Invited Commentary

Invited Commentary: The Use of Sibship Studies to Detect Familial Confounding

Ezra Susser*, Martha G. Eide, and Melissa Begg

* Correspondence to Dr. Ezra Susser, Room 1030E, Allan Rosenfield Building, Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10038 (e-mail: ess8@columbia.edu).

Initially submitted February 3, 2010; accepted for publication May 13, 2010.

The authors discuss how the sibship design can be used to detect and control for familial confounding. Family-level confounding is especially problematic when estimating modest individual-level effects in the presence of familial confounders with large effects. This circumstance arises frequently in studies which relate indicators of fetal growth, such as birth weight, to outcomes that are strongly associated with parental socioeconomic status and genes. The study by Eriksen et al. in this issue of the Journal (Am J Epidemiol. 2010;172(5):530–536) uses the sibship design to capture the relation between birth weight, gestational age, and intelligence score among Norwegian males born as singletons at 37–41 completed weeks’ gestation during 1967–1984. Their study illustrates how valuable the design can be in this kind of scenario. It also illustrates the potential complexity of sibship studies and the challenges they present for appropriate interpretation.

birth weight; cohort studies; fetal development; gestational age; intelligence; Norway; siblings

Abbreviation: IQ, intelligence quotient.

For well more than half a century, epidemiologists have sought to elucidate the relation between birth weight and general intelligence or intelligence quotient (IQ). The question is important, because we need to understand the relation of fetal growth to subsequent experience over the life course. Since birth weight and IQ have both been recorded for numerous large populations, one might wonder why the question has not yet been resolved. It has proven extremely difficult, however, to disentangle a modest effect of birth weight from the larger effects of confounders such as parental socioeconomic status and genes. In addition, many investigators have focused only on the relation of low birth weight to IQ, which is also important but does not address the broader question of the relation of birth weight to IQ in the majority of the population.

The design adopted by Eriksen et al. (1) exemplifies how this question can be resolved. In order to minimize confounding, they studied male siblings, for all of whom birth weight and gestation were recorded in the Medical Birth Register of Norway and a proxy for IQ was recorded by the National Conscript Service. They focused sharply on the broader question, by limiting their study to gestational ages of 37–41 weeks and to birth weight z scores (z scores for birth weight standardized to gestational age) between the 10th and 90th percentiles. They analyzed the relation of birth weight z score to IQ within as well as between families. The extremely large sample size made it possible to achieve precision for modest effects.

Taken together with a previous study carried out in the same Norwegian cohort (2), their results allow us to draw an important conclusion. In at least some populations, and at least for males, birth weight z score is related to (a proxy for) young adult IQ. Nonetheless, their study too has flaws, which need to be understood before one can justify this conclusion.

Begg and Parides (3) have articulated the ways in which different statistical models answer different types of questions for clustered data. They used for illustration the data from a previous sibship study of birth weight and IQ (4). They showed that in this design, the individual-level effects of birth weight on IQ “within families” should be estimated after adjustment for family-level effects of birth weight on IQ “between families.” They argued that family-level effects can be captured, at least in part, by family-averaged...
birth weight. Sibship studies which overlook this point may yield inflated estimates of the relation between individual birth weight and IQ. Eriksen et al. are cognizant of this point, as reflected in their differentiation of “within-family” and “between-family” effects of birth weight z score (1).

Their approach, however, also introduces complications to the analysis which could render their findings noncomparable with those of other studies. The authors present models for IQ as a function of birth weight z score, appropriately using regression analysis via the generalized estimating equations technique, specifying the identity link and exchangeable correlation. The models they present all include birth weight z score as a main effect/predictor variable, as well as the square of birth weight z score. This conforms with the quadratic shape of the relation between birth weight and IQ observed in many studies, in which there is a steeper, increasing slope at the lower end of the birth weight scale which gradually becomes less steep, ultimately reaching a plateau and starting to decline at the high end of the birth weight scale. The use of the quadratic term, however, alters the appropriate interpretation of their statistical model, as does the inclusion of interaction terms of birth weight with birth order and age difference. Indeed, we will show that these additions require a more careful interpretation of analytic results and pose a barrier for comparisons with other studies.

Consider a linear regression model (generalized estimating equations or other type) that attempts to relate predictor variable x to outcome variable y. By a “linear” relation, we mean to say that for every 1-unit change in x, the expected change in y is the same, anywhere along the scale of the predictor. When a squared term (x²) is added as a predictor to a regression model, it begins to capture any nonlinear component of the relation between the predictor variable and the outcome (as is commonly observed between birth weight and IQ). A direct consequence of adding the squared term is that the expected change in y changes depending on one’s location along the distribution of x values. That is, the expected difference in y for a 1-unit change in x will vary, based on where one is along the x scale.

While Eriksen et al. recognize the impact of the addition of the squared term on the shape of the curve (1), they do not draw out the implications for “effect size” very explicitly for the reader. This might be best done by creating a series of estimated effects at various points along the birth weight z score scale. We can demonstrate this approach using the results provided in the authors’ Table 3 (1). Suppose that one fitted a model for intelligence score as a function of birth weight z score, the square of birth weight z score, and the mean birth weight z score for the family, plus other variables. Then the expected difference in intelligence scores between a subject whose birth weight z score was equal to x₁ and another subject whose birth weight z score was x₂ would be computed as

\[
\beta_1(x_1 - x_2) + \beta_2(x_1^2 - x_2^2),
\]

where \(\beta_1\) is the estimated regression coefficient for birth weight z score and \(\beta_2\) is the estimated regression coefficient for the square of birth weight z score. Given the estimated coefficients from a particular model, then, we can calculate the expected change in intelligence score for a 1-unit difference in birth weight z score at various points along the birth weight z score scale (see Table 1).

