

Sibling Comparison Designs

Bias From Non-Shared Confounders and Measurement Error

Thomas Frisell,^{a,b} Sara Öberg,^c Ralf Kuja-Halkola,^{a,b} and Arvid Sjölander^a

Abstract: Twins, full siblings, and half-siblings are increasingly used as comparison groups in matched cohort and matched case-control studies. The “within-pair” estimates acquired through these comparisons are free from confounding from all factors that are shared by the siblings. This has made sibling comparisons popular in studying associations thought likely to suffer confounding from socioeconomic or genetic factors. Despite the wide application of these designs in epidemiology, they have received little scrutiny from a statistical or methodological standpoint. In this paper we show, analytically and through a series of simulations, that the standard interpretation of the models is subject to several limitations that are rarely acknowledged.

Although within-pair estimates will not be confounded by factors shared by the siblings, such estimates are more severely biased by non-shared confounders than the unpaired estimate. If siblings are less similar with regard to confounders than to the exposure under study, the within-pair estimate will always be more biased than the ordinary unpaired estimate. Attenuation of associations due to random measurement error in exposure will also be higher in the within-pair estimate, leading within-pair associations to be weaker than corresponding unpaired associations, even in the absence of confounding. Implications for the interpretation of sibling comparison results are discussed.

(*Epidemiology* 2012;23: 713–720)

Submitted 12 June 2011; accepted 12 March 2012; posted 9 July 2012.

From the ^aDepartment of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; ^bCentre for Violence Prevention, Karolinska Institutet, Stockholm, Sweden; and ^cDepartment of Epidemiology, Harvard School of Public Health, Boston, MA.

Conflicts of Interest and Source of Funding: T.F. and R.K.-H. are supported by the Swedish Prison and Probation Service R&D. SÖ is supported by the Swedish Council for Working Life and Social Research [2011–1477]. AS is supported by the Swedish Research Council [2008–5375]. The authors report no conflict of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy edited; it is the sole responsibility of the author.

Correspondence: Thomas Frisell, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, PO Box 281, 171 77, Stockholm, Sweden. E-mail: Thomas.Frisell@ki.se.

Copyright © 2012 by Lippincott Williams & Wilkins

ISSN: 1044-3983/12/2305-0713

DOI: 10.1097/EDE.0b013e31825fa230

Sibling comparison designs, such as the co-twin control design, have been considered simple, yet ingenious, extensions of the matched case-control design. Instead of explicitly matching on a set of measured variables, the use of siblings as controls will automatically match on many unmeasured factors, including cultural background, parental characteristics and child-rearing practices, and—particularly for monozygotic twins—genetics. This promise of adjustment for a wide range of unmeasured, even unknown, confounders has made co-twin control methods popular, especially in etiologic research on controversial topics. Recent applications include the associations of body mass index with mortality,¹ sexual orientation with mental disorder,² and birth weight with cardiovascular disease.³ Twins are relatively rare, and the degree to which associations among twins are generalizable to the entire population has been questioned. Recent commentaries have therefore advocated matching of other relatives, primarily full- and half-siblings.^{4,5} Indeed, sibling comparisons have already been used in economics research for at least 30 years,⁶ and recent applications in epidemiology include the association of mother’s smoking during pregnancy with offspring psychologic traits,^{7–9} and the association of birth weight with intelligence.¹⁰

Methodologic aspects of sibling-comparison designs have been relatively unexplored. Common analytic strategies for paired data include the flexible between-within model^{11–13} and conditional logistic regression. In the absence of nonshared confounding and other biases, these methods produce estimates of the causal effect of exposure free from confounding by factors shared by the siblings.¹³ The absence of bias is rather unlikely in real applications, but what impact these biases might have on sibling comparisons has, to our knowledge, not been discussed in the epidemiologic literature. In contrast, economists have long been aware that random measurement error and omitted variables can have specific influence on within-pair estimates under a linear model.^{6,14}

We explore the impact of 2 common sources of bias: confounding by factors not perfectly shared by siblings, and random measurement error of exposure. We will show, analytically and with simulations, how sibling comparison methods in these situations may actually lead to increased bias. Specifically, we give a brief summary of how sibling com-

parisons are analyzed and explain why the designs may lead to increased bias. Next, we show the impact analytically under a linear model, and we illustrate the impact through simulations under a logistic model where the outcome and exposure are dichotomous. We end with a discussion of how results from sibling comparisons may be interpreted in light of the caveats raised here.

STATISTICAL ANALYSES OF SIBLING COMPARISON DESIGNS

For binary outcomes, the sibling comparison is often considered an extension of the matched case-control design. Sibling-pairs discordant for the studied outcome are selected, and the association of exposure with outcome in the paired data is analyzed using McNemar's test or conditional logistic regression. Because the analysis is essentially stratified by sibling pair, only pairs that also differ in exposure level will contribute to the estimated association. Although conditional logistic regression is the method of choice in cross-sectional case-control settings, most sibling comparison studies are nested in a larger cohort or population register and may thus be analyzed using the more general between-within model.

