IM (see Odumade et al and references therein [4]). To avoid inflation of the estimated rate ratios (RRs) associated with having an affected relative, we therefore excluded follow-up during the first year after IM in a relative in all analyses.

Log-linear Poisson regression was used to model incidence rates and RRs and thereby assess the associations between risk of IM and having an affected relative (ie, a relative having had IM). The variables investigated included having an affected relative of a given type and years since the most recent occurrence of IM in such type of relative examined as a log-linear trend. All these variables were time-dependent, that is, potentially changing when new relatives were born or became affected. All RRs were adjusted for birth cohort (calendar year of birth), sex-specific age (1-year groups), sex-specific period (calendar year), and having a relative of the type considered in the exposure (eg, the RR when having an affected older sibling was adjusted for “having an older sibling”).

Associations were assessed for all ages (0–19 years) and by age group (0–6 and 7–19 years). All analyses were performed using SAS statistical software version 9.3 (SAS Institute, Cary, North Carolina). Two-sided P values and 95% confidence intervals (CIs) were based on Wald tests. Tests for effect modification
were performed by inclusion of interaction terms and were all likelihood-ratio based. To maximize power, the test for homogeneity by age (0–6 vs 7–19) of the RRs associated with having affected relatives were based solely on estimates regarding the 4 major groups of affected relatives (parents, siblings, half-siblings, and uncles/aunts) in a joint model. Similarly, all tests for effect modification were performed without stratification by age group.

RESULTS

Overall, 16 870 IM cases occurred during 40.4 million person-years of follow-up in our cohort. Overall, 88% of cases had IM as the main diagnosis in the relevant hospital contact, and 85% were inpatient admissions. In all the following, we exclude follow-up time during the first year since IM in any affected relative, thereby excluding 7.4% of the outcome events among exposed, corresponding to 0.38% of all outcome events in the cohort.

The table shows the RR of IM according to type of affected relative. The RR varied from a large and statistically highly significant, 9.3 for same-sex twins to 1.00 for third-degree relatives. Mostly, the estimated RRs associated with having an affected relative were a little larger in children aged 0–6 years than in children aged 7–19 years (P = .03). RRs associated with having different types of affected relatives with the same degree of relatedness could not be assumed to be homogeneous (P < .0001). Equivalent maternal and paternal RRs could not be assumed to be homogeneous (P = .006, data not shown). This seemed due to different RRs associated with having affected maternal and affected paternal half-siblings. Other equivalent maternal and paternal familial associations (ie, through maternal/paternal parents, uncles and aunts, grandparents, and cousins) could be assumed similar (P = .06, data not shown). There was no variation by sex of the exposed in the RR associated with having a given type of affected relative (P = .59, data not shown).

Next we assessed variation in associations by time since IM in the affected relative. Based on the above results, we only considered first-degree affected relatives as an input to the discussion. The changes in RR per year since IM in a relative were RR = 1.01 (95% CI, .99–1.04; P = .34) for parents as affected relative, and RR = 0.94 (95% CI, .91–.97; P = .0002) for siblings as affected relative. Following on from this, we also assessed the RRs when follow-up started at least 5 years after the relative was diagnosed. As expected from the above trends, this yielded slightly higher estimates for the affected parent association than those shown in Table 1 (data not shown), and mostly somewhat lower estimates for the affected sibling association: 0–19 years, RR = 2.57 (95% CI, 2.08–3.17); 7–19 years, RR = 2.46 (95% CI, 1.98–3.06); 0–6 years, RR = 4.39 (95% CI, 1.82–10.6).

DISCUSSION

This is the first study to address familial clustering of IM, apart from a twin study [26] by Hwang et al. In the cohort of all
Danes born in 1971–2011, having relatives who had been hospitalized with IM was associated with an increased risk of being hospitalized with IM before the age of 20 years. The relative risk of IM varied by degree of relatedness, from an imprecise RR of 9.3 in same-sex twins to 1.0 for third-degree relatives. The observed familial clustering is unlikely to be a chance phenomenon, but rather points to shared environmental or genetic risk factors for IM or a combination thereof.

Sociodemographic affluence has long been considered to be associated with an increased risk of IM through simple postponement of primary EBV infection until adolescence [32]. Therefore, the observed associations in our study could be partly due to sociodemographic correlation between relatives. However, the inconspicuous RR associated with affected third-degree relatives (cousins) and also second-degree relatives excluding maternal half-siblings suggests that this is a negligible contribution. Also, we observed similar RRs for children (aged 0–6 years) and older children/adolescents (aged 7–19 years) within all categories of exposed relatives, contradicting the proposed "postponement of infection" mechanism. Hence, IM risk appears to vary little by sociodemographic characteristics in an affluent and fairly equal society such as Denmark in 1977 onward. Accordingly, sociodemographic correlation between relatives seems an implausible explanation for our results. Beyond affluence, shared environmental characteristics or behavior also seem insufficient to completely explain the familial clustering of IM. The median time since parental IM in IM cases was 20 years (data not shown), and the associated RR did not vary by time since parental IM.

Because the RRs associated with affected parents, uncles, or cousins presumably are not to any substantial degree due to similarities in behavior or environment, the observed estimates may be interpreted as estimates of the genetic effects of having affected first-, second-, and third-degree relatives. Having seen the familial clustering of IM in first- and second-degree relatives, one would expect the RR in third-degree relatives of affected persons to be at least slightly elevated, and some leeway is provided in its upper confidence limit of 1.20. Hwang et al found the RR for monozygotic vs dizygotic twins to be 1.9 (95% CI, 1.2–5.3) [26], which fits the overall picture from the present study of modest genetic associations with IM risk.

