American Journal of Epidemiology

Vol. 190, No. 5
DOI: 10.1093/aje/kwaa183
Advance Access publication:
August 24, 2020

Invited Commentary

Invited Commentary: Sibling-Comparison Designs, Are They Worth the Effort?

Thomas Frisell*

* Correspondence to Dr. Thomas Frisell, Clinical Epidemiology Division, Eugeniahemmet T2, Karolinska University Hospital, Stockholm 171 76, Sweden (e-mail: Thomas.Frisell@ki.se).

Initially submitted June 25, 2020; accepted for publication August 13, 2020.

In this issue of the *Journal*, von Ehrenstein et al. (*Am J Epidemiol*. 2021;190(5):728–737) add to the large and growing literature on the potentially causal association between prenatal exposure to maternal smoking and neuropsychiatric health. In addition to statewide, prospectively collected data, a particular strength was their ability to perform a sibling-comparison design, contrasting the rate of autism spectrum disorder in siblings discordantly exposed to maternal smoking. Unfortunately, the estimate from the sibling pairs could neither confirm nor refute the conclusions based on the full cohort. Interpretation was hampered by broad confidence limits, and even had power been higher, the authors acknowledge a range of potential biases that would have made it difficult to draw any firm conclusions from a similarity or difference in the sibling-pair estimate and estimate from the full cohort. Was the addition of the sibling comparison actually worth the effort? In this commentary, I will briefly summarize the benefits and limitations of this design, and, with some caveats, argue that its inclusion in the study by von Ehrenstein et al. was indeed a strength and not just an ornamentation.

autism spectrum disorder; causal inference; pregnancy; sibling design; smoking

Abbreviations: ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorder.

Editor's note: *The opinions expressed in this article are those of the author and do not necessarily reflect the views of the* American Journal of Epidemiology.

It has been convincingly shown over the past decades that offspring to mothers who smoked during pregnancy will on average have poorer neuropsychiatric health, including an increased rate of attention-deficit hyperactivity disorder (ADHD) (1). In this issue of the *Journal*, von Ehrenstein et al. (2) showed that this association can also be found for autism spectrum disorder (ASD), with and without intellectual disability, in a large cohort study using prospectively collected data on subjects born between 2007 and 2010 in California, with ASD diagnosis assessed until the end of 2013.

Despite the widely accepted association of maternal smoking during pregnancy and offspring neuropsychiatric traits, there is still no strong consensus on whether this association reflects a causal effect. As von Ehrenstein et al. outline in the case of ASD, it is easy to conceive of possible mechanisms for such a causal effect. On the other hand, it is even easier to conceive of potential confounders, including social determinants of health and dietary or lifestyle factors. Above all, many neuropsychiatric traits are both highly heritable and associated with a higher rate of smoking, so that mothers who smoke during pregnancy should be more likely to have a neuropsychiatric condition, which they can pass on to their offspring. These conceivable confounders are notoriously difficult to measure, and even if the association remains in adjusted regression model estimates (as in von Ehrenstein et al. (2)) one could convincingly argue that this association was due to residual confounding.

Attempting to sidestep this confounding, von Ehrenstein et al. performed a sibling-comparison design, comparing the outcome between siblings born of the same mother, where the mother smoked in some but not all pregnancies. This is an elegant design; because the mother remains the same, any association found between prenatal smoking and offspring ASD in the sibling comparison must be free from confounding by factors that are constant between pregnancies, including maternal genetics, neuropsychiatric

traits, and socioeconomic status. It is also subject to several limitations, however, and in the end, the results from this analysis neither confirmed nor refuted the results of their main analysis. Was including it actually worth the effort?

THE PROMISE OF THE SIBLING-COMPARISON DESIGN

The sibling-comparison design is intuitively appealing: If matching controls to cases by sex and age is an efficient way of controlling for confounding from the same factors, wouldn't matching a case to her noncase sister be a brilliant way of adjusting for all the confounders siblings share, whether we can measure them or not?

Methodologically, a sibling comparison can be conceived as a matched cohort study, following exposure-discordant siblings with regards to an outcome, or as a matched case-control study where exposure is compared in outcome-discordant sibling pairs. With binary exposure and outcome, these 2 descriptions coincide, illustrating a key feature of sibling-comparison designs: Only sibling pairs discordant in both exposure and outcome will contribute to the estimated "within-pair" association. Common analytical options include conditional logistic regression for binary outcomes and stratified Cox regression for time-to-event outcomes (conditioning or stratifying on the family). In the case of binary exposure and outcome, the same association could also be obtained without explicitly conditioning on family by restricting the sample to doubly discordant sibling pairs (3–5). Alternatively, any generalized linear model could be fitted as a "between-within" model, by including family means for all covariates (3). The effect estimates from these different models will in practice be very similar (acknowledging subtle differences in the target population (5, 6)) and must by design be an association of exposure on outcome, free from confounding by factors that are perfectly shared by the siblings. This has led investigators to first present the association from an ordinary cohort or case-control analysis and then contrast it with the association seen in a sibling-comparison design, assuming that any change would be due to adjusting away the confounders shared by the siblings. There are, however, many other reasons that the within-pair estimate might differ from the estimate in the source population (4).