To describe this model, let Y_{ij} and X_{ij} be the outcome and exposure, respectively, of individual j in pair i . The pairs are symmetric, so the ordering of j is exchangeable. First consider an ordinary generalized linear regression model, ignoring the fact that the data consist of pairs:

$$g\{E(Y_{ij}|X_{ij})\} = \alpha + \beta X_{ij} \quad (1)$$

The between-within model is a simple extension of this, where the individual's exposure and the siblings' mean exposure are included as independent variables:

$$g\{E(Y_{ij}|X_{ij}, \bar{X}_i)\} = \alpha_{BW} + \beta_W X_{ij} + \beta_B \bar{X}_i \quad (2)$$

where \bar{X}_i is the average exposure in pair i . The link function, $g\{\}$, puts the sibling comparison in a generalized linear model framework. Different link functions enable for instance linear, logistic, and log-linear regression. The exposure-outcome association is split into 2 parts: a within-effect β_W and a between-effect β_B . When the exposure is dichotomous, the between-within model will give the same result as analyzing only exposure-discordant twin pairs.¹⁵

Recently, we showed that in the absence of nonshared confounders, the within effect may be interpreted as a causal exposure effect in the subpopulation consisting of exposure discordant pairs.¹³ The between-effect contains information on how strongly factors shared by the family affect the exposure-outcome association. However, this information is generally considered difficult to interpret and is ignored in most applications.

Although some sibling comparison studies estimate only within effects, most aim to evaluate the presence and

degree of confounding due to shared factors by comparing the unpaired estimate, β , to β_W . The customary interpretation has been that if $\beta = \beta_W$ there is no confounding by shared factors; if β is different from β_W , there is at least some confounding by shared factors; and if $\beta \neq \beta_W = 0$, the entire association is caused by shared confounders. As we will show in the following sections, none of these conclusions is necessarily true.

HOW SIBLING COMPARISON DESIGNS CAN INCREASE BIAS

To understand how sibling comparison designs could increase bias, it may help to consider a hypothetical situation in which the exposure, confounder, and outcome are binary, there is no causal effect of exposure on outcome, and the exposure and outcome are both influenced by sibling-shared risk factors, but the confounder is not. As noted earlier in the text, the only pairs that contribute to the estimated "within-pair" association are those discordant in exposure. Although such discordant pairs may share many causes of the exposure, the fact these pairs are differently exposed implies a selection of pairs that also differ in nonshared causes of the exposure, including common causes of both exposure and outcome.

In other words, with regard to nonshared confounders, the members of exposure-discordant pairs (although they are siblings) are likely to differ more from each other than 2 randomly selected persons from the same population having the same exposure levels. If the effect of the confounder is to increase both the probability of being exposed and the probability of developing the outcome, unexposed persons are less likely to have the confounder, and consequently less likely to develop the outcome. When picked as part of a discordant pair, a sibling is also selected to be different from the exposed relative, further reducing their likelihood of having the confounder. As an effect, the confounder-exposure association will be strengthened, thus increasing any spurious association due to nonshared confounding bias.

This problem can be illustrated using causal diagrams.¹⁶ Suppose that the data-generating mechanism is given by the causal diagram in the Figure. X is the exposure of interest, Y is the outcome, and C denotes the full set of unmeasured confounders, both shared and nonshared. C_{ij} represents C for the j th individual in the i th sibling pair. F_C and F_X denote the familial factors influencing C and X , respectively. The hypothetical situation discussed in the preceding paragraph is a special case of this scenario, where F_C is completely absent. Suppose that all variables in the Figure are binary, and that all causal influences are "positive," so that $F_C = 1$ increases the likelihood for $C_{ij} = 1$, which in turn increases the likelihood for $X_{ij} = 1$, etc. A crude (ie, non-paired) analysis compares the exposed ($X_{ij} = 1$) with the unexposed ($X_{ij} = 0$), in terms of the outcome Y_{ij} . Due to the path $C_{ij} \rightarrow X_{ij}$, those with $X_{ij} = 1$ are more likely to have

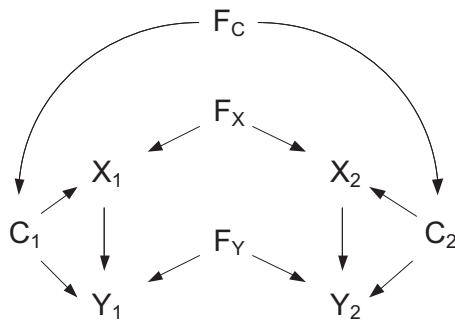


FIGURE. Causal diagram of exposure-outcome association in sibling pairs. X, C, and Y are the exposure, confounder, and outcome, respectively, for individuals 1 and 2. Familial factors (F_X , F_C , and F_Y) cause X, C, and Y to be clustered in families.

