Epidemiologic Methods

Volume 1, *Issue* 1 2012 *Article* 10

Causal Interpretation of Between-Within Models for Twin Research

Arvid Sjölander Mr, *Karolinska Institute*
Thomas Frisell, *Karolinska Institute*
Sara Öberg, *Harvard School of Public Health*

Recommended Citation:

Sjölander, Arvid Mr; Frisell, Thomas; and Öberg, Sara (2012) "Causal Interpretation of Between-Within Models for Twin Research," *Epidemiologic Methods*: Vol. 1: Iss. 1, Article 10.
DOI: 10.1515/2161-962X.1015

Causal Interpretation of Between-Within Models for Twin Research

Arvid Sjölander Mr, Thomas Frisell, and Sara Öberg

Abstract

The between-within (BW) model is a popular regression model for twin data. Despite its popularity, the properties of the BW model have not yet been thoroughly investigated, and most reviews are largely heuristic. The aim of this paper is to provide a formal guide to the causal interpretation of BW models. First, we show that the BW model produces a specific subpopulation causal effect, in the absense of non-shared confounding. Second, we discuss how to adjust for non-shared confounders in the BW model. We show that if the adjustment is done in a `standard fashion', then the desired causal interpretation is lost. We further show that the causal interpretation can be retrieved with a minor modification of the model.

KEYWORDS: Between-within, causal inference, twin studies

Author Notes: Sjölander acknowledges financial support from the Swedish Research Council [2008-5375], Frisell acknowledges financial support from the Swedish Prison and Probation Service R&D, Öberg acknowledges financial support from the Swedish Council for Working Life and Social Research [2011-1477].

1 Introduction

A common aim of epidemiological research is to estimate the causal effect of a particular exposure on a particular outcome. In observational (i.e. non-randomized) studies, the exposure-outcome association is often confounded by extraneous factors and can therefore generally not be given a causal interpretation. Thus, a crucial and non-trivial task for epidemiologists is to identify influential confounders to adjust for in the analysis. Twin studies are of special interest because they provide pairs which have been naturally matched on a large number of potential confounders, e.g. genetic make-up (for monozygotic twins) and childhood environment (for twins reared together). We refer to confounders that are constant within each twin pair as shared, and to confounders that may vary within a twin pair as non-shared.

To assess the exposure-outcome association, regression models are commonly used. In a review of various types of regression models for twin data, Carlin et al (2005) presented the between-within (BW) model, which decomposes the exposure-outcome association into a between-effect and a within-effect. In recent years, the BW model have gained considerable popularity, and has been used in the analysis of various twins cohorts (Morley et al, 2005; Skidmore et al, 2006, 2008; Öberg et al, 2007, 2009). There are two versions of the BW model, one with a fixed intercept and one with a random intercept (Carlin et al, 2005). Whereas the random intercept BW model has been carefully scrutinized in the theoretical literature (e.g. Neuhaus and McCulloch (2006); Brumback et al (2010)), the properties of the fixed intercept BW model have not yet been thoroughly investigated, and most reviews are largely heuristic (Begg and Parides, 2003; Carlin et al, 2005; Dwyer and Blizzard, 2005). While previous reviews seem to indicate a causal interpretation for the within-effect (in the absence of non-shared confounders) the wording used to support such a claim is at best vague (Begg and Parides, 2003; Carlin et al, 2005; Dwyer and Blizzard, 2005). For example, it has not been established what type of causal effect the within-effect would represent. Although it could be inferred that a causal interpretation would only hold in the absence of non-shared confounders, there is no reference to if and how such confounders could be handled in the fixed intercept BW model. Lastly, the fixed intercept BW model has mostly been applied to continuous exposures and outcomes, and it has not been made clear whether the desired causal interpretation holds for non-continuous (e.g. binary) variables as well. For instance, Carlin et al (2005) stated; 'With dichotomous covariates the

concept of between-pair regression effects may not be useful', but they did not provide a formal argument to support this claim.

The aim of this paper is to provide a formal guide to the causal interpretation of fixed intercept BW models. First, we show that the fixed intercept BW model produces a specific subpopulation causal effect, in the absence of non-shared confounders. Second, we discuss how to adjust for non-shared confounders in the fixed intercept BW model. We show that if the adjustment is done in a 'standard fashion', then the desired causal interpretation is lost. We further show that the causal interpretation can be retrieved with a minor modification of the model.

The paper is organized as follows. In Section 2 we describe the data structure in twin studies, and the assumptions typically made regarding the data generating mechanism. In Section 3 we introduce the potential outcome framework, which is used to define causal effects. We make the distinction between population causal effects and subpopulation causal effects. In Section 4 we discuss how to adjust for non-shared confounders in generalized linear models. In Section 5 we present the generalized linear fixed effect BW model and give a formal motivation as to why and how this model adjusts for shared confounders. We briefly mention the random intercept BW model, and discuss in what sense it differs from the fixed intercept BW model. In Section 6 we discuss how to adjust for shared and non-shared confounders simultaneously in a generalized linear fixed effect BW model. In Section 7, we carry out a small simulation study. Our focus is on the interpretation of model parameters, not on the finite sample properties of their statistical estimates. Thus, our main exposition will be void of any statistical measures of sampling variability, such as p-values, standard errors, and confidence intervals. In Section 8 we make concluding remarks, discuss how to fit the proposed models, and how to obtain standard errors for the parameter estimates. Since our focus is exclusively on fixed intercept BW models, we use 'BW model' from now on to abbreviate 'fixed intercept BW model'.

2 Data structure and assumptions

In prospective or cross-sectional twin studies, data is recorded on the exposure of interest, X, and the outcome of interest, Y, for each twin in n pairs. We arbitrarily label the two twins in each pair as '1' and '2', and we let X_j and Y_j denote the exposure and outcome, respectively, for twin j (1 or 2) in

a given pair. We let U denote the set of all shared potential confounders, and we let V denote the set of all non-shared potential confounders. We let V_j denote the value of V for twin j in a given pair. We will supress the index j for X, Y and V when not needed.