LIMITATIONS OF THE SIBLING-COMPARISON DESIGN

Siblings tend to be similar on all things, yet a sibling comparison depends on the exclusion of sibling pairs with the same outcome/exposure. This strong selection has several consequences for the effect estimate.

First, the sibling design can amplify confounding by factors not shared by siblings. If most siblings have the same exposure, then those exposed despite having unexposed siblings are more likely to have some nonshared factor that caused the exposure. This selection thus strengthens the association of exposure and non–sibling-shared causes of exposure. If the nonshared factor also causes the outcome, this will increase confounding among discordant sibling pairs. This might be an issue with regard to prenatal smoking, where most mothers either never smoke or smoke similar amounts in all their pregnancies. If women who stop smoking between pregnancies do so due to an incident disease, or start smoking linked to a traumatic life event, this could be a significant confounder when considering pregnancies within a mother, despite being negligible in the source population. Ultimately, sibling comparisons will reduce confounding by factors more shared by siblings than the exposure is, while increasing confounding by less-shared factors (4).

Random measurement error in exposure will also attenuate the association among discordant siblings more than in the full cohort, and increasingly so for exposures increasingly shared by siblings (4). This can be intuited by imagining that true exposure causes the outcome, and is almost perfectly shared by siblings, but is measured with substantial random error. Most discordant pairs would in this scenario only appear to be discordant due to measurement error, and no association with the outcome would be seen among them.

The sibling comparison is also subject to bias if the exposure or outcome of the first sibling influences the second sibling, a so-called a cross-over effect (7). Imagine if mothers who smoked and had a child with ASD in a first pregnancy are more likely to stop smoking in a second (possibly out of newly sparked awareness of risks). The presence of such mothers among the doubly discordant pairs in the sibling comparison will lead to an inverse association between smoking and autism.

To summarize, the sibling-comparison design will ingeniously remove all confounding from factors that are identically shared by the siblings. This comes at a cost: amplified confounding from factors unique to each sibling, amplified attenuation from random measurement error, and possibly bias introduced by cross-sibling interactions. Further complications could arise from effect modification, given that discordant sibling pairs are likely to differ in the distribution of many factors compared with the source population, and results might be sensitive to modeling assumptions. All in all, there are many reasons that the association in a sibling comparison might differ from association in the source population, in addition to the removal of family-level confounders.

TRIANGULATING WITH A SIBLING-COMPARISON DESIGN

Despite these limitations, it has been argued that sibling comparisons should be included in triangulation between different designs with different limitations, complementing each other to form a consistent picture (8). And the limitations should not be overstated: Confounding from factors not shared by families is an issue in both sibling comparisons and in the source population, and it can be mitigated by adjusting for such factors. Once the increased attenuation from measurement error is recognized, it might not be a major obstacle when comparing different results. And asymmetries between siblings, suggesting cross-sibling interactions, can be assessed in the data at hand.

Although the exact estimates are difficult to compare between the full cohort and the sibling sample, it is generally informative to test whether an association seen in the full

cohort disappears in the sibling design. Such a disappearance is highly suggestive that the association was due to confounding and not a causal effect, at least if we can convincingly argue (as I believe we often could) that this is not due to increased attenuation from measurement error, cross-sibling interactions, or non–sibling-shared confounding working in the opposite direction of a causal effect. If, on the other hand, an association remains, we have not shown that it is causal, but we have shown that it is not entirely explained by confounding shared by the siblings. Sibling comparisons' truly unique contribution is that they allow us to test the hypothesis that a family-level confounder can explain all or most of an observed association. It is thus imperative that we have sufficient power to demonstrate such a null finding.

This was the weakness in von Ehrenstein et al. The sample was large, with over 2,000,000 births, and 12,000 children with ASD and 2,600 unaffected siblings to ASD cases were identified. Only 58 of these were exposed to prenatal smoking, however, and many of these were likely from uninformative pairs where the mother smoked also when pregnant with the ASD case. The resulting broad confidence limits were consistent with both increased and decreased association compared with the full cohort, and a null association is certainly possible.

Could we have foreseen this lack of power? Possibly, but it is not easy to anticipate the power of a sibling-comparison design. For binary exposures and outcomes, it will be decided by the number of doubly discordant pairs. This in turn depends on the prevalence of exposure and outcome, their association with each other and between siblings. This information is not likely to be at hand before collecting the data. Illustrating that it was conceivable that power would be sufficient in von Ehrenstein et al., a recent nationwide study in Denmark on the same topic had access to a smaller sample of births, but thanks to higher proportions with ASD and prenatal smoking, they were powered to demonstrate that the association seen in the full cohort (hazard ratio = 1.17, 95% confidence interval: 1.13, 1.22) was not present in the sibling comparison (hazard ratio = 0.86, 95% confidence interval: 0.64, 1.15) in their material (9).