$C_{ij} = 1$ than those with $X_{ij} = 0$, which induces an imbalance in C_{ij} across these groups. A sibling analysis makes a similar comparison, but restricted to the pairs with $X_{i1} \neq X_{i2}$. In other words, a comparison is made between those with ($X_{ij} = 1$, $X_{ij'} = 0$) and those with ($X_{ij} = 0$, $X_{ij'} = 1$), in terms of the outcome Y_{ij} . What does the extra information about the sibling's exposure, $X_{ij'}$, tell us about the confounder C_{ij} for the person under study? Due to the path $C_{ij} \leftarrow F_C \rightarrow C_{ij'} \rightarrow X_{ij'}$, we have that $X_{ij'} = 0$ decreases the likelihood for $C_{ij} = 1$ and that $X_{ij'} = 1$ increases the likelihood for $C_{ij} = 1$, that is, the path induces a positive correlation between C_{ij} and $X_{ij'}$. Thus, ignoring other paths, a restriction to pairs with $X_{i1} \neq X_{i2}$ counteracts the imbalance in C_{ij} across the comparison groups with $X_{ij} = 1$ and $X_{ij} = 0$. In other words, because the confounder is influenced by familial factors, confounding will be decreased by this restriction. However, we must also take the path $C_{ij} \rightarrow X_{ij} \leftarrow F_X \rightarrow X_{ij'}$ into account. By conditioning on the collider X_{ij} , we induce a spurious (ie, noncausal) association between C_{ij} and $X_{ij'}$.¹⁷ Suppose now that $X_{ij'} = 0$. It is then likely that $F_X = 0$, which makes it more likely that $X_{ij} = 0$. If, despite this, $X_{ij} = 1$ (ie, the pair is discordant), then this is likely due to $C_{ij} = 1$, and so this path induces a negative correlation between C_{ij} and $X_{ij'}$. Thus, ignoring other paths, a restriction to pairs with $X_{i1} \neq X_{i2}$ induces an even larger imbalance in C_{ij} across the comparison groups with $X_{ij} = 1$ and $X_{ij} = 0$. In other words, because the exposure is influenced by familial factors, confounding will be increased by this restriction. In reality, both these paths are likely to be present, and the magnitude of the resulting bias depends on the relative strength of the arrows along both paths, that is, on whether the exposure or the confounder is most strongly influenced by familial factors. As shown below, this makes the sibling correlation in exposure and confounders important for the interpretation of sibling comparisons.

If there is random measurement error (misclassification) in the observed exposure, then the association of the observed exposure and the outcome will be weaker than the

association of the true exposure and the outcome. Among the discordant pairs, some are in truth concordant but appear to be discordant due to misclassification. As argued in the beginning of this section, analyzing discordant pairs implies a selection for pairs that differ on nonshared causes of the exposure. Because random measurement error is not shared by siblings, we are selecting for pairs that differ in the direction of such error. In other words, among individuals in discordant pairs, fewer will be correctly classified on exposure than in the general population. This will lead to an increased attenuation of the observed exposure-outcome association.

The reasoning above is valid because only sibling pairs that differ in exposure will be informative on the within-pair exposure-outcome association. It applies to case-control and cohort settings, and to continuous, binary, or count measures of exposure or outcome, and is not dependent on the statistical analysis procedures used. Although sibling comparison methods have been used under a variety of names (eg, "sibling discordance study,"⁴ "sibship design,"⁵ "discordant-twin design,"¹⁸ "cotwin control,"^{19,20} "cotwin-matched case-control,"²¹), this reasoning applies to all.

IMPACT UNDER A LINEAR BETWEEN-WITHIN MODEL

If the outcome is continuous and the true causal effects are linear, it is possible to derive exact mathematical expressions for the crude and within-pair regression coefficients, showing explicitly how sibling comparison models may increase bias in parameter estimates. We show only the results in this section; for full derivations, see the Appendix.

Consider the case in which Y and X are continuous and centered (so we may ignore intercepts), and all causal effects in the Figure are linear. The true causal model can then be written as:

$$Y_{ij} = \beta_{YX}X_{ij} + \beta_{YC}C_{ij} + \epsilon_{Yij} \quad (3)$$

$$X_{ij} = \beta_{XC}C_{ij} + \epsilon_{Xij} \quad (4)$$

We define $\sigma_C^2 = Var(C)$, $\sigma_{\epsilon_Y}^2 = Var(\epsilon_Y)$, and $\sigma_{\epsilon_X}^2 = Var(\epsilon_X)$, and note that (due to F_C , F_X , and F_Y) C, X, and Y may all be correlated within pairs: $Cov(C_{i1}, C_{i2}) = \rho_C \sigma_C^2$, $Cov(\epsilon_{Yi1}, \epsilon_{Yi2}) = \rho_{\epsilon_Y} \sigma_{\epsilon_Y}^2$, and $Cov(\epsilon_{Xi1}, \epsilon_{Xi2}) = \rho_{\epsilon_X} \sigma_{\epsilon_X}^2$. According to the Figure, all confounding between X and Y is captured by C, and C is associated with X and Y only through its causal influence. Thus, person ij 's deviation terms (ϵ s) are independent of X, Y, C and each other, and there are no within-family cross-variable correlations ($Cov(\epsilon_{Yi1}, \epsilon_{Xi2}) = 0$, etc).

In this situation, the parameter of interest—the causal effect of one unit's increase in X on Y—is β_{YX} . A linear regression of X and C on Y would produce an unbiased estimate, although correct inference requires robust standard errors to account for the correlated structure of the data.

The regression of X on Y would suffer from confounding bias.

