Invited Article

Using Twins to Assess What Might Have Been: The Co-twin Control Design

Research on Social Work Practice
1–11

DOI: 10.1177/10497315221120604
journals.sagepub.com/home/rsw
SAGE

Wilhelmina van Dijk[1], Cynthia U. Norris[2,3], and Sara A. Hart[2,3]

Abstract

Randomized control trials are considered the pinnacle for causal inference. In many cases, however, randomization of participants in social work research studies is not feasible or ethical. This paper introduces the co-twin control design study as an alternative quasi-experimental design to provide evidence of causal mechanisms when randomization is not possible. This method maximizes the genetic and environmental sameness between twins who are discordant on an "exposure" to provide strong counterfactuals as approximations of causal effects. We describe how the co-twin control design can be used to infer causality and in what type of situations the design might be useful for social work researchers. Finally, we give advantages and limitations to the design, list a set of Twin Registries with data available after application, and provide an example code for data analysis.

Keywords

quasi-experimental designs, twin studies, genetics, environment, co-twin control design

Many social work researchers evaluate and develop interventions, guidance, and public policies to improve the lives of children, adults, and families dealing with difficulties (e.g., Drisko et al., 2020; Greeson et al., 2015; Wilson et al., 2020; Zlotnik & Solt, 2006). It is important that these interventions, guidance, and policies are grounded in research that has determined the causal mechanisms behind these difficulties and assessed the potential of interventions and policies to decrease these difficulties. Randomized control trials (RCTs) are regarded as "the gold standard" for intervention research (Mosteller & Boruch, 2002) and the most robust research design to determine causal effects (Shadish et al., 2002). Before spending substantial resources (i.e., time and money) on designing and implementing a rigorous RCT, however, researchers should determine if an RCT is the correct design for their situation and consider alternative ways to assess the hypothesized causal mechanisms.

RCTs are regularly used in social work research (Thyer, 2015). For example, an RCT was done to examine the effects of an outreach program as compared to traditional intensive foster care services on independent living outcomes of young adults in the foster care system (see Greeson et al., 2015). However, in many instances in social work research, the use of an RCT may not be feasible, or even desirable. Take a researcher interested in exploring if there is a causal relation between chronic child malnutrition and later mathematics achievement. It would be unethical to sample a group of infants and randomly assign them to either a condition where they are malnourished or a condition where they are receiving a balanced diet for a number of years, and then assess differences in math achievement years later. Comparing groups without randomizing participants, however, limits the inferences researchers can make about causality because of potential confounding variables (Shadish et al., 2002). Groups may be different in many other ways besides their status on the hypothesized cause. For example, children who have been chronically malnourished may have grown up in poverty with less developed home mathematics environments, while children provided with a well-balanced diet grew up in a mathematics-rich environment. When comparing these two groups, it would be impossible to say if lower mathematics achievement was due to malnutrition or the home mathematics environment. By randomizing participants, the use of RCTs ensures that these possible confounding factors are randomly distributed

[1]Department of Special Education and Rehabilitation Counseling, Utah State University, Logan, UT, USA
[2]Department of Psychology, Florida State University, Tallahassee, FL, USA
[3]Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA

Corresponding Author:
Wilhelmina van Dijk, Department of Special Education and Rehabilitation Counseling, Utah State University, 2865 Old Main Hill, Logan, UT 84322-2865, USA.
Email: willa.vandijk@usu.edu

across the groups, making the groups, on average, similar to each other (Shadish et al., 2002).

This puts the social work researcher in a predicament. If randomization is unethical and not randomizing may lead to invalid inferences, how can we still make valid and strong causal inferences? Instead of RCTs, social work researchers can use quasi-experimental designs to approximate the causal mechanism when either random assignment or true manipulation of the situation is not possible (Thyer, 2012). Several quasi-experimental designs exist, such as propensity score models and regression discontinuity designs. Here, we will expand on the logic behind a different, less widely known, quasi-experimental design: the co-twin control design (CTCD), or discordant twin pair design (McGue et al., 2010). The CTCD includes twin pairs who are discordant on an independent variable hypothesized to have a causal effect and uses this discrepancy as a natural experiment to examine its influence on an outcome of interest. In a discordant twin pair, the unexposed twin serves as his or her exposed twin's "counterfactual" or control comparison, similar to matching pairs of control and treatment participants. The CTCD is especially strong in its ability to control for possible confounding factors because the twins are matched on both genetic and environmental backgrounds and thus can provide compelling inferences on possible causation. For this reason, CTCD designs are widely used in other fields, such as epigenetics and public health (Baird & Hysi, 2019), and we think they can be of similar value as an alternative to RCTs in social work research.