As a concrete example, consider the study by Morley et al (2005). In this study, the aim was to determine whether there is a causal effect of cord blood erythropoietin (EPO) on birth weight. The authors analyzed a data set consisting of both monozygotic (MZ) and dizygotic (DZ) twins. In this setting, U would pertain to all factors that are necessary equal for both twins in a twin pair, including observed factors such as length of gestation and maternal age, but also a large set of unobserved factors such as maternal nutrition, lifestyle, and health during pregnancy. For MZ twins, U would also include genetic make up. For DZ twins, V would include infant sex (observed) and genetic make up (unobserved), since these factors may differ within a DZ twin pair.

Several of the potential confounders (both shared and non-shared) may be difficult or even impossible to observe (e.g. maternal lifestyle and genetic make up). Twin data offer a unique 'natural' resource to adjust for unobserved confounders, as long as they are perfectly shared within each pair (see Section 5). Thus, when collecting data the focus is often on non-shared potential confounders, with the ultimate aim of making V completely observed.

The causal diagram (Pearl, 1995; Greenland et al, 1999, 2; Pearl, 2000) in Figure 1 displays a possible data generating mechanism. In Figure 1, each arrow represents a possible causal influence. For instance, both X_j and Y_j are possibly affected by V_j and/or U, in which case the association between X_j and Y_j is confounded. Figure 1 encodes two structural (i.e. causal) assumptions. 1) (X_1,Y_1) have no causal influence on (X_2,Y_2), and vice versa, and 2) V_1, V_2, and U do not cause each other, and have no common causes. The first assumption implies that (X_1, Y_1) and (X_2, Y_2) are conditionally independent, given U. In statistical notation, we express this as

$$(X_1, Y_1) \amalg (X_2, Y_2)|U, \tag{1}$$

where we have used $A \amalg B|C$ for 'A and B conditionally independent, given C'. Assumption (1) is standard in the literature, and is used in several common analysis methods, such as conditional logistic regression and random effects models. The second assumption is made only to simplify the exposition; none of the results presented in this paper rely on this assumption.

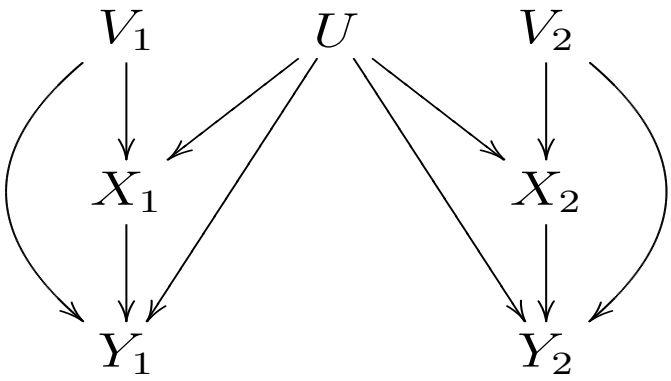

Figure 1: A possible causal mechanism

3 Population causal effects and subpopulation causal effects

To quantify the association between exposure X and outcome Y, it is common to measure how the mean outcome changes when the exposure changes with 1 unit. Specifically, it is common to focus on contrasts on the form

$$g\{E(Y|X = x + 1)\} - g\{E(Y|X = x)\}, \tag{2}$$

where $g(\cdot)$ is a smooth link function. For continuous outcomes, the identity link is a natural choice, which gives the mean difference $E(Y|X = x + 1) - E(Y|X = x)$. For binary (0/1) outcomes, the log link and logit link are natural, which for binary (0/1) exposures give the log risk ratio $\log\left\{\frac{\Pr(Y=1|X=1)}{\Pr(Y=1|X=0)}\right\}$ and log odds ratio $\log\left\{\frac{\Pr(Y=1|X=1)\Pr(Y=0|X=0)}{\Pr(Y=0|X=1)\Pr(Y=1|X=0)}\right\}$, respectively. Other common links are the probit link and the complementary log-log link.

The contrast in (2) compares the mean outcome for two different groups; those exposed to level $X = x$ and those exposed to level $X = x + 1$. In the absence of confounders (i.e. $V = U = \emptyset$), individuals are exchangeable (i.e. 'comparable') across exposure levels, and the contrast in (2) can be given a causal interpretation. To formalize, we use the potential outcome framework (Rubin, 1974; Pearl, 2000; Hernan and Robins, 2006). Towards

this end we conceive of each individual as having one potential outcome for each possible level of the exposure. We let Y^x denote the potential outcome for a given individual, if the individual is potentially exposed to level $X = x$. In practice, each individual is only observed under one exposure level. If the individual is factually exposed to level $X = x$, then the potential outcome Y^x is observed and equal to the factual outcome Y for that individual:

$$X = x \Rightarrow Y^x = Y. \tag{3}$$

The potential outcomes under all other exposure levels are then unobserved, or counterfactual. The relation in (3) is often referred to as the 'consistency assumption' (Pearl, 2000). In the potential outcome framework, exchangeability is defined as

$$Y^x \amalg X. \tag{4}$$

Let $E(Y^x)$ denote the average outcome in the study population, if all individuals would, contrary to fact, be exposed to level $X = x$. Under (3) and (4), we have that

$$E(Y|X = x) = E(Y^x|X = x) = E(Y^x), \tag{5}$$

where the first equality follows from (3) and the second equality follows from (4). It follows from (5) that the contrast in (2) equals

$$g\{E(Y^{x+1})\} - g\{E(Y^x)\}. \tag{6}$$

The contrast in (6) is by definition a causal effect, since it compares the same population under two exposure levels, and thus 'like with like'. It is a population (marginal) causal effect, since it applies to the whole twin study population.