Impact of Confounding

The ordinary, or crude, estimate is simply the coefficient from a linear regression of X on Y. Expressed in terms defined above, the crude estimate can be shown (see Appendix) to be:

$$\beta = \beta_{YX} + \frac{\beta_{YC}\beta_{XC}\sigma_C^2}{\beta_{XC}^2\sigma_C^2 + \sigma_{\epsilon X}^2} \tag{5}$$

We see that β is a sum of the causal effect, β_{YX} , and a term due to confounding bias. The degree of confounding depends on the strength of C’s effect on Y and X but does not depend on the strength of the familial clustering of either Y, X, or C. The expression for the within-effect is similar, but not identical:

$$\beta_w = \beta_{YX} + \frac{\beta_{YC}\beta_{XC}\sigma_C^2}{\beta_{XC}^2\sigma_C^2 + \sigma_{\epsilon X}^2} \frac{1 - \rho_{\epsilon X}}{1 - \rho_C} \tag{6}$$

The only difference from β is the $\frac{1 - \rho_{\epsilon X}}{1 - \rho_C}$ in the denominator of the bias term. If all confounders are perfectly shared by siblings, $\rho_C = 1$, and the denominator of the confounding term will be infinite, the confounding term will be zero and β_w equal the causal effect. However, if $\rho_C = \rho_{\epsilon X}$ then $\beta = \beta_w$, even though there may be confounders shared by siblings. We also see that if $\rho_C > \rho_{\epsilon X}$, the confounding bias in β_w will be less than in β , and if $\rho_C < \rho_{\epsilon X}$, then β_w is more biased than β . Because $Cor(X_{i1}; X_{i2})$ is a linear combination of ρ_C and $\rho_{\epsilon X}$ (weighted by the proportion of variance in X due to C and to other causes, respectively), it follows that these conclusions also hold when substituting $\rho_{\epsilon X}$ with $Cor(X_{i1}; X_{i2})$. Thus, we have shown that under the linear model, the within estimate will be less biased than the crude estimate only when the set of all confounders is more strongly shared by siblings than the exposure. Below we will show that the same conclusion holds under the logistic model.

Impact of Measurement Error

In general, measurements of X are likely to be somewhat inaccurate. In the presence of random measurement error, the association between observed X (X^*) and Y is lower than the association of X and Y, and this attenuation will be increased in the within-pair association. Let the X^* equal true X plus an error term

$$X_{ij}^* = X_{ij} + \epsilon_{Mij} \tag{7}$$

Then, the attenuated association can be shown (see Appendix) to be

$$\beta^* = \frac{Var(X_{ij})}{Var(X_{ij}^*)} \beta = \gamma \beta \tag{8}$$

where γ is the reliability of our measure of X. Thus, a reliability of 90% will attenuate the regression coefficient by 10%. For the within estimate, the solution is

$$\beta_w^* = \gamma \left(\beta_{YX} \frac{\beta_{XC}^2\sigma_C^2(1 - \rho_C) + \sigma_{\epsilon X}^2(1 - \rho_{\epsilon X})}{\beta_{XC}^2\sigma_C^2(1 - \gamma\rho_C) + \sigma_{\epsilon X}^2(1 - \gamma\rho_{\epsilon X})} + \frac{\beta_{YC}\beta_{XC}\sigma_C^2(1 - \rho_C)}{\beta_{XC}^2\sigma_C^2(1 - \gamma\rho_C) + \sigma_{\epsilon X}^2(1 - \gamma\rho_{\epsilon X})} \right) \tag{9}$$

Like β , the within-estimate is attenuated by a multiple of the reliability. However, β_w is further attenuated by the γ in the denominators. Consider a situation where the within-estimate is not confounded (eg, when $\beta_{XC} = 0$).

$$\beta_w^*(\beta_{XC} = 0) = \beta_{YX} \left(1 - \frac{1 - \gamma}{1 - Cor(X_{i1}^*, X_{i2}^*)} \right) \tag{10}$$

Although β is attenuated by $1 - \gamma$, β_w is attenuated by $(1 - \gamma)/(1 - Cor(X_{i1}^*, X_{i2}^*))$. This factor has been reported previously,^{6,18} but we point out that it holds only under no confounding. Otherwise, the attenuation of the within estimate will be even stronger, as described by Eq. 9.

IMPACT UNDER A LOGISTIC BETWEEN-WITHIN MODEL

If the causal effects follow a logistic model, a model commonly used for dichotomous outcomes, we do not have a closed-form solution of the regression coefficients. To show that the same qualitative conclusions still hold, we simulated paired data following the causal structure in the Figure, under a logistic model with binary exposure, confounder, and outcome. For details on the simulation set-up, please see the eAppendix (<http://links.lww.com/EDE/A596>). The influence of the within-pair correlation of confounders and measurement error in the exposure is shown in the Table. Note that under the logistic model, due to the noncollapsibility of the odds ratio, the “within effect” is a causal exposure effect in the subpopulation consisting of exposure discordant pairs, rather than a pair-specific effect.¹³ This means that in the Table, the $OR_w = 5$ in rows where $Cor(C1, C2) = 1$ (eg, rows 1 and 5) is a causal, nonconfounded effect of X on Y, but due to noncollapsibility, it would not generally equal the causal effect in the whole population. In the absence of confounding, though, the within effect will equal the causal effect in the whole population.¹⁵

TABLE. Impact of Confounding and Measurement Error on Crude and Within-Pair Odds Ratios