At the same time, as our results suggest genetic risk factors for IM, they also indicate that these act in the context of environmental exposures. This is evidenced by the variation in RR within degrees of relatedness and within maternal/paternal half-siblings, and the declining RR by time since IM in an affected sibling. The most likely explanation for the latter trend is that the closer in age siblings are, the more similar their environments are likely to be. Accordingly, a diagnosis of IM in an older sibling is suggestive of an environment conducive of IM, and the closer in age to the affected sibling, the higher the likelihood that an exposed sibling shares this particular environment. Alternative explanations seem unconvincing. For example, the higher the birth order of a child, the smaller the likelihood of hospitalization upon IM, because more experienced parents are more relaxed about disease in their children. The larger the difference in birth order, the larger the likely age differential between the siblings. Hence, a decreasing tendency for hospital-seeking by birth order could translate into a decreasing relative risk of hospitalization for IM by time since hospitalization for IM in an older sibling. However, when we adjusted our analyses for birth order, the resulting changes in estimates were inconspicuous, leaving explanations based on sibship structure unlikely (data not shown). The relationship between small age interval and similarity of environment may also explain why having an affected sibling is associated with a larger RR than having an affected parent. It would also contribute especially to elevate the RRs for twins. This point can be illustrated by calculating $R_{ent}$, the expected RR in same-sex twins entirely due to genetics, under assumptions regarding the monozygotic (MZ) and dizygotic (DZ) twins in our study. If we assume $R_{MZ}/R_{DZ}$ = 1.9 (from [26]) and $R_{DZ}$ = 1.9 (from the observed parent–offspring association), and $R_{ent} = (R_{MZ} + R_{DZ})/2$, then $R_{ent}$ = 2.8, a small fraction of the observed RR of 9.3. The difference in RR between having affected maternal and paternal half-siblings may be also explained the same way: that maternal half-siblings presumably have much more in common than paternal half-siblings regarding environment and behavior.

The present study has several strengths and weaknesses to consider. We performed a purely register-based study, thus by design avoiding recall bias. The free access to hospitalization in Denmark [30] limits trivial sociodemographic biases in the ascertainment of outcomes and exposures in the study. Secular trends were adjusted for using very detailed sex-specific information on calendar period and birth cohort. There was no suggestion that our results were confounded by birth order characteristics. Inflation of familial RRs due to contagion by the affected relative was dealt with by discarding follow-up during the year following IM in a relative.

Many family studies have considered outcomes, and by implication, exposures that were very rare. In that case, the fact that the reference groups of no presumed exposure are polluted by a minimal fraction of unrecognized exposed has no practical statistical consequence. Not so in this study. The fact that we only ascertained hospitalized cases of IM and only did so during a limited time period, rather than from the birth of each relative, means that most IM cases in relatives of our study cohort go unnoticed by us. Figures from the Centers for Disease Control and Prevention suggest the lifetime risk of IM to be at least 3% [33]. Other estimates of IM lifetime risk can be obtained from the prevalence of having had IM in population controls in HL.
case-control studies; for example, 4.7% in the United Kingdom [34] and 4.1% in Denmark [35] (authors’ personal observation). Even with this small lifetime risk of IM, the number of relatives of various types would lead to noticeably deflated RRs in the present analysis.

We have not been able to locate a study on the validity of the diagnosis of IM in the Danish National Patient Register. Nor was it possible to validate the diagnoses based on registers. The proportion of false-negative IM diagnoses is likely to be nonnegligible, based on the experience of the diagnostic precision within many specialties in the register [36]. Importantly, however, misclassification of IM, which is both outcome and exposure here, would normally deflate association estimates and is therefore unlikely to explain our findings of familial aggregation. It is not obvious how our results will generalize to nonhospitalized IM. However, the distribution of hospitalized IM cases by sex and age seems no different from the distribution of self-reported IM in the Danish Blood Donor Study [37] (personal observation), whereas one would expect hospitalized IM patients to be older than nonhospitalized IM patients based on the dogma that later infection means more severe disease [38].

Our study supports the notion that there is a noticeable genetic component to the familial aggregation of IM—that is, that it cannot be entirely explained by contagion and sociodemographic characteristics. Even modest RRs from family studies may translate into substantial RRs for those with the least favorable genetic profile compared with those with the most favorable genetic profile [39]. Besides human leukocyte antigen (HLA) loci, which determine immune responses on chromosome 6, loci on chromosomes 1, 2, 7, 12, 19, and 22 have been associated with IM risk [22–25]. Therefore, thorough genetic association studies of IM cannot be restricted to the HLA region. At least in white individuals, the genetic associations should be stronger than suggested by our RR of having an affected parent, uncle/aunt, or first cousin. In the present context of large associations with environmental factors or gene–environment interactions, GWA studies would probably be statistically most efficient using a within-family paired design, whereas access to an existing infrastructure of genotyped controls may suggest other designs.

In conclusion, we found familial aggregation of IM, which may be explained by a mixture of genetics, shared environmental characteristics, and shared behaviors between relatives. The presumed genetic components are large enough to warrant GWA studies on disease etiology, especially to examine commonalities with causal pathways in other EBV-related diseases so as to provide clues to their prevention and treatment.

Notes

Financial support. This work was supported by the Danish Cancer Society (grant number D08155) and the Danish Council for Independent Research (grant number 09-072164).