In this model, the coefficient for the simple birth weight z score term, reported in Eriksen et al.’s Table 3 (e.g., 0.12 from one model), is not on its own an accurate depiction of the relation between birth weight and intelligence score. It is important to emphasize that the relation will be different at each point along the birth weight scale due to the inclusion of the squared term. In addition, it is not meaningful to directly compare it with estimates obtained from other results in the literature which did not include a squared term. The introduction of the squared birth weight z score term renders the association more complicated, and the presentation and interpretation of results should reflect this level of complexity.

Another complication results from the inclusion of interaction terms. In some models, Eriksen et al. included interactions between birth weight z score and birth order and between birth weight z score and the (mean) age difference between siblings (1). Consider for a moment just the addition of the interaction between birth weight z score and birth order. Combined with the simple and squared terms for birth weight, this implies that the expected difference in intelligence score for 2 subjects with birth weight z scores of x₁ and x₂, respectively, and a given birth order value (z) and corresponding regression coefficient estimate of \(\beta_3\), must be calculated as

\[
\beta_1(x_1 - x_2) + \beta_2(x_1^2 - x_2^2) + \beta_3(z(x_1 - x_2)) = (\beta_1 + \beta_3z)(x_1 - x_2) + \beta_2(x_1^2 - x_2^2),
\]

such that the expected change in intelligence score varies according to where one is along the scale of birth weight z scores, as well as according to the assumed value of the birth order term. (Note that the statement above assumes that we are calculating expected change in intelligence score for 2 persons with the same birth order; for different birth orders, the expression would be even more complex.) Again, the presentation and interpretation of results from the study should reflect this. In addition, the main effect coefficient for birth weight z score is not directly comparable to findings from other studies that did not include squared terms or interaction terms (without including additional details about the expected change in y at different points along the x scale, as we have presented in Table 1).

| Table 1. Illustration of How Effect Size Changes With Birth Weight in a Nonlinear Model |
|---------------------------------|---------|---------|-----------------|-----------------|-----------------|
| x₁    | x₂    | x₁ – x₂ | x₁² – x₂² | 0.12(x₁ – x₂) – 0.02(x₁² – x₂²) |
| -3    | -4    | 1       | -7          | 0.26           |
| 0     | -1    | 1       | -1          | 0.14           |
| 0.5   | -0.5  | 1       | 0           | 0.12           |
| 1     | 0     | 1       | 1           | 0.10           |
| 3     | 2     | 1       | 5           | 0.02           |
Nonetheless, Eriksen et al.’s paper (1) does make an important contribution to our understanding of the relation between birth weight and IQ. An interesting aspect of their analysis is the nonsignificant contribution of the “family-level” effect, as captured by the coefficient of the mean birth weight across siblings. Why is there no apparent effect of the surrogate for family-level factors? In their Table 3, Eriksen et al. show that after adjustment for parental education and birth order, there are no detectable “between-family” effects (1). In other words, in this population, adjusting for parental education and birth order would appear to be sufficient to remove almost all of the family-level confounding from the association being investigated.

This result enables us to reinterpret the findings from a previous study of birth weight and IQ in Norwegian conscripts (2). The previous study was conducted in essentially the same population but did not identify sibships. Eide et al. (2) reported a modest association between birth weight and IQ after adjustment for maternal education and birth order, among other factors. They could not, however, rule out confounding by unmeasured familial factors. The results of Eriksen et al. (1) provide a necessary complement to the prior study, by showing that unmeasured familial factors do not make a detectable contribution to the association between birth weight and IQ in this population. As Eriksen et al. noted (1), the prior study was not restricted to gestational ages of 37–44 weeks (Eide et al. used 37–44 weeks) and was not restricted to birth weight z-scores within the 10th–90th percentiles. A reanalysis of Eide et al.’s data after imposing these same restrictions, however, did not appreciably change the coefficient for the birth weight z-score (results available from M. G. E.). Thus, the totality of evidence suggests that in this population, birth weight z-score is indeed related to IQ in male conscripts after exclusion of preterm and postterm babies as well as low and high birth weight babies.

Does this result mean that a sibship study is unnecessary, after all, for tight control of confounding? We do not think so. The additional control derived from using sibships will vary across populations, depending (among other things) on the nature of social inequalities and the association being investigated. In this instance it proved superfluous, in that the sibship study (1) validated the result of the previous study (2). However, one could not have known that without conducting the sibship study.

The hazard of overlooking unmeasured familial confounding is neatly illustrated in a recent study from Sweden (5). Lundberg et al. (5) examined the relation of prenatal smoking to low intellectual performance in male offspring, by linking the Swedish Medical Birth Register to the Swedish Conscript Register. In an analysis of unrelated persons, adjustment for parental education and other factors reduced the association between prenatal smoking and poor intellectual performance in male offspring, but the association remained evident and statistically significant. In an analysis of male sibships, however, Lundberg et al. demonstrated that prenatal smoking was not related to poor intellectual performance. The authors inferred that unmeasured familial factors accounted for the association they observed among unrelated persons.

Studies of sibships offer a powerful and underutilized method for enhancing causal inference from observational studies. As a rule of thumb, investigators should seriously consider the advantages of sibships in the design of their studies and should include a sibship analysis when the data allow for it (the most important requirement is a sufficient number of siblings who differ on exposure status). We also propose that the logic of the sibship design and the appropriate analytic strategies be incorporated into the teaching of core epidemiologic methods.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York (Ezra Susser); Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, New York (Ezra Susser); New York State Psychiatric Institute, New York, New York (Ezra Susser); Division of Mental Health, Norwegian Institute of Public Health, Bergen, Norway (Martha G. Eide); Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway (Martha G. Eide); and Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York (Melissa Begg).

Conflict of interest: none declared.

REFERENCES