Confounding			
Cor(X1, X2)	Cor(C1, C2)	OR	OR_w
0.6	1	8.0	5.0
0.6	0.6	8.0	8.0
0.6	0.3	8.0	12.2
0.6	0	8.0	19.9
0.3	1	8.0	5.0
0.3	0.6	8.0	6.5
0.3	0.3	8.0	8.0
0.3	0	8.0	10.1
Inverse Confounding			
Cor(X1, X2)	Cor(C1, C2)		
0.6	1	2.0	5.0
0.6	0.6	2.0	2.0
0.6	0.3	2.0	1.0
0.6	0	2.0	0.5
0.3	1	2.0	5.0
0.3	0.6	2.0	2.8
0.3	0.3	2.0	2.0
0.3	0	2.0	1.4
Measurement Error in Exposure			
Sensitivity	Specificity		
1	1	5.0	5.0
1	0.8	4.4	3.5
1	0.6	3.9	3.0
0.8	1	3.0	2.6
0.8	0.8	2.3	1.9
0.8	0.6	1.8	1.5
0.6	1	2.3	1.9
0.6	0.8	1.7	1.5

Results from between-within model on simulations under a logistic model. Non-confounded OR = 5.0.

Simulations follow the Figure. In all simulations, the prevalence of X, C, and Y was set to 20%, 50%, and 10%, and non-confounded OR to 5.0. Confounders create positive association between X and Y, inverse confounders create inverse association. In measurement error simulations, the Pearson correlation of true X1-X2 was set to 0.3.

OR indicates crude odds ratio; OR_w, within sibling-pair odds ratio.

With this simulation, we can also illustrate a point made above. Let the prevalence of exposure, confounder, and outcome be 20%, 50%, and 10%; the Pearson correlation of X1-X2, 0.3; the correlation of C1-C2, 0.0; and the nonconfounded odds ratio (OR) of X on Y be 5, the OR of C on X be 3, and the OR of C on Y be 26. The prevalence of the confounder among unexposed individuals in the population is then 45%. But the prevalence of the confounder among unexposed individuals who are part of an exposure discordant pair is 37%. Further, let the exposure be measured with a sensitivity and specificity of 0.8. In the entire population, 80% will be correctly classified on exposure, but among individuals in dis-

cordant pairs, only 67%. For readers interested in trying out other parameter values, we provide R-code used in our simulations in eAppendix 1 (<http://links.lww.com/EDE/A596>).

Impact of Confounding

The exact magnitude of confounding is a function of the prevalence of exposure, confounder, and outcome, the strength of the causal effect of confounder on exposure and outcome and the causal effect of exposure on outcome. In the Table simulations, these values were set to produce a confounded crude OR of 8 (confounding) and 2 (inverse confounding). The extent that sibling comparisons remove or increase confounding bias is decided by the relative correlation of exposure and confounder within families. From the Table, we see that the same qualitative conclusion holds for within-pair OR as for β_w in the linear setting. If all confounders are perfectly shared by siblings, that is, $Cor(C1, C2) = 1$, then OR_w equals a causal OR. However, if $Cor(C1, C2) = Cor(X1, X2)$, then $OR = OR_w$, although there are confounders shared by siblings. Further, if $Cor(C1, C2) > Cor(X1, X2)$, the bias in OR_w will be less than the bias in OR; if $Cor(C1, C2) < Cor(X1, X2)$, then OR_w is more biased than OR.

Impact of Measurement Error

Measurement error, or misclassification, of a dichotomous variable is often described in terms of the measurement's sensitivity and specificity.²² Whether low specificity or low sensitivity is most influential in attenuating the exposure-outcome association will depend on the prevalence of the variables. Regardless, as illustrated in the Table, both will lead to an attenuation of the OR, and further attenuation of the OR_w .

DISCUSSION

In line with previous econometric research,^{6,7} we have shown that within-pair estimates from sibling comparisons, β_w , can be more biased than crude unpaired estimates, β . In particular, we have shown that the usual interpretation of results from sibling comparison studies will be valid only in the absence of confounding not perfectly shared by siblings and measurement error in the exposure. Since many real associations are likely to suffer from one or both of these biases, this should be considered an important caveat for sibling comparison applications and may call for re-evaluation of previous work using these designs.

Implications for Interpretation

To summarize, β_w is unbiased only if all confounders are perfectly shared by the members of the pair, and there is no random measurement error of the exposure. If there is measurement error, β_w is expected to be lower (closer to the null) than β , even without confounding. If there is confound-

ing, β_w may be either less or more confounded than β . When siblings are less similar in exposure than in confounders, confounding bias will be lower in β_w . When siblings are more similar in exposure than in confounders, confounding bias will be higher in β_w . When siblings are similarly correlated in exposure and confounders, $\beta_w \approx \beta$, although confounders are to some extent shared by siblings. This makes interpretation of sibling comparisons more complicated than usually assumed, particularly for co-twin control designs. Comparisons of within-pair estimates from dizygotic (DZ) and monozygotic (MZ) twins are often used to assess whether confounding comes from family environment or genetic sources. Because MZ twins have a stronger correlation of both the exposure and confounders whenever these are heritable, this comparison will be sensitive to the influence of nonshared confounding and measurement error. For instance, if there is a positive causal effect of the exposure and measurement error in the exposure, but no confounding, we would find $\beta_{w(MZ)} < \beta_{w(DZ)} < \beta$, which would typically (and in this scenario erroneously) be interpreted as evidence for genetic confounding.^{23,24}