Finally, even when power is less limiting, triangulation of evidence can be challenging. It has been argued that the association of prenatal smoking with birthweight most likely is causal, because an association remains in sibling comparisons and fits evidence from alternative designs (10). In contrast, a series of studies has found prenatal smoking to be associated with offspring hyperactivity or ADHD in the full cohort but not when performing a sibling-comparison study (11–13). If this rejects the causal hypothesis, then why is there such a clear dose-response pattern, and why did some (14), but not all (12), negative control studies find an association specific to maternal (vs. paternal) smoking? And what exactly would the family-level confounder be? When evidence seems conflicting, standard adjustment for potential confounders, demonstrating whether or not each could explain the association between exposure and outcome, has superior interpretability. Some studies have suggested that the association of prenatal smoking and offspring ADHD is indeed greatly attenuated by adjustment for measured maternal factors, but most have not found this to be the case (11–14). Perhaps a firm consensus will not be reached unless it is possible to identify a specific confounding factor where conditioning on it (reproducibly!) removes the association.

CONCLUDING REMARKS

Sibling comparisons do indeed add unique value, but only when the power is moderate to high, and they should not be included at the cost of proper investigation of the influence of measured confounders in the full cohort. In von Ehrenstein et al. (2), statistical power might have been too low to make any inference based on their sibling comparison, but this was difficult to predict beforehand. With the somewhat noninformative results, it might have been tempting to drop the sibling design from the paper and focus on the more robust adjusted odds ratios. We should be thankful they did not, and can only hope other authors do the same, or else publication bias will further complicate the triangulation between results from different designs.

On a final note, von Ehrenstein et al. also observed that maternal smoking during pregnancy was quite rare in California. This reduced the power of the sibling comparison but is unequivocally good news for public health. Controversy might remain on the effect of maternal prenatal smoking on ASD, but smoking before, during, or after pregnancy should clearly not be recommended.

ACKNOWLEDGMENTS

Author affiliation: Clinical Epidemiology Division, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden (Thomas Frisell).

This work was supported by the Swedish Research Council (grant DNR 2016–01355).

REFERENCES

1. Dong T, Hu W, Zhou X, et al. Prenatal exposure to maternal smoking during pregnancy and attention-deficit/hyperactivity disorder in offspring: a meta-analysis. *Reprod Toxicol*. 2018; 76:63–70.
2. von Ehrenstein OS, Cui X, Yan Q, et al. Maternal prenatal smoking and autism spectrum disorder in offspring: a California statewide cohort and sibling study. *Am J Epidemiol*. 2021;190(5):728–737.
3. Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Stat Med*. 2003;22(16):2591–2602.
4. Frisell T, Öberg S, Kuja-Halkola R, et al. Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiology*. 2012;23(5):713–720.
5. Sjölander A, Frisell T, Öberg S. Causal interpretation of between-within models for twin research. *Epidemiologic Methods*. 2012;1(1):217.
6. Petersen AH, Lange T. What is the causal interpretation of sibling comparison designs? *Epidemiology*. 2020;31(1): 75–81.

7. Sjölander A, Frisell T, Kuja-Halkola R, et al. Carryover effects in sibling comparison designs. *Epidemiology*. 2016; 27(6):852–858.
8. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45(6): 1866–1886.
9. Kalkbrenner AE, Meier SM, Madley-Dowd P, et al. Familial confounding of the association between maternal smoking in pregnancy and autism spectrum disorder in offspring. *Autism Res*. 2020;13(1):134–144.
10. Keyes KM, Davey Smith GD, Susser E. Commentary: smoking in pregnancy and offspring health: early insights into family-based and 'negative control' studies? *Int J Epidemiol*. 2014;43(5):1381–1388.
11. Skoglund C, Chen Q, D'Onofrio BM, et al. Familial confounding of the association between maternal smoking during pregnancy and ADHD in offspring. *J Child Psychol Psychiatry*. 2014;55(1):61–68.
12. Gustavson K, Ystrom E, Stoltenberg C, et al. Smoking in pregnancy and child ADHD. *Pediatrics*. 2017;139(2): e20162509.
13. Obel C, Zhu JL, Olsen J, et al. The risk of attention deficit hyperactivity disorder in children exposed to maternal smoking during pregnancy—a re-examination using a sibling design. *J Child Psychol Psychiatry*. 2016;57(4):532–537.
14. Keyes KM, Davey Smith G, Susser E. Associations of prenatal maternal smoking with offspring hyperactivity: causal or confounded? *Psychol Med*. 2014;44(4):857–867.