In this paper, we will introduce the CTCD as a potential methodology to use in social work research. We will first categorize the CTCD among other experimental and quasi-experimental designs by situating them within the counterfactual model. Then, we will explain the logic behind the CTCD and how it can be used to infer causality. We will provide examples of different ways twins can be discordant on an exposure. Finally, we will provide the advantages and limitations of CTCD in social work research.

Counterfactuals

As stated above, researchers can make strong claims about causality in RCTs because the treatment and control groups are considered probabilistically similar to each other. Stated differently, the control group acts as a *counterfactual* for the treatment group. The counterfactual is a proxy for what would have happened to the treatment group if they had not received the treatment. Estimating the counterfactual from a proxy is necessary because we cannot observe the *actual* counterfactual of the treatment group. This idea is explained through the counterfactual model of causation (see, e.g., Rubin, 2004). At the basis of establishing causation is the goal to understand if an individual will be better or worse on a specific outcome, given a cause. For example, we might be interested in establishing the effect of an intervention aimed to reduce anxiety in children who have experienced a traumatic event. To estimate the effect of the intervention on anxiety levels for an individual, researchers would need two *simultaneous* pieces of information: (a) an individual's anxiety level in presence of the intervention and (b) the individual's anxiety level in absence of the intervention. By comparing the difference between the anxiety levels of each individual, the researcher would know the exact effect of the intervention on children's anxiety levels. In real life, however, we can only observe one of the two situations for each individual, not both simultaneously, and we need to use a proxy for the unobserved situation.

Stated more technically, an individual has two *potential outcomes* resulting from a cause: the fact (or that what was observed) and the counterfactual (that what would have been) (see Table 1) (Rubin, 2004). In the case of the anxiety-reducing intervention, a child has the following two potential outcomes: the anxiety level that was observed when they received the intervention (i.e., the fact) and the anxiety level that would have been observed if they did not receive the intervention (i.e., the counterfactual). The effect of a cause can be estimated by taking the difference between the fact and the counterfactual. Because the counterfactual outcome is missing when looking at causality of a treatment for *an individual effect*, causal inference designs often utilize group designs with a treatment group and a control group as a counterfactual to estimate causality for *an average treatment effect*. Participants in the control group provide observed information on the counterfactual condition but have missing information on the fact. In the case of our example, the control group has observed anxiety levels in absence of the anxiety reducing intervention, and

Table 1. Example of Potential Outcomes in the Counterfactual Model for One Participant.

	Anxiety Level After Receiving Anxiety Reducing Intervention	Anxiety Level Not Receiving Anxiety Reducing Intervention
Receiving anxiety reducing intervention	Observed	Missing

Table 2. Example of Potential Outcomes in the Counterfactual Model for More Than One Participant.

	Anxiety Level After Receiving Anxiety Reducing Intervention	Anxiety Level Not Receiving Anxiety Reducing Intervention
Receiving anxiety reducing intervention	Observed	Missing
Not receiving anxiety reducing intervention	Missing	Observed

unobserved levels of anxiety in the presence of the anxiety reducing intervention (see Table 2). If participants in the treatment and control group are the same in every respect except the cause, in this case receiving the anxiety reducing intervention, the average difference between the anxiety level of treatment participants and of control participants is a strong approximation of the true effect of the anxiety reducing intervention on anxiety levels of children who have experienced a traumatic life event.

In RCTs, the control group counts as a strong counterfactual because, as mentioned before, randomly assigning participants to treatment and control groups assures the participants in both groups are equal on expectation since all potential confounding variables are randomly distributed across the two groups (Shadish et al., 2002). Often, however, randomization is not desirable or feasible. In this case, researchers can employ quasi-experimental designs in which the cause is still manipulated, but no random assignment occurs. A consequence of this lack of randomization is that participants in treatment and control groups cannot be considered equal on expectation (Shadish et al., 2002). Different quasi-experimental designs approach ensuring appropriate counterfactuals in distinct ways. Propensity score models, for instance, estimate how prone each individual is to experience the treatment based on their value on a set of covariates selected by the researcher and then match cases with similar propensities (Leite, 2016). The regression discontinuity design takes a completely different approach by considering the regression line as a counterfactual, instead of the participants in the control group. Participants are assigned to treatment based on a cut-off score on an assignment variable, such as a pretest. The difference in regression lines or relation between the assignment variable and the outcome variable, for the treatment group and control group shows the effect of the treatment. If there is no effect, then there should be no difference in the intercept or slope of the line (Shadish et al., 2002). Both these quasi-experimental methods are valuable options in a methodological toolkit, but still, have limitations when it comes to providing counterfactuals. The CTCD design is an interesting alternative approach to provide a counterfactual by employing the natural design inherent to twins.