Sibling comparisons cannot in themselves prove causality or establish whether there is sibling-shared confounding. However, adding previous subject matter knowledge or complementary analyses, we may still be able to use sibling comparisons to support such reasoning. If we perform an adequately powered sibling comparison and observe β stronger than β_w , this is compatible with several causal scenarios. First, the association may be partly or wholly caused by confounders that are more shared by siblings than the exposure is. Second, there may be random measurement error in the exposure. Third, there may be a combination of these 2. Fourth, the association may be partly or wholly caused by confounders less shared by siblings than the exposure is. Although this would have made β_w stronger than β , the increased attenuation due to measurement error counteracts this. Fifth, there may be confounding causing an inverse (with respect to β) association, masking part of the causal effect, less shared by siblings than the exposure is, with or without measurement error. Sixth, there may be confounding causing an inverse association, more shared by siblings than the exposure is; this would have made β_w stronger than β , but the increased attenuation due to measurement error counteracts this.

In given settings, one may be able to convincingly argue that overall confounding is unlikely to cause an inverse association, disqualifying the fifth and sixth explanations. Further, if one could show that the measurement error is not strong enough to explain the lowered β_w under the scenario with no confounding, then only the first and the third explanations are possible. The association must be, at least partly, caused by confounders shared by siblings. Further, unless $\beta_w = 0$, it is not possible that the association is

completely caused by confounders that are perfectly shared by siblings.

The relative strength of sibling correlation in exposure and confounder turns out to be a central concept in these designs. Thinking in terms of the degree of correlation among variables may be more useful than the usual distinction between shared and nonshared factors. If shared factors are those that have a sibling correlation of one, and the nonshared factors have a correlation of zero, it would seem that most variables of interest are somewhere in between.

As shown earlier in the text, the sibling-correlation in exposure leads to increased bias due to collider stratification. Sibling-correlation in the confounders is what allows a reduction in bias. These correlations set sibling comparisons apart from ordinary matched designs in which the index person and reference are selected to be perfectly correlated on the matching variables, but are not correlated in other causes of exposure or confounders. Although the correlation in exposure can be estimated, the correlation in the set of all confounders can only be hypothesized, and requires a discussion of what they may be, their relative strength, and how heritable or influenced by shared environment they are. If there are 2 confounders—one weak and highly heritable or influenced by shared environment, and the other strong but only very weakly familial—then the set of confounders will have an intermediate correlation.

As a rule of thumb, sibling comparison may be used in situations where we believe that the confounding is more shared than the exposure, and avoided in the opposite scenario. An example of the first situation may be twin studies of prenatal exposures and adult outcomes. Although these exposures may be strongly correlated in twins (eg, birth weight), most hypothetical confounders will be perfectly shared by the twins (eg, gestational length, maternal age, parity, maternal substance use). A situation that warrants more caution may be studies of adult lifestyle factors and disease (eg, body mass index and mortality). In these, it is easy to think of potential confounders that, though heritable, may be less correlated than the exposure (eg, personality factors, diet, exercise, health-seeking behavior). Sibling comparisons of, for example, maternal smoking during pregnancy may also warrant some caution. Although mothers share many characteristics from one birth to the other, several will also differ by default (parity, maternal age, and paternal age). It would then be informative to estimate the sibling correlation in maternal smoking during pregnancy and contrast it to the sibling correlation in some of the putative confounders.

In general, it may be difficult to make assumptions of the extent to which unmeasured confounders are shared. However, we may feel more comfortable arguing whether they should, on the whole, cause a positive or negative association. As for measurement error, we may be able to

directly measure the reliability of our instruments, or at least put some upper or lower limit on how well our measure captures the true exposure. Together, assumptions like these may aid us in deciding what causal scenario is best supported by our data.

FUTURE DIRECTIONS

All results presented in this paper assume that, with the exceptions of nonshared confounding and random measurement error, estimates from sibling comparisons are unbiased. There are of course many other potential sources of bias, such as reverse causation or systematic measurement error in exposure or outcome. The relatives may further influence each other in some way, or the association may not be linear on the scale we are modeling. Studies of whether such errors need special consideration in the sibling comparison setting would make a welcome contribution to the literature.

Finally, do the caveats pointed out in this article render sibling comparison designs useless? We would say no; sibling comparison designs remain a unique tool to adjust associations for unmeasured confounding by factors shared by siblings. However, their interpretation rests on several assumptions that should be made explicit, and that may be tested in some situations, particularly concerning measurement error. In any sibling comparison study, it would be prudent to include a discussion of the (hypothesized) extent to which the exposure and the set of all confounders are shared by family members; whether confounding is likely to create a positive or negative association; and how well the observed exposure measures the causal exposure. Although it may not be possible to retrieve an unbiased estimate of the association adjusted for family-level confounders, it may often be possible to argue that the association could not be completely due to shared confounders, or that the association to some extent must be caused by such factors. Although perhaps not the definite test researchers were hoping for, there are many applications where sibling studies may prove helpful, particularly in combination with evidence from other study designs, in drawing causal conclusions about specific associations.