In the presence of confounders (either shared or non-shared), the contrast in (2) cannot be given a causal interpretation. However, if both U and V are fully observed, a causal interpretation can in principle be retreived by conditioning (e.g. stratifying) on (U, V). In Figure 1, there is conditional exchangeability, given (U, V):

$$Y^x \amalg X|(U, V). \tag{7}$$

Under (3) and (7) we have that $E(Y|X = x, U, V) = E(Y^x|U, V)$, so that the contrast

$$g\{E(Y|X = x + 1, U, V)\} - g\{E(Y|X = x, U, V)\} \tag{8}$$

equals

$$g\{E(Y^{x+1}|U,V)\} - g\{E(Y^x|U,V)\}. \tag{9}$$

The contrast in (9) is a subpopulation (conditional) causal effect, since it applies to the subpopulation (stratum) defined by a distinct level of (U,V).

In the presence of confounders, population (marginal) causal effects can be obtained through standardization (Pearl, 2000; Hernan and Robins, 2006). Under (3) and (7) we have that

$$E\{E(Y|X=x,U,V)\} = E\{E(Y^x|U,V)\} = E(Y^x),$$

where the outer expectation is taken over the population distribution for (U,V), and the second equality follows from the law of total probability. For instance, suppose that the only confounder is a gene with alleles $G=0$ and $G=1$, having allele frequencies 0.95 and 0.05, respectively. The population causal effect of increasing X from x to $x+1$ is then obtained as

$$\begin{aligned}
&g\{E(Y^{x+1})\} - g\{E(Y^x)\} \\
&\quad = g[E\{E(Y|X=x+1,G)\}] - g[E\{E(Y|X=x,G)\}] \\
&\quad = g[\{0.95\times E(Y|X=x+1,G=0) + 0.05\times E(Y|X=x+1,G=1)\}] \\
&\qquad - g[\{0.95\times E(Y|X=x,G=0) + 0.05\times E(Y|X=x,G=1)\}].
\end{aligned}$$

We end this section by noting that standardization can also be used to compute subpopulation effects. For instance, suppose that 'sex' is contained in U or V. From the law of total probability we then obtain the causal effect in the subpopulation of males by averaging over the distribution for (U,V) among the males:

$$\begin{aligned}
&g\{E(Y^{x+1}|\text{Sex=male})\} - g\{E(Y^x|\text{Sex=male})\} \\
&\quad = g[E\{E(Y|X=x+1,U,V)|\text{Sex=male}\}] \\
&\quad - g[E\{E(Y|X=x,U,V)|\text{Sex=male}\}].
\end{aligned} \tag{10}$$

4 Adjusting for non-shared confounders in generalized linear models

In practice, the set of shared potential confounders U is often largely unobserved, and the set of non-shared potential confounders V is rarely completely

observed. Thus, stratification on (U, V) is typically not feasible. Furthermore, stratification may often produce highly inefficient estimates due to sparse data. A standard way to gain efficiency is to use a regression model. Let V_{obs} denote the observed subset of V. Generalized linear models (GLMs) is a widely used class of models, which states that the mean of Y, transformed with a smooth link function $g(\cdot)$, is linear in X and V_{obs}, e.g.

$$g\{E(Y|X, V_{obs})\} = \alpha + \beta X + \gamma V_{obs}. \tag{11}$$

Suppose that there are no shared confounders, and that all non-shared confounders are observed (i.e. $U = \emptyset, V = V_{obs}$). Under (3) and (7) we then have that

$$g\{E(Y^{x+1}|V)\} - g\{E(Y^x|V)\} = \beta.$$

That is, β is a subpopulation (V-specific) causal effect of increasing X from x to $x + 1$, on the mean of Y. In principle, this parameter could vary with V (or with x). However, the absence of interaction terms between X and V in (11) encodes the assumption that the effect is constant across levels of V.

Using standardization, we obtain the population causal effect, on the scale defined by $g(\cdot)$, as

$$\begin{aligned} &g\{E(Y^{x+1})\} - g\{E(Y^x)\} \\ &\quad = g[E\{E(Y|X = x + 1, V)\}] - g[E\{E(Y|X = x, V)\}] \\ &\quad = g[E\{g^{-1}(\alpha + \beta x + \beta + \gamma V)\}] - g[E\{g^{-1}(\alpha + \beta x + \gamma V)\}]. \end{aligned} \tag{12}$$

When $g(\cdot)$ is the identity link or the log link, the expression in (12) simplifies to β. Thus, for these links the population effect equals the common subpopulation effect; we say that the mean difference and the log risk ratio are are 'collapsible' (Greenland et al, 1999, 1). Although collapsibility may seem intuitively reasonable, it does not hold for all link functions. For instance, when $g(\cdot)$ is the logit link, the expression in (12) does not simplify further, and is not in general equal to β. Thus, the log odds ratio is non-collapsible.

5 Adjusting for shared confounders in generalized linear BW models

Explicit confounder adjustment, as in (11), is only possible for confounders that are observed. However, twin data offer a unique 'natural' resource to

adjust for shared confounders that are unobserved, through implicit techniques.

One common analysis tool is the BW model (Carlin et al, 2005). Let $\bar{X}$ denote the average exposure level for the given pair, i.e. $\bar{X} = (X_1 + X_2)/2$. The linear BW model is defined as

$$E(Y_j|X_j, \bar{X}) = \beta_0 + \beta_W X_j + \beta_B \bar{X},$$

where β_W is the 'within-effect' and β_B is the 'between-effect' (Carlin et al, 2005). Informally, it has been argued that by including the between-effect in the model , the within-effect becomes adjusted for all shared confounders (Begg and Parides, 2003; Carlin et al, 2005; Dwyer and Blizzard, 2005). In this section we provide a formal underpinning. We first note that the linear model is most natural for continuous outcomes. In line with standard GLMs, we define the generalized linear BW model as

$$g\{E(Y_j|X_j, \bar{X})\} = \beta_0 + \beta_W X_j + \beta_B \bar{X}, \tag{13}$$

where $g()$ is a smooth link function. A variant of the generalized linear BW model replaces the fixed intercept β_0 with a pair-specific random intercept. The random intercept BW model was discussed in detail by Neuhaus and McCulloch (2006). They showed that by treating the intercept as random, β_W can be interpreted as being conditioned on U. Sjölander et al (2011) provided an in-depth comparison between the fixed intercept BW model and the random intercept BW model for binary exposures and outcomes. We restrict our attention to the fixed intercept BW model in (13).