The CTCD

Before getting into the details of the CTCD, it will be worthwhile to explain the basics behind the classical twin design. Researchers use twin samples to partition the variance in a trait into genetic and environmental sources of variance by comparing the similarity between monozygotic (MZ or identical twins) and dizygotic (DZ or nonidentical) twins. Identical twins share all their genes and nonidentical twins share, on average, half of their genes. Both types of twins share aspects of their environment such as home and school (i.e., the "shared environment"), and also both types of twins experience some aspects of the environment independently, such as illness or separate sets of friends (i.e., the "non-shared environment"). Variance attributable to additive genetic influences (or heritability) is assumed when MZ twins are more similar in their reading scores than DZ twins, and shared environmental influences (nongenetic influences that make siblings more alike), are assumed when MZ twins are less than two times as similar as DZ twins, and non-shared environmental influences (those that decrease similarities between twins; plus error) are assumed when the correlation between MZ twins is less than one (see Hart et al., 2020 for a visual representation; also see Knopik et al., 2017 for more details on the classical twin design methodology).

Classical twin designs partition genetic and environmental influences to estimate which of those is mainly responsible for differences in outcomes. For example, Scaini et al. (2021) estimated that 40% of the variability in childhood anger in a sample of Italian twins was due to their genes, 25% to their shared environment, and 35% to their non-shared environment. These results suggest that differences in twins' environment are a substantial driver of their anger. The CTCD takes a different approach, instead using a counterfactual model to infer causality by comparing one member of a twin pair that has an environmental exposure to the other twin who does not have this exposure to an outcome. This approach sets up discordant twin pairs, and the nonexposed twin serves as the counterfactual to the exposed twin, serving what the exposure twin would have looked like had they not been exposed. Coming back to our earlier example of anxiety reducing intervention for children who have experienced a traumatic event, one of the twins would receive the intervention and serve as the fact, while the second twin would not receive the intervention and serve as the counterfactual. As twins share all, or at least some, of their additive genetic influences and all of their shared environmental influences, the nonexposed twin member serves a genetic and shared environmental control.

In conventional causal modeling research, individuals' exposure to the intervention and the subsequent outcome measurement could be due to a direct causal relation or possibly genetic and/or environmental confounding (Hart et al., 2021). Conversely, in the CTCD design, the environmental exposure solely is hypothesized as a causal mechanism, and the genetic and shared environmental similarity within a twin pair is used to control for genetic and shared environmental confounding outside of that exposure. As a note, this design cannot control for non-shared environmental confounding, as this is by nature not shared within a twin pair.

Exposure Types in the CTCD

The critical feature of the CTCD is that twins are discordant on the exposure. Exposures can be classified under three distinct types: experimental interventions, rare happenings, or natural events (Segal, 2019). In experimental interventions,

one twin is subjected to an intervention, whereas the other twin does not receive the intervention. NASA has employed the experimental design to understand how prolonged spaceflight impacts human physiology (Garrett-Bakelman et al., 2019). In their study, NASA sent one twin astronaut to space for 340 days while the identical twin stayed on earth. Twins were compared pre- and postspaceflight on various outcomes, including several cognitive skills. One of the more notable outcomes was that the space twin decreased accuracy and precision in the cognitive tasks postspaceflight as compared to his earth twin (Garrett-Bakelman et al., 2019). In this study, the earth twin represented what would have happened to the space twin had he not spent a year in space.

It is also possible to use the occurrence of rare events in CTCD. In keeping with our earlier example of the effect of childhood trauma on later anxiety, one twin might have experienced trauma early in childhood, such as being in a car accident, whereas the other twin did not experience this. With exception of this event, the twins are considered similar in all other aspects. A researcher could then check the differences in anxiety levels between the twins in a number of twin pairs to get an estimate of the influence of trauma on later anxiety levels. Using data from Vietnam veterans, Koenen et al. (2003) used twin discordance in exposure to combat and combat-related PTSD to assess if this caused depression and dependency disorders after controlling for shared familiar vulnerability. The results from the CTCD showed twins with higher combat exposure and combat-related PTSD had more risk for later mental illness and substance abuse disorders, even when the genetic risks were taken into account.

The last type of discordance that can be investigated with CTCD is due to natural occurrences. Differences in the behavior and health of twins are considered natural events if they are not a result of interventions. For example, CTCD has been used to assess the health consequences of smoking tobacco (Carmelli & Page, 1996) and the effects of drinking alcohol on cognitive development when aging (McGue et al., 2010). In psychological sciences, researchers have used the design to explore the effect of teacher quality on student reading performance (Hart et al., 2013) and the effect of birthweight on ADHD symptoms (Lim et al., 2018). These examples highlight the meaningful ways CTCD can be used to answer research questions. The use of any of these three types of exposure will result in a particularly strong quasi-experimental design that may lead to new insights on risk factors of negative life outcomes and implications for interventions and public policy.