APPENDIX

Proof of Eq. 5:

$$\beta = \frac{Cov(Y_{ij}, X_{ij})}{Var(X_{ij})} = \frac{Cov(\beta_{YX}X_{ij} + \beta_{YC}C_{ij} + \epsilon_{Yij}, X_{ij})}{Var(X_{ij})}$$

$$= \frac{\beta_{YX}Var(X_{ij}) + Cov(\beta_{YC}C_{ij}, X_{ij})}{Var(X_{ij})} = \beta_{YX}$$

$$+ \frac{\beta_{YC}Cov(C_{ij}, \beta_{XC}C_{ij})}{Var(X_{ij})} = \beta_{YX} + \frac{\beta_{YC}\beta_{XC}\sigma_C^2}{\beta_{XC}^2\sigma_C^2 + \sigma_{\epsilon X}^2}$$

Proof of Eq. 6:

In the linear between-within model, we may exclude β_B from the model by reparameterizing Eq 2:

$$E(Y_{ij}|X_{ij}, \bar{X}_i) = \beta_{Bw*} + \beta_w(X_{ij} - \bar{X}_i) \tag{11}$$

where β_w from Eq 11 and Eq 2 will be identical.¹¹ Thus:

$$\beta_w = \frac{Cov(Y_{ij}, (X_{ij} - \bar{X}_i))}{Var(X_{ij} - \bar{X}_i)} = \frac{Cov\left[Y_{ij}, \left(\frac{X_{ij} - X_{ij'}}{2}\right)\right]}{Var\left(\frac{X_{ij} - X_{ij'}}{2}\right)}$$

$$= \frac{\frac{1}{2} [Cov(Y_{ij}, X_{ij}) - Cov(Y_{ij}, X_{ij'})]}{\frac{1}{4} [2Var(X_{ij}) - 2Cov(X_{ij}, X_{ij'})]}$$

$$= \frac{Cov(\beta_{YX}X_{ij} + \beta_{YC}C_{ij} + \epsilon_{Yij}, X_{ij}) - Cov(\beta_{YX}X_{ij} + \beta_{YC}C_{ij} + \epsilon_{Yij}, X_{ij'})}{Var(X_{ij}) - Cov(X_{ij}, X_{ij'})}$$

$$= \frac{\beta_{YX}(Var(X_{ij}) - Cov(X_{ij}, X_{ij'})) + \beta_{YC}(Cov(C_{ij}, X_{ij}) - Cov(C_{ij}, X_{ij'}))}{Var(X_{ij}) - Cov(X_{ij}, X_{ij'})}$$

$$= \beta_{YX} + \beta_{YC} \frac{Cov(C_{ij}, \beta_{XC}C_{ij} + \epsilon_{Xij}) - Cov(C_{ij}, \beta_{XC}C_{ij'} + \epsilon_{Xij'})}{Var(\beta_{XC}C_{ij} + \epsilon_{Xij}) - Cov(\beta_{XC}C_{ij} + \epsilon_{Xij}, \beta_{XC}C_{ij'} + \epsilon_{Xij'})}$$

$$= \beta_{YX} + \frac{\beta_{YC}\beta_{XC}(\sigma_C^2 - \sigma_C^2\rho_C)}{\beta_{XC}^2\sigma_C^2 + \sigma_{\epsilon X}^2 - \beta_{XC}^2\sigma_C^2\rho_C - \sigma_{\epsilon X}\rho_{\epsilon X}}$$

$$= \beta_{YX} + \frac{\beta_{YC}\beta_{XC}\sigma_C^2}{\beta_{XC}^2\sigma_C^2 + \sigma_{\epsilon X}^2 \frac{1 - \rho_{\epsilon X}}{1 - \rho_C}}$$

Proof of Eq. 8:

$$\beta^* = \frac{Cov(Y_{ij}, X_{ij}^*)}{Var(X_{ij}^*)} = \frac{Cov(Y_{ij}, X_{ij} + \epsilon_{Mij})}{Var(X_{ij}^*)}$$

$$= \frac{Cov(Y_{ij}, X_{ij})Var(X_{ij})}{Var(X_{ij})Var(X_{ij}^*)} = \beta \frac{Var(X_{ij})}{Var(X_{ij}^*)} = \beta\gamma$$

Proof of Eq. 9:

$$\begin{aligned} \beta_w^* &= \frac{\text{Cov}(Y_{ij}, (X_{ij}^* - \bar{X}_i^*))}{\text{Var}(X_{ij}^* - \bar{X}_i^*)} = \frac{\text{Cov}\left[Y_{ij}, \left(\frac{X_{ij}^* - X_{ij'}^*}{2}\right)\right]}{\text{Var}\left(\frac{X_{ij}^* - X_{ij'}^*}{2}\right)} \\ &= \frac{\frac{1}{2} [\text{Cov}(Y_{ij}, X_{ij}) - \text{Cov}(Y_{ij}, X_{ij'})]}{\frac{1}{4} [2\text{Var}(X_{ij}^*) - 2\text{Cov}(X_{ij}, X_{ij'})]} \\ &= \frac{\text{Cov}(Y_{ij}, X_{ij}) - \text{Cov}(Y_{ij}, X_{ij'})}{\frac{\text{Var}(X_{ij})}{\gamma} - \text{Cov}(X_{ij}, X_{ij'})} \\ &= \gamma \frac{\text{Cov}(Y_{ij}, X_{ij}) - \text{Cov}(Y_{ij}, X_{ij'})}{\text{Var}(X_{ij}) - \gamma \text{Cov}(X_{ij}, X_{ij'})} \\ &= \gamma \left(\beta_{YX} \frac{\beta_{XC}^2 \sigma_C^2 (1 - \rho_C) + \sigma_{\epsilon X}^2 (1 - \rho_{\epsilon X})}{\beta_{XC}^2 \sigma_C^2 (1 - \gamma \rho_C) + \sigma_{\epsilon X}^2 (1 - \gamma \rho_{\epsilon X})} + \frac{\beta_{YC} \beta_{XC} \sigma_C^2 (1 - \rho_C)}{\beta_{XC}^2 \sigma_C^2 (1 - \gamma \rho_C) + \sigma_{\epsilon X}^2 (1 - \gamma \rho_{\epsilon X})} \right). \end{aligned}$$