Let j' denoting the co-twin of j, i.e. $j' = 1$ if $j = 2$ and $j' = 2$ if $j = 1$. To interpret β_W, note that model (13) implies the relation

$$\begin{aligned} \beta_W \Delta x &= g\{E(Y_j|X_j = x_0 + \Delta x, X_{j'} = x_0)\} \\ &- g\{E(Y_j|X_j = x_0, X_{j'} = x_0 + \Delta x)\} \end{aligned} \tag{14}$$

for an arbitrary reference level x_0 and increment Δx. In the absence of non-shared confounders (i.e. $V = \emptyset$), we can use standardization arguments (see Appendix A) to show that

$$E(Y_j|X_j = x, X_{j'} = x') = E(Y^x|X_j = x, X_{j'} = x'). \tag{15}$$

That is, among the pairs where twin j is exposed to level $X = x$ and twin j' is exposed to level $X_{j'} = x'$, the observed mean of Y in twin j is identical to the mean of Y in both twins, had they both been exposed to level $X = x$. Combining (14) and (15) gives that

$$\begin{aligned} \beta_W \Delta x &= g\{E(Y^{x_0+\Delta x}|X_j = x_0, X_{j'} = x_0 + \Delta x)\} \\ &- g\{E(Y^{x_0}|X_j = x_0, X_{j'} = x_0 + \Delta x)\}. \end{aligned} \tag{16}$$

That is, in the absence of non-shared confounders we may interpret $\beta_W \Delta x$ as the causal effect of increasing X from x_0 to $x_0 + \Delta x$ (on the scale defined by $g()$), in the subpopulation of twin pairs where one twin has $X = x_0$, and the other twin has $X = x_0 + \Delta x$. This argument shows that it may be somewhat misleading to refer to β_W as a 'within-pair effect', when in fact β_W is a subpopulation effect and not a pair- (i.e. $U-$) specific effect.

An important special case occurs when X is binary (0/1). If one twin has exposure level $x_0 = 1$, then the other twin cannot have exposure level $x_0 + 1 = 2$. Thus, with binary exposures, $\beta_W \Delta x$ is only defined for reference level $x_0 = 0$ and increment $\Delta x = 1$. The subpopulation of twin pairs where one twin has exposure level $x = 0$ and the other twin has exposure level $x = 1$ is the subpopulation of exposure discordant twin pairs. Thus, with binary exposures and in the absence of non-shared confounders, β_W can be interpreted as the causal effect of X on Y, for the subpopulation of exposure discordant pairs. This special result was derived by Sjölander et al (2011). For continuous exposures, β_W is defined for all exposure levels.

Two subtle points merit attention.

1. Suppose that we use the identity link or log link. These links do not induce non-collapsibility, which in the present context has the following implication. Consider the population which consists of all twin pairs where the difference in exposure levels is Δx, i.e. $|X_1 - X_2| = \Delta x$. This population is the union of all subpopulations where one twin has $X = x_0$, and the other twin has $X = x_0 + \Delta x$, where the union is taken over all possible reference levels x_0. Under collapsibility, the effect in the union population is identical to the common effect in each of the subpopulations ($= \beta_W \Delta x$). Hence, for the identity link and the log link, the interpretation of β_W simplifies; we may, in the absence of non-shared confounders, interpret $\beta_W \Delta x$ as the causal effect of increasing

X with Δx (on the scale defined by $g()$), in the population of twin pairs where $|X_1 - X_2| = \Delta x$. However, as discussed in Section 4, the logit link induces non-collapsibility. Thus, when we use the logit link, this interpretational simplification is not valid.

2. The model in (13) postulates a linear effect of X (on the scale defined by $g()$). However, the linearity applies *across* subpopulations of twin pairs, not *within* these subpopulations. To clarify this point, suppose that X is three-level numerical (0/1/2), and that we use the identity link. The effect of increasing X with 1 unit, among the twin pairs with $|X_1 - X_2| = 1$, is to increase the mean of Y with $\beta_W \times 1$ units. Similarly, the effect of increasing X with 2 units, among the twin pairs with $|X_1 - X_2| = 2$, is to increase the mean of Y with $\beta_W \times 2$ units. Thus, there is a linear effect *across* subpopulations of twin pairs. However, the effect of increasing X with 2 units, among the twin pairs with $|X_1 - X_2| = 1$, is unspecified by the model. In particular, the model does not imply that this effect is equal $\beta_W \times 2$ units. Similarly, the effect of increasing X with 1 unit, among the twin pairs with $|X_1 - X_2| = 2$, is unspecified by the model, and in particular not necessarily equal to $\beta_W \times 1$ units. Thus, the model does not imply a linear dose-response relationship *within* subpopulations of twin pairs. To put it generally, the model is agnostic to the effect of any increases in X, other than the one actually observed, for each subpopoulations of twin pairs.

6 Adjusting for shared and non-shared confounders simultaneously

In Section 4 we discussed how to explicitly adjust for observed non-shared confounders in a standard GLM. In Section 5 we discussed how to implicitly adjust for unobserved shared confounders in a generalized linear BW model. In this section we discuss how these techniques can be combined to adjust for both an observed subset of V and an unobserved set U.

A naive attempt to adjust for both V_{obs} and U, is to include $V_{obs,j}$ in the BW model, e.g.:

$$g\{E(Y_j|X_j, \bar{X}, V_{obs,j})\} = \beta_0 + \beta_W X_j + \beta_B \bar{X} + \gamma V_{obs,j}. \quad (17)$$

We will argue that model (17) does not adjust properly for U and V_{obs}. In

particular, even if model (17) is correctly specified, and $V = V_{obs}$, β_W does not have a causal interpretation.

We first give a heuristic explanation of the problem. In Appendix B we give a more rigorous explanation. In Figure 1, both X_j and Y_j are conditionally independent of $V_{j'}$, given V_j and U. However, the BW model (17) does not adjust explicitly for U; the adjustment is implicit by using $\bar{X}$ as a proxy for U. Conditioning on X_j and $\bar{X}$, as in model (17), is equivalent to conditioning on X_j and $X_{j'}$. By conditioning on $X_{j'}$ we 'open' the paths $X_j \leftarrow U \rightarrow X_{j'} \leftarrow V_{j'}$ and $Y_j \leftarrow U \rightarrow X_{j'} \leftarrow V_{j'}$, since $X_{j'}$ is a collider on both these paths (Pearl, 2000; Cole et al, 2010). This means that adjusting for $\bar{X}$ will induce a 'spurious' (i.e. non-causal) association between X_j and $V_{j'}$, and between Y_j and $V_{j'}$, so that $V_{j'}$ becomes an 'artificial' confounder for the X_j-Y_j association. Notably, this heuristic explanation does not make any reference to the path $V_j \rightarrow Y_j$. This suggests that the induced bias does not depend on whether and how V_j affects Y_j.

Further insight can be made by considering a concrete example. Suppose that the structural (i.e. causal) equations for X_j and Y_j are given by

$$\begin{aligned} X_j &= \alpha_U U + \alpha_V V_j + \epsilon_{X_j} \\ Y_j &= \delta_U U + \delta_V V_j + \delta_X X_j + \epsilon_{Y_j}, \end{aligned} \tag{18}$$

with V_1, V_2, U, ϵ_{X_1}, ϵ_{X_2}, ϵ_{Y_1}, and ϵ_{Y_2} being independent standard Normal. δ_X is the causal effect of X on Y, both conditional on (V, U), and (due to collapsibility) in the whole population. It can then be shown (see Appendix C) that if $V = V_{obs}$, then model (17) holds with $g(\cdot)$ being the identity link, but

$$\beta_W = \delta_X + \frac{\delta_U \alpha_U \alpha_V^2}{(1+\alpha_U^2)(1+\alpha_V^2) + \alpha_U^2}. \tag{19}$$

Thus, β_W does not equal the causal effect δ_X. In particular, $\beta_W \neq 0$ even in the absence of a causal effect ($\delta_X = 0$). The bias depends on the effect of U on X_j (through α_U), the effect of V_j on X_j (through α_V), and the effect of U on Y_j (through δ_U). However, the absence of δ_V in the bias term confirms what the heuristic explanation suggested; the bias does not depend on whether and how V_j affects Y_j. As a comparison, suppose that we don't adjust for V_{obs} at all, thus fitting model (13) with identity link. It can be

shown that this model holds as well, with

$$\beta_W = \delta_X + \frac{\delta_V \alpha_V}{1 + \alpha_V^2}. \tag{20}$$

The bias term in (20) depends on the effect of V_j on X_j (through α_V), and the effect of V_j on Y_j (through δ_V). If the effect of V_j on Y_j is sufficiently weak (δ_V close to 0), then it is actually possible to increase bias by adjusting for $V_{j,obs}$ using model (17), as compared to not adjusting for $V_{j,obs}$ at all.

Fortunately, there is a simple way to eliminate the bias induced by adjusting for $V_{obs,j}$, namely by including $V_{obs,j'}$ in the model as well, e.g.:

$$g\{E(Y_j|X_j, \bar{X}, V_{obs,j}, V_{obs,j'})\} = \beta_0 + \beta_W X_j + \beta_B \bar{X} + \gamma V_{obs,j} + \gamma' V_{obs,j'}. \tag{21}$$

If $V = V_{obs}$, then we can use standardization arguments (see Appendix D) and express $\beta_W \Delta x$ under model (21) as

$$\begin{aligned} \beta_W \Delta x \;\; = \;\; & g\{E(Y^{x_0+\Delta x}|X_j = x_0 + \Delta x, X_{j'} = x_0, V_j = V_{j'} = v\} \\ & - g\{E(Y^{x_0}|X_j = x_0, X_{j'} = x_0 + \Delta x, V_j = V_{j'} = v\}, \end{aligned} \tag{22}$$

for any fixed level v. Thus, if $V = V_{obs}$, then we may interpret $\beta_W \Delta x$ under model (21) as the causal effect of increasing X from x_0 to $x_0 + \Delta x$ (on the scale defined by $g()$), in the subpopulation of twin pairs where one twin has $X = x_0$, and the other twin has $X = x_0 + \Delta x$, and where both twins have $V = v$, for any fixed level v. Arguing as in Appendix C, it can be shown that under the data generating mechanism (18), and with $V = V_{obs}$, model (21) holds with $\delta_X = \beta_W$.

7 Simulation

To empirically verify the theoretical conclusions in Section 6, we carried out a small simulation study. We draw 1000 samples of 100 individuals each, from the model in (18), with $\alpha_U = 0.5$, $\alpha_V = 0.6$, $\delta_U = 0.7$, and $\delta_V = 0$. The true causal effect was set to $\delta_X = 0.2$. Each sample was analyzed with models (13), (17), and (21), with $V = V_{obs}$ and $g(\cdot)$ being the identity link. For each of the three models, the mean (over the 1000 samples) estimate of β_W was computed. This procedure was repeated for a

sequence of values for δ_V, ranging from 0 to 0.3. Figure 2 displays the result. In Figure 2, the solid tilted line is the theoretical value for β_W under model (13) as a function of δ_V (given by equation (20)), and the solid horizontal lines at $\hat{\beta}_W = 0.26$ and $\hat{\beta}_W = 0.2$ are the theoretical values for β_W under model (17) (given by equation (19)) and model (21) ($= \delta_X$). The dashed lines are the corresponding mean estimates. The simulation confirms the theoretical derivations in Section (6). The mean estimates are very close to the theoretical values. As predicted by theory, the model that only adjusts for V_j (model (17)) exhibits more bias than the model that makes no adjustment for V at all (model (13)), when the effect of V_j on Y_j is small (i.e. for small values of δ_V).

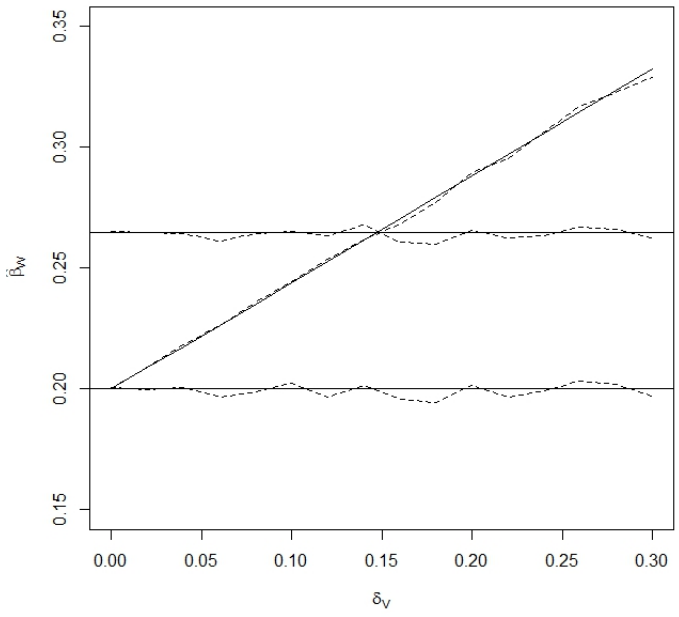

Figure 2: Simulation results.

8 Discussion

The BW model is a versatile tool in the analysis of twin data. In this paper we have shown that in the absence of non-shared confounders, the within-effect can be interpreted as a specific subpopulation causal effect. We have further shown that this interpretation is valid for a wide class of data (e.g. continuous, categorical, binary etc). We have shown that the causal interpretation is lost when non-shared confounders are incorporated in the model in a 'standard fashion'. Finally, we have shown that the causal interpretation can easily be retrieved, by including the non-shared confounders for both twins simultaneously in the model.

In our exposition, we have focused on the interpretation of model parameters. All models that we have considered are special cases of GLMs, and can be fitted with standard software. By default, the software typically output standard errors that assume independent observations. Robust (sandwich) standard errors, that take the clustered nature of the twin data into account, can typically be obtained from the software as well.

The subpopulation for which the within-effect applies may often be small, as compard to the whole population. Thus, it can be argued that the within-effect may, at times, have limited scientific relevance. However, if U is completely unobserved (which is often the case), then this subpopulation may be the only population for which we can identify any causal effect, without making strong untestable assumptions. As an example, consider the situation when there are no non-shared confounders ($V = \emptyset$), both X and Y are binary (0/1), and $g(\cdot)$ is the logit link. In this case, the within-effect β_W equals the causal effect of X on Y, on the log odds ratio scale, in the exposure discordant twin pairs ($X_1 \neq X_2$). Naturally, there are several instances where most twin pairs will be exposure concordant ($X_1 = X_2$). It would thus be desirable to find an alternative analysis method which does not restrict attention to the exposure discordant pairs. One standard alternative is conditional logistic regression (CLR), which estimates the conditional log odds ratio

$$\beta(U) = \log\left\{\frac{\Pr(Y^1 = 1|U)\Pr(Y^0 = 0|U)}{\Pr(Y^1 = 0|U)\Pr(Y^0 = 1|U)}\right\},$$

under the assumption that $\beta(U)$ is constant across levels of U, i.e. equal to some constant β^* (Sjölander et al, 2011). Neuhaus and McCulloch (2006) showed that the within-effect estimate obtained by fitting a random intercept

BW model will typically be very close to the conditional logistic regression estimate $\hat{\beta}^*_{CLR}$. Because U is by definition constant within each pair, we may interpret β^* as a pair-specific causal effect. It applies to all pairs, not only to those that are exposure discordant. Thus, at first glance it may seem as if both CLR and the random intercept BW model allow the analyst to infer causal effects for exposure concordant pairs, whereas the fixed intercept BW model does not. This conclusion is misleading, for the following reasons. An analyst is always at the liberty to assume *a priori* that $\beta(U) = \beta^*$. But equally well (s)he may simply assume that the effect in the exposure concordant pairs is equal to β_W. Neither of these assumptions are stronger than the other, since neither of them implies the other. Furthermore, with paired (e.g. twin) data and U being unobserved, both assumptions are untestable. In any realistic scenario, both assumptions are likely to be violated, to some extent. Sjölander et al (2011) showed that in general (i.e. regardless of whether $\beta(U) = \beta^*$) $\hat{\beta}^*_{CLR}$ converges to the logarithm of a weighted average of U-specific odds ratios, among the exposure discordant pairs:

$$\hat{\beta}^*_{CLR} \rightarrow \log\left[E\left\{W(U)\beta(U)|X_1 \neq X_2\right\}\right].$$

Thus, it is clear that neither CLR nor the random intercept BW model informs the analyst about causal effects for exposure concordant pairs, at least not to any wider extent than does the fixed intercept BW model.

The (fixed intercept) BW model was introduced and has been advocated predominantly as a method to evaluate the influence of shared factors on a given exposure-outcome association (by comparing between and within estimates). However, there has been surprisingly little emphasis on formal definitions for the interpretation of the obtained effects (Begg and Parides, 2003; Carlin et al, 2005; Dwyer and Blizzard, 2005). Aiming to fill this gap in the previous literature, the present work should be of great consequence to both past and future within-twinpair analyses. The conclusions presented should facilitate the interpretation of forthcoming studies as well as previously published results. Whereas previous reviews have focused on continuous data, we have shown that the causal interpretation applies to arbitrary kinds of numeric data, thus allowing for greater flexibility in the analyses. Moreover, the influence and handling of non-shared confounders has largely been neglected in previous work. The present illustration of how non-shared confounders could be adjusted for in the analyses so to maintain a causal interpretation therefore provides an important contribution to the future practice of within-twinpair analysis.

References

Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Stats in Med* 2003; **22**: 2591-2602.

Brumback BA, Dailey AB, Brumback LC, Livingston MD, He Z. Adjusting for confounding by cluster using generalized linear mixed models. *Stat Probabil Lett*; **80**: 1650-1654.

Carlin JB, Gurrin LC, Sterne JAC, Morley R, Dwyer T. Regression models for twin studies: a critical review. *Inl J Epi* 2005; **34**: 1089-1099.

Cassidy A, Skidmore P, Rimm EB, Welch A, Fairweather-Tait S, Skinner J, Burling K, Richards JB, Spector TD, MacGregor AJ. Plasma adiponectin concentrations are associated with body composition and plant-based dietary in female twins. *J Nutr* 2009; **139**: 353-358.

Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C, Illustrating bias due to conditioning on a collider. *Int J Epidemiol* 2010; **39**: 417-20.

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 Ep* 2005; **19**: 48-53.

Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; **14**: 29-46.

Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiological research. *Epidemiology* 1999; **10**: 37-48.

Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006; **60**: 578-586.

Morley R, Moore VM, Dwyer T, Owens JA, *U*mstad MP, Carlin JB. Association between erythropoietin in cord blood of twins and size at birth; does it relate to gestational factors or to factors during labor or delivery? *Pediatr Res* 2005; **57**: 680-684.

Neuhaus JN, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society* B 2006; **68**(5): 859-872.

Pearl J. Causal diagrams for empirical research. *Biometrika* 1995, **82**(4): 669-710.

Pearl J. *Causality: Models, Reasoning and Inference* 2000, Cambridge University press: Cambridge.

Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974; **66**(5): 688-701.

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 revision*. 2011.

Skidmore PML, Cassidy A, Swaminathan R, Falchi M, Spector TD, MacGregor AJ. Intrauterine, environmental, and genetic influences in the relationship between birth weight and lipids in a female twin cohort. *Arterioscler Thromb Vasc Biol* 2006; **26**: 2373-2379.

Skidmore PML, Cassidy A, Swaminathan R, Richards JB, Spector TD, MacGregor AJ. Relation of birth weight, body mass index, and change in size from birth to adulthood to insulin resistance in a female twin cohort. *J Clin Endoctrinol Metab* 2008; **93**: 516-520.

Öberg S, Cnattingius S, Sandin S, Lichtenstein P, Illiadou A. Birth weight-breast cancer revisited: is the association confounded by familial factors? *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 2447-2452.

Öberg S, Ge D, Cnattingius S, Svensson A, Treiber F, Sneidder H, Illiadou A. Ethinc differences in the association of birth weight and blood pressure the Georgia cardiovascular twin study. *Am J Hypertens* 2007; **20**: 1235-1241.

Appendix

A

First note that Figure 1 implies that

$$Y^x \amalg (X_1, X_2) | (U, V). \tag{23}$$

We assume

$$V = \emptyset. \tag{24}$$

We have that

$$\begin{aligned}
&E(Y_j | X_j = x, X_{j'} = x') \\
&\quad = E\{E(Y_j | X_j = x, X_{j'} = x', U) | X_j = x, X_{j'} = x'\} \\
&\quad \overset{(1)}{=} E\{E(Y_j | X_j = x, U) | X_j = x, X_{j'} = x'\} \\
&\quad = E\{E(Y | X = x, U) | X_j = x, X_{j'} = x'\} \\
&\quad \overset{(3),(7),(24)}{=} E\{E(Y^x | U) | X_j = x, X_{j'} = x'\} \\
&\quad \overset{(23)}{=} E\{E(Y^x | X_j = x, X_{j'} = x', U) | X_j = x, X_{j'} = x'\} \\
&\quad = E(Y^x | X_j = x, X_{j'} = x').
\end{aligned}$$

A crucial step is the third equality, where the removal of index j on Y_j and X_j is motivated by symmetry.

B

First note that Figure 1 implies that

$$Y_j \amalg X_{j'} | (U, V_j). \tag{25}$$

We have that

$$\begin{aligned}
&E(Y_j|X_j = x, X_{j'} = x', V_j = v)\\
&\quad = E\{E(Y_j|X_j = x, X_{j'} = x', U, V_j = v)|X_j = x, X_{j'} = x', V_j = v\}\\
&\quad \overset{(25)}{=} E\{E(Y_j|X_j = x, U, V_j = v)|X_j = x, X_{j'} = x', V_j = v\}\\
&\quad = E\{E(Y|X = x, U, V = v)|X_j = x, X_{j'} = x', V_j = v\}\\
&\quad \overset{(3),(7)}{=} E\{E(Y^x|U, V = v)|X_j = x, X_{j'} = x', V_j = v\}.
\end{aligned}$$

The removal of index j on Y_j and X_j in the third equality is motivated by symmetry. Thus, if $V = V_{obs}$, then model (17) implies that

$$\begin{aligned}
\beta_W \Delta x \;=\; & g[E\{E(Y^{x_0+\Delta x}|U, V = v)|X_j = x_0 + \Delta x, X_{j'} = x_0, V_j = v\}]\\
& - g[E\{E(Y^{x_0}|U, V = v)|X_j = x_0, X_{j'} = x_0 + \Delta x, V_j = v\}].
\end{aligned} \tag{26}$$

In the right hand side of (26), $E(Y^{x_0+\Delta x}|U, V = v)$ is averaged over $\Pr(U|X_j = x_0 + \Delta x, X_{j'} = x_0, V_j = v)$, whereas $E(Y^{x_0}|U, V = v)$ is averaged over $\Pr(U|X_j = x_0, X_{j'} = x_0 + \Delta x, V = v)$. If these distributions would be equal, then the contrast in (26) would be comparing 'like with like', and could be given a causal interpretation. Unfortunately, this is typically not the case. To see this, we can use Bayes rule to write

$$\begin{aligned}
\Pr(U = u|X_j, X_{j'}, V_j = v) &= \frac{\Pr(V_j = v|U = u, X_j, X_{j'})\Pr(U = u|X_j, X_{j'})}{E\{\Pr(V_j = v|U, X_j, X_{j'})|X_j, X_{j'}\}}\\
&= \frac{\Pr(V_j = v|U = u, X_j)\Pr(U = u|X_j, X_{j'})}{E\{\Pr(V_j = v|U, X_j)|X_j, X_{j'}\}},
\end{aligned} \tag{27}$$

where the second equality follows because $V_j \amalg X_{j'}|(U, X_j)$ in Figure 1. From (27) it follows that $\Pr(U = u|X_j = x_0 + \Delta x, X_{j'} = x_0, V_j = v) = \Pr(U = u|X_j = x_0, X_{j'} = x_0 + \Delta x, V_j = v)$ if either 1) $V_j \amalg X_j|U$ (i.e. if $\alpha_V = 0$ in (18)), in which case they both equal $\frac{\Pr(V_j=v|U=u)\Pr(U=u|X_1=x_0,X_2=x_0+\Delta x)}{E\{\Pr(V_j=v|U)|X_1=x_0,X_2=x_0+\Delta x\}}$, or if 2) $V_j \amalg U|X_j$ (i.e. if $\alpha_U = 0$ in (18)), in which case they both equal $\Pr(U = u|X_1 = x_0, X_2 = x_0 + \Delta x)$. However, neither of these criteria hold in Figure 1. 1) is violated since V_j has a causal effect on X_j. 2) is violated since both V_j and U have a causal effect on X_j.

C

The model in (17) can, with identity link be expressed as

$$E(Y_j|X_j, X_{j'}, V_j) = \beta_0 + \left(\beta_W + \frac{\beta_B}{2}\right) X_j + \frac{\beta_B}{2} X_{j'} + \beta_{V_j} V_j. \tag{28}$$

Under the data generating mechanism in (18), we have that

$$E(Y_j|X_j, X_{j'}, V_j) = \delta_U E(U|X_j, X_{j'}, V_j) + \delta_V V_j + \delta_X X_j. \tag{29}$$

It follows from standard theory for Normally distributed variables that U can be expressed as

$$U = \psi_{X_j} X_j + \psi_{X_{j'}} X_{j'} + \psi_{V_j} V_j + \epsilon_U, \tag{30}$$

where ϵ_U is standard normal and independent of V_1, V_2, ϵ_{X_1}, ϵ_{X_2}, ϵ_{Y_1}, and ϵ_{Y_2}. Thus,

$$E(U|X_j, X_{j'}, V_j) = \psi_{X_j} X_j + \psi_{X_{j'}} X_{j'} + \psi_{V_j} V_j. \tag{31}$$

Combining (28), (29), and (31) gives that $\beta_W = \delta_X + \delta_U(\psi_{X_j} - \psi_{X_{j'}})$. It thus remains to compute $\psi_{X_j} - \psi_{X_{j'}}$. Multiplying both hand sides of (30) with X_j, $X_{j'}$, and V_j, respectively, and averaging, gives three equations:

$$\begin{aligned}
\text{cov}(X_j, U) &= \psi_{X_j}\text{var}(X_j) + \psi_{X_{j'}}\text{cov}(X_j, X_{j'}) + \psi_{V_j}\text{cov}(X_j, V_j) \\
\text{cov}(X_{j'}, U) &= \psi_{X_j}\text{cov}(X_{j'}, X_j) + \psi_{X_{j'}}\text{var}(X_{j'}) + \psi_{V_j}\text{cov}(X_{j'}, V_j) \\
\text{cov}(V_j, U) &= \psi_{X_j}\text{cov}(V_j, X_j) + \psi_{X_{j'}}\text{cov}(V_j, X_{j'}) + \psi_{V_j}\text{var}(V_j)
\end{aligned} \tag{32}$$

From the data generating mechanism in (18), we have that $\text{cov}(X_j, U) = \alpha_U$, $\text{var}(X_j) = \alpha_U^2 + \alpha_V^2 + 1$, $\text{cov}(X_j, X_{j'}) = \alpha_U^2$, $\text{cov}(X_j, V_j) = \alpha_V$, $\text{cov}(X_{j'}, U) = \text{cov}(X_{j'}, V_j) = \text{cov}(V_j, U) = 0$, and $\text{var}(V_j) = 0$. Solving the system in (32) now gives that

$$\psi_{X_j} - \psi_{X_{j'}} = \frac{\alpha_U \alpha_V^2}{(1 + \alpha_U^2)(1 + \alpha_V^2) + \alpha_U^2}.$$

D

First note that Figure 1 implies that

$$Y_j \amalg X_{j'} | (U, V_1, V_2), \tag{33}$$

and

$$Y^x \amalg (X_1, X_2) | (U, V_1, V_2). \tag{34}$$

For any fixed value v we have that

$$\begin{aligned}
&E(Y_j | X_j = x, X_{j'} = x', V_j = V_{j'} = v) \\
&= E\{E(Y_j | X_j = x, X_{j'} = x', U, V_j = V_{j'} = v) | X_j = x, X_{j'} = x', V_j = V_{j'} = v\} \\
&\overset{(33)}{=} E\{E(Y_j | X_j = x, U, V_j = V_{j'} = v) | X_j = x, X_{j'} = x', V_j = V_{j'} = v\} \\
&= E\{E(Y | X = x, U, V_j = V_{j'} = v) | X_j = x, X_{j'} = x', V_j = V_{j'} = v\} \\
&\overset{(3),(7)}{=} E\{E(Y^x | U, V_j = V_{j'} = v) | X_j = x, X_{j'} = x', V_j = V_{j'} = v\} \\
&\overset{(34)}{=} E\{E(Y^x | X_j = x, X_{j'} = x', U, V_j = V_{j'} = v) | X_j = x, X_{j'} = x', V_j = V_{j'} = v\} \\
&= E(Y^x | X_j = x, X_{j'} = x', V_j = V_{j'} = v).
\end{aligned}$$

A crucial step is the third equality, where the removal of index j on Y_j and X_j is motivated by symmetry. Thus, if $V = V_{obs}$, then model (21) implies that for any fixed value v,

$$\begin{aligned}
\beta_W \Delta x \;\; = \;\; & g\{E(Y^{x_0 + \Delta x} | X_j = x_0 + \Delta x, X_{j'} = x_0, V_j = V_{j'} = v\} \\
& - g\{E(Y^{x_0} | X_j = x_0, X_{j'} = x_0 + \Delta x, V_j = V_{j'} = v\}.
\end{aligned}$$