Determining Causality With CTCD

Causality in the CTCD is determined by checking if the association between the exposure, or hypothesized cause, and the outcome is similar or different across three groups: all individuals in the sample, MZ twins, and same-sex DZ twins (see e.g., Bergen et al., 2008; Hart et al., 2013; McGue et al., 2010). First, the relation between exposure and outcome is assessed across all participants in the sample, disregarding their twin status. This examination will show if a relation exists, to begin with. Then the same relation between exposure and outcome is examined within MZ twins. The resulting estimate represents the association between the exposure and the outcome without possible confounding of the effect of genetic and shared environmental influences since the MZ twins share all of their additive genetic and shared environmental influences. Therefore, if the effect is still present in discordant MZ pairs, the causal claim is much stronger. Finally, the same relation can be tested in DZ twins (preferably same-sex DZ pairs for the strongest control). The resulting estimate represents the possible relation between exposure and outcome controlling for shared environmental influences and some genetic influences (since DZ twins share about half of their additive genetic influences).

The causal relation from exposure to outcome is established by estimating different regression models commensurate with the three tests just described. First, to estimate the individual, or sample, effect, researchers need to estimate a linear regression using all twins without considering if they come from the same family. For this model let Y_{ij} be the observed outcome for the ith twin ($i = 1, 2$) in the jth twin pair ($j = 1, 2, \ldots, N$) and x_{ij} the corresponding value on the exposure for each individual twin. In general, exposure in the CTCD is conceptualized as a dichotomous variable (0 = not exposed, and 1 = exposed). In cases where researchers are interested in the effect of a continuous variable (e.g., birthweight, alcohol use, anxiety) on an outcome, they will have to make a decision as to what the cut-off for exposure is, or what constitutes a difference great enough to be considered discordant. The individual regression model is then represented as follows

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij} \quad (1)$$

where β_1 is the individual level effect of the exposure on the outcome, β_0 represents the intercept, and ε_{ij} is the residual, or error term, which is correlated within each twin pair in this model. Imagine a study examining the relation between being bullied and exhibiting externalizing problem behaviors in middle school. Y_{ij} would represent each twin's value on an externalizing behavior measure, and x_{ij} if they were a victim of bullying. The value of β_1 is the estimate of the effect of being bullied on later externalizing problems in the sample. If this estimate is statistically significant, researchers can then move to control for possible genetic and environmental confounding by modeling the exposure to outcome relation taking the twin pairs into account.

The twin models are performed within the hierarchical linear modeling framework. Within this framework,

individual twins are considered nested within a twin pair, making each individual twin a level 1 variable and the twin pairs the level 2 variables. This leads to the following model

$$Y_{ij} = \beta_0 + \beta_W(x_{ij} - \bar{x}_j) + \beta_B\bar{x}_j + \upsilon_j + \varepsilon_{ij} \qquad (2)$$

where β_W represents the within twin pair effect, β_B represents the between twin pair effect, $\bar{x}_j$ represents the mean value on the exposure of the jth twin pair, υ_j is the level 2 variance, and ε_{ij} is the individual residual, or error term, which is correlated within each twin pair. In our imagined study, β_W is the direct estimate of the effect of being bullied on externalizing behavior within each twin pair, thus accounting for genetic and environmental confounds. To further probe the influence of genes of the effect of bullying on externalizing problem behavior, researchers can disaggregate the within twin pair parameter estimate by zygosity status. This disaggregation involves adding zygosity as a covariate and an interaction term. In addition, any nonshared environmental factors that may influence the relation should be added as control variables in the models to reduce the possibility of confounding the effect (see e.g., Koenen et al., 2003). The final model serves as a way to estimate the differences in marginal means, or expected outcomes, for each of the groups (i.e., all individuals in the sample, discordant MZ twins, and discordant DZ twins). These differences in marginal means are the basis for the comparisons to determine causality. Table 3 provide the example code that includes the CTCD model and estimation of marginal means for both SAS and *R*.

Comparing the association between the exposure and the outcome across these three groups (i.e., all individuals in the sample, discordant MZ twins, and discordant DZ twins) based on the difference in marginal means leads to four outcome scenarios: (a) there is a possible causal relation between the exposure and the outcome, (b) there is a possible confounding effect caused by the shared environment, (c) there is a possible confounding effect of caused by genes, and (d) there is a possible confounding effect of both environmental and genetic nature. These outcome scenarios are represented in Figure 1A to D.

Only scenario A provides evidence of a causal relation between an exposure and an outcome; all bars representing the estimates of the difference in marginal means should be of (near) equal height with overlapping error bars (see Figure 1A). In other words, the association between exposure and outcome is seen even when controlling for genetic and environmental influences within twin pairs, indicating neither of these influences mattered for the relation. For example, if the relation between being bullied and externalizing problem behaviors is completely causal, the association should not be different in the sample, discordant MZ twins, or discordant DZ twins.

Scenarios B–D are all noncausal. Figure 1B illustrates the second outcome scenario where shared environmental factors confound the causal relation. The association estimated for both types of twins should be similar to each other, but lower than the sample estimate. In Figure 1B, the difference in marginal means for the sample is about 10, whereas the difference in marginal means for both twin pairs (i.e., discordant MZ and DZ twins) is about 8. Additionally, while the discordant twin pair groups have overlapping error bars, neither overlap with the sample error bar. This divergence between the difference in marginal means for the sample and discordant twin pairs is a result of the same shared environment of twins, for example, their family environment, and these shared environmental influences are controlled for when examining within twin pairs. In the case of the relation between being bullied and exhibiting externalizing behaviors, we would conclude that there are additional triggers in the shared environment of children that lead to higher levels of externalizing behaviors, but that bullying is not the cause of the externalizing behaviors.

The third hypothetical outcome scenario is also noncausal and involves genetic influences confounding a potential association (see Figure 1C). In this case, the difference in marginal means will diverge across all three groups. In the sample, it may seem as if an association exists, because the individuals in the sample are considered unrelated and genetic influences are, therefore, not being controlled for. In the figure, this is represented by high differences in marginal means. The association should be almost nonexistent for discordant MZ twins because they share all of their genetic influences, and the genetic confounds are thus controlled for when examining this group. The differences in marginal means are near zero in this scenario. Finally, discordant DZ twin pairs should fall somewhere in the middle of the two estimates, since they share about half of their genetic influences. While the outcomes for the sample still suggest an association between being bullied and externalizing problem behavior, the fact that this association is much lower in DZ twins and almost absent in MZ twins would suggest bullying has no influence on externalizing behavior problems. In fact, the results in this scenario suggest externalizing behavior problems are mostly genetic in nature.

The final hypothetical outcome scenario is also a noncausal scenario where both genetic and shared environmental factors are confounding a potential association. In this case, the association is only present in the sample but does not show up for either of the twin types. This scenario is represented in Figure 1D. The differences in marginal means for the sample are still high and suggest an association exists between the exposure and the outcomes. For both twin pairs, however, there is no relation. The difference in marginal means for both groups is near zero. This scenario implies that while it looks like an association between bullying and externalizing behaviors exist based on the sample data, this association is in fact confounded by both genetic and shared environmental causes. Children display higher levels of externalizing behavior

Table 3. Example Code to Run co-twin Control Designs.

Platform	Code	Comments
SAS	Title "WithinBetweenInteraction model";	This model estimates the effect of being bullied (bul) on externalizing behavior problems (ExtBeh)
	proc mixed data=YourData method=ml covtest noclprint empirical method;	Use ML estimator with empirical method to get sandwich estimator, covtest makes sure we get asymptotic standard errors, noclprint surpresses any class level information
	class twinid bul zyg;	The model includes the effect of zygosity and its interaction with the exposure (bul) needed to estimate the marginal means. ddfm residual indicates SAS will use the residual degrees of freedom.
	model ExtBeh = bul zyg bul* zyg/solution ddfm=residual;	
	random intercept/subject = twinid type = un;	This is a random intercept model, with an unstructured covariance matrix that is useful got correlated models
	lsmeans bul* zyg bul/diff;	We ask for the least-square means (marginal means) for different groups based on exposure (bul) and zygosity (zyg) and based on exposure only (bul), and ask for the differences in the marginal means.
	run;	
R	## WithinBetweenInteraction model##	Comment on code
	library(lme4) library(lmerTest)	The installed R packages to fit model and give statistical test. Run packages in this order.
	model <-lmer(ExtBeh ~ bul + zyg + bul* zyg + (1|twinid), data=YourData, REML=FALSE)	The model indicates the effect of zygosity and its interaction with exposure(bul), the "(1|twinid)" specifies random intercept ("1" means intercept and "|" means correlated), and a command not to use default REML. This model will be used to estimate marginal means.
	summary(model)	Asks for model results
	anova(model)	Asks for random effects *p*-values
	confint(model)	Ask for confidence intervals for random effect
	library(lsmeans)	Installed package for lsmeans function (emmeans package can also be used to get least-square means)
	lsmeans(model, pairwise ~ bul* zyg)	Asking for the least-square means based on exposure (bul) and zygosity only, this provides contrasts with differences in marginal means.
	lsmeans(model, pairwise ~ bul)	Asking for least-square means based on exposure only (bul)

problems because of their genes and how they grew up, not because they were bullied.

Advantages and Limitations of the CTCD

Compared to other quasi-experimental designs, the CTCD has one main advantage as alternative to experimental designs: the ability to control for genetic and shared environmental confounds. This control leads to the provision of a near-perfect counterfactual. Almost any relation between an exposure and an outcome might be confounded by genetic or shared environmental influences. For example, the home math environment is often considered related to kindergarteners' mathematics skills (2016). The home math environment, which are math-related activities that parents do in the home with their children such as counting objects and cooking, could be considered a natural exposure that can lead to better mathematics outcomes. It is possible this relation is due to a causal mechanism between home numeracy activities and math outcomes in school; however, it is also possible there is a common environmental factor, such as having parents with higher educational backgrounds (Thompson

et al., 2017). Alternatively, the lower mathematics skills may be a function of other underlying problems, such as processing speed (Willcutt et al., 2013). A combination between genetics and the environment might also be at play. For example, students genetically at risk for lower mathematics skills have worse outcomes the more chaotic their home environment (Docherty et al., 2011). Besides the randomized control trial, where these genetic and environmental confounders are expected to be equally distributed across the treatment and control groups, only the CTCD and several other twin design studies can rule out genetic or shared environmental factors as possible confounders (Hart et al., 2021).

It may be clear that CTCD's greatest advantage is the use of a powerful counterfactual that can be used to assess causation while accounting for genetic and environmental influences without the need for experimental manipulation or randomization. As mentioned above, this characteristic of the design does not preclude the use of experimental manipulation. For example, to better understand the effect of play therapy on internalizing behaviors (e.g., Drisko et al., 2020) researchers might randomly assign twins within each twin pair to either a play therapy condition or a control condition. Combining experimental manipulation with a twin design increases the power to detect effects while requiring fewer participants. This happens because much of the unexplained variance in scores present for a random sample of participants, possibly due to genetic and environmental factors, is not expected to exist in the scores of twin pairs. Researchers are left with a highly powerful model capable of providing strong evidence for causation.

Collecting data from twins may seem like a daunting task. After all, twins are not ubiquitous and finding, recruiting, and surveying them can take a long time. Fortunately, many other researchers have started doing this since 1954, when the Danish Twin Registry was established (Hur et al., 2019) and researchers interested in using the CTCD can capitalize on these registries. Currently, there are twin registries all over the world. A recent special issue in *Twin Research and Human Genetics* (Hur et al., 2019) contains information on more than 60 such registries. Several registries contain contact information on twins that can be used to recruit twins for new studies (e.g., Mid-Atlantic Twin Registry of Virginia Commonwealth University, Lilley et al., 2019). Beyond new data collection, researchers can take advantage of existing twin registries by using already collected data. Many registries contain a wealth of variables, often across multiple years, on behavioral, social, and academic characteristics of the twins themselves and many also include info on their families, teachers, and environments (Hur et al., 2019). Because of the wide range of scope of contextual and outcome measures within these registries, researchers are likely to find a data set that includes the constructs of their interest. Many national and international twin study registries have publicly available data or at least will make data available upon request. Table 4 has an overview of twin registries with data that is available through an online application process. There are also registries that may not have an online application process, but generally, because of a long tradition of data sharing within the twin research community, principal investigators are often highly receptive to sharing data or collaborating on new projects. Examples of such registries include the FinnTwin16 (Kaidesoja et al., 2019), the Murcia Twin Registry (Ordoñana et al., 2019), the Florida State Twin Registry (Taylor et al., 2019), the National Project on Achievement of Twins (Hart et al., 2019), and the Project Talent Twin and Sibling Study (Prescott et al., 2019), amongst others. Using extant data to answer research questions using the CTCD is advantageous because it can save researchers and participants valuable time and monetary resources.

Conversely, the use of extant data from twin registries can also be seen as a limitation to the CTCD design. This is because the researcher is limited to the variables and collection methods used by the original investigators of these registries. It is possible the construct of interest is not directly measured, but the data set may contain several proxies of the construct. In some cases, this may constitute a threat to construct validity in a study (Huggins-Manley et al., 2019). A second possible consequence of using extant data is a lack of control over the timing of the data collection. When selecting variables, it is very important to ensure that the hypothesized cause was measured before the outcome in order to claim causation. In some relations, however, a reverse causation may be at work. That is, the outcome construct may have influenced the cause construct at an earlier time. The CTCD cannot distinguish between this reverse causation (Bergen et al., 2008; McGue et al., 2010). Using data from longitudinal twin studies can help to rule out the possibility of reverse causation.

Besides reverse causation, as briefly mentioned earlier, CTCDs cannot control for possible confounding non-shared environmental influences (Bergen et al., 2008; McGue et al., 2010). For example, if one twin received play therapy and reduced internalizing behaviors at a higher rate than their twin, but this twin was also bullied while not receiving play therapy and consequently had higher rates of internalizing behaviors, the effects of the treatment are confounded with this unshared life event. As a consequence, our perception of the causal mechanism of play therapy on internalizing behaviors would be muddled.

Finally, if researchers use the CTCD as a natural experiment, without experimental manipulation, it is impossible to estimate the effect of environmental exposures that are shared between twins on outcomes. The essential ingredient of the CTCD is that twins are discordant on the exposure. The diet in a household, for example, is unlikely to be different for twins during childhood and therefore cannot be explored as a hypothesized cause for later body weight problems with a CTCD. Nonetheless, these few restrictions on the types of researcher questions that can be answered using the CTCD do not detract from their utility in social work research.

Table 4. Twin Registries With Data on Behavioral, Environmental, and Social Aspects of Life Available by Applications.

Registry	Country	Brief Description	Data Availability
Danish Twin Registry	Denmark	Data from 175,000 Danish twins, starting with twins born in 1870. Datasets include information on socio-demographics, health outcomes, cognitive performance, psychological well-being, personality, and lifestyle.	All data is available through an application process online (fees for data transfer apply) https://www.sdu.dk/en/om_sdu/institutter_centre/ist_sundhedstjenesteforsk/centre/dtr/researcher/guidelines A subset of the data is also available through ICPSR (21041) https://www.icpsr.umich.edu/web/NACDA/studies/21041
German Twin Family Panel	Germany	Data from more than 4,097 German families with twins. Data collection focused on the emergence and development of social inequality. Datasets include information on educational achievement, cognitive development, personality, social, cultural and political life, physical and psychological health, behavior disorders.	All data is available online through an application process https://paneldata.org/twinlife
Michigan State University Twin Registry	USA	Data from more than 30,000 US twins. Datasets include information on behavioral disorders, psychological well-being, home environment, substance use, and academic performance.	All data is available online through an application process https://msutwinstudies.com/msutr-data The registry can also be used as recruitment for new studies (fees apply).
NAS-NRC Twin Registry	USA	Data from 16,000 US twins. All twins were white males participating in military service in World War II. Datasets include information on demographics, service records, substance use, health outcomes, and home environments.	Data is available through application process at ICPSR (36234). https://www.icpsr.umich.edu/web/NACDA/studies/36234
Netherlands Twin Register	Netherlands	Data from 122,652 Dutch twins and their families. Data includes information behavior disorders, development, health outcomes, and psychological well-being.	Data is available through an application process https://tweelingenregister.vu.nl/information_for_researchers/working-with-ntr-data
Norwegian Twin Registry	Norway	Data from 47,989 Norwegian twins, starting with twins born in 1895. Datasets include information on socio-demographics, health outcomes, neurological disorders, and psychological well-being.	Data is available through an application process https://www.fhi.no/en/more/health-studies/norwegian-twin-registry/
Québec Study of Newborn Twins	Canada	Data from 662 Canadian families with twins. Data include information on physical development, cognitive achievement, behavioral disorders, socio-demographics, and the home environment.	Data is available through an application process https://www.maelstrom-research.org/mica/individual-study/ejnq
SpeADy Twin Family Study	Germany	Data from 573 German families with twins. Data includes information on personality, social, cultural and political life.	Data is available through an application process
Swedish Twin Registry	Sweden	Data from 87,000 Swedish twin pairs. Datasets include information on family life, health outcomes, psychological well-being, behavior disorders, and psychopathology.	Data is available through an application process (fees may apply) https://ki.se/en/research/swedish-twin-registry-for-researchers A subset of the data is also publicly available through ICPSR (3843) https://www.icpsr.umich.edu/web/NACDA/studies/3843
Twin Research Australia	Australia	Data from 1,500 Australian twin pairs. Datasets include information on education, health outcomes, lifestyles, socioeconomics, psychological well-being	Data is available through an application process https://www.twins.org.au/research/research-with-us/81-how-to-work-with-us Registry can also be used as recruitment for new studies.
Twins Early	Britain	Data from 10,000 British twin pairs. Datasets	Data is available through an application process

(continued)

Table 4. (continued)

Registry	Country	Brief Description	Data Availability
Development study		contain information on the home and school environment, academic achievement, cognitive development, psychological well-being, personality, and physical health outcomes.	http://www.teds.ac.uk/researchers/teds-data-access-policy
Washington State Twin Registry	USA	Data from 9,771 US twin pairs. Datasets include information on physical health outcomes, psychological well-being, stress, and personality.	Data is available through an application process https://wstwinregistry.org/for-researchers/policies-procedures-for-accessing-the-wstr/

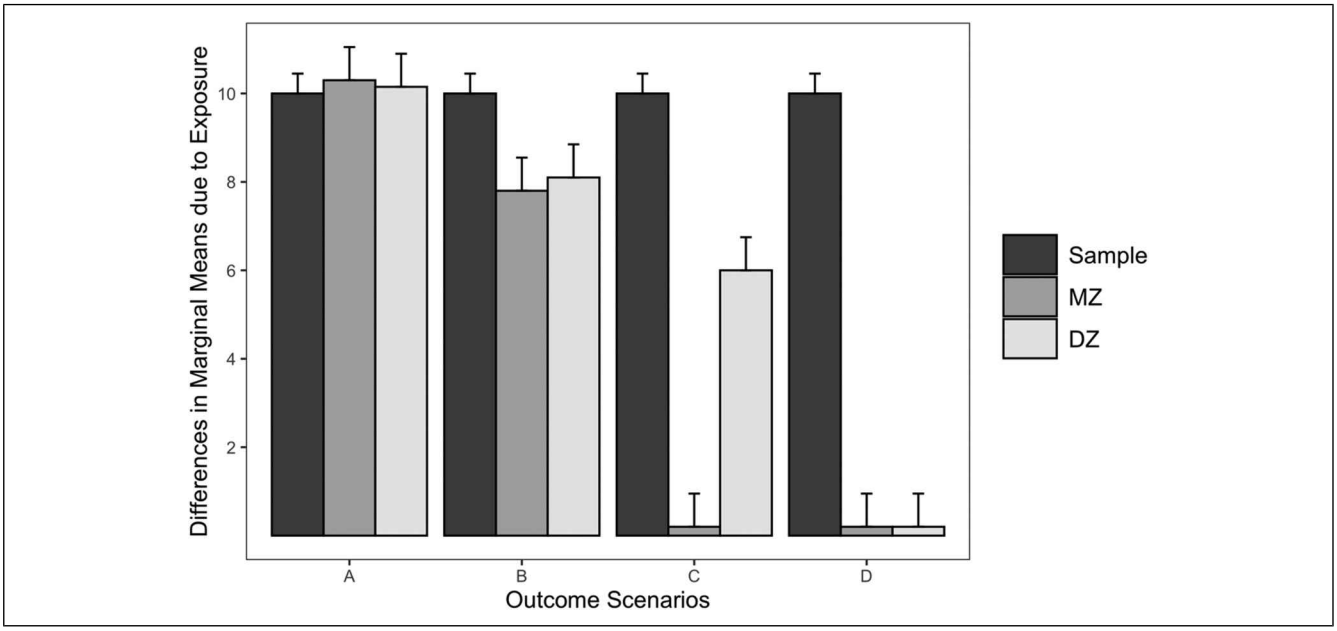

Figure 1. Hypothetical outcome scenarios of a Co-twin control design study.
Note. Sample = all individuals in the sample; MZ = monozygotic; DZ = dizygotic. A represents a possible causal effect, represented by a similar effect across individuals no matter their relatedness. B represents the possibility of shared environmental confounding, with the effect in both twin pairs lower than in the sample. C represents the genetic confounding, with the effect dissipating as subjects become more related. D represents the genetic and shared environmental confounding with no effect for twins, while the effect remains in the sample. Outcome scenarios B, C, and D do not provide evidence of a causal relation between the exposure and the outcome

Conclusion

Many causal mechanisms that are of interest to social work researchers may not be feasible or ethical to examine within experimental RCTs, and finding quasi-experimental designs with counterfactuals able to provide compelling evidence of causality can also be challenging. In such situations, the CTCD, with the added benefit of using extant data from twin registries and the ability to eliminate genetic and environmental confounds, is a viable and attractive option for social work researchers wanting preliminary evidence for a causal inference.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Cynthia Norris is supported by the McKnight Fellowship at the Florida State University and the FIREFLIES training fellowship funded by the Institute of Education Sciences (grant nos. P50HD052120, R01HD072286, R305B200020).

ORCID iD

Wilhelmina van Dijk https://orcid.org/0000-0001-9195-8772

References

Baird, P. N., & Hysi, P. (2019). Twin registries moving forward and meeting the future: A review. *Twin Research and Human Genetics*, *22*(4), 201–209. https://doi.org/10.1017/thg.2019.53

Bergen, S. E., Gardner, C. O., Aggen, S. H., & Kendler, K. S. (2008). Socioeconomic status and social support following illicit drug use: Causal pathways or common liability? *Twin Research and Human Genetics*, *11*(3), 266–274. https://doi.org/10.1375/twin.11.3.266

Carmelli, D., & Page, W. F. (1996). Twenty-four year mortality in World War II US male veteran twins discordant for cigarette smoking. *International Journal of Epidemiology*, *25*(3), 554–559. https://doi.org/10.1093/ije/25.3.554

Docherty, S. J., Kovas, Y., & Plomin, R. (2011). Gene-environment interaction in the etiology of mathematical ability using SNP sets. *Behavior Genetics*, *41*(1), 141–154. https://doi.org/10.1007/s10519-010-9405-6

Drisko, J., Corvino, P., Kelly, L., & Nielson, J. (2020). Is individual child play therapy effective? *Research on Social Work Practice*, *30*(7), 715–723. https://doi.org/10.1177/1049731519854157

Garrett-Bakelman, F. E., Darshi, M., Green, S. J., Gur, R. C., Lin, L., Macias, B