REFERENCES

1. Carlsson S, Andersson T, de Faire U, Lichtenstein P, Michaelsson K, Ahlbom A. Body mass index and mortality: is the association explained by genetic factors? *Epidemiology*. 2011;22:98–103.
2. Frisell T, Lichtenstein P, Rahman Q, Langstrom N. Psychiatric morbidity associated with same-sex sexual behaviour: influence of minority stress and familial factors. *Psychol Med*. 2010;40:315–324.
3. Öberg S, Cnattingius S, Sandin S, Lichtenstein P, Iliadou AN. Birth weight predicts risk of cardiovascular disease within dizygotic but not monozygotic twin pairs: a large population-based co-twin-control study. *Circulation*. 2011;123:2792–2798.
4. Donovan SJ, Susser E. Commentary: advent of sibling designs. *Int J Epidemiol*. 2011;40:345–349.
5. Susser E, Eide MG, Begg M. Invited commentary: the use of sibship studies to detect familial confounding. *Am J Epidemiol*. 2010;172:537–539.

6. Griliches Z. Sibling models and data in economics—beginnings of a survey. *J Polit Econ*. 1979;87:S37–S64.
7. D’Onofrio BM, Singh AL, Iliadou A, et al. Familial confounding of the association between maternal smoking during pregnancy and offspring criminality: a population-based study in Sweden. *Arch Gen Psychiatry*. 2010;67:529–538.
8. Kuja-Halkola R, D’Onofrio BM, Iliadou AN, Langstrom N, Lichtenstein P. Prenatal smoking exposure and offspring stress coping in late adolescence: no causal link. *Int J Epidemiol*. 2010;39:1531–1540.
9. Obel C, Olsen J, Henriksen TB, et al. Is maternal smoking during pregnancy a risk factor for hyperkinetic disorder?—Findings from a sibling design. *Int J Epidemiol*. 2011;40:338–345.
10. Eriksen W, Sundet JM, Tambs K. Birth weight standardized to gestational age and intelligence in young adulthood: a register-based birth cohort study of male siblings. *Am J Epidemiol*. 2010;172:530–536.
11. Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Stat Med*. 2003;22:2591–2602.
12. Carlin JB, Gurrin LC, Sterne JAC, Morley R, Dwyer T. Regression models for twin studies: a critical review. *Int J Epidemiol*. 2005;34:1089–1099.
13. Sjölander A, Frisell T, Öberg S. Causal interpretation of between-within models for twin research. In press.
14. Bound J, Solon G. Double trouble: on the value of twins-based estimation of return to schooling. *Econ Educ Rev*. 1999;18:169–182.
15. Sjölander A, Johansson A, Lundholm C, Altman D, Pawitan Y. Analysis of 1:1 matched cohort studies and twin studies, with binary exposures and binary outcomes. In press.
16. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press; 2000.
17. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010;39:417–420.
18. McGue M, Osler M, Christensen K. Causal inference and observational research: the utility of twins. *Perspect Psychol Sci*. 2010;5:546–556.
19. Tettamanti G, Altman D, Pedersen NL, Bellocco R, Milsom I, Iliadou AN. Effects of coffee and tea consumption on urinary incontinence in female twins. *Br J Obstet Gynaecol*. 2011;118:806–813.
20. Sadler ME, Miller CJ, Christensen K, McGue M. Subjective wellbeing and longevity: a co-twin control study. *Twin Res Hum Genet*. 2011;14:249–256.
21. Xu WL, Atti AR, Gatz M, Pedersen NL, Johansson B, Fratiglioni L. Midlife overweight and obesity increase late-life dementia risk: a population-based twin study. *Neurology*. 2011;76:1568–1574.
22. Szklo M, Nieto J. Information bias. *Epidemiology: Beyond the Basics*. Jones & Barlett Learning; 2007.
23. Dwyer T, Blizzard L. A discussion of some statistical methods for separating within-pair associations from associations among all twins in research on fetal origins of disease. *Paediatr Perinat Epidemiol*. 2005;19(suppl 1):48–53.
24. Groen-Blokhus MM, Middeldorp CM, van Beijersveldt CEM, Boomsma D. Evidence for a causal association of low birth weight and attention problems. *J Am Acad Child Adolesc Psychiatry*. 2011;50:1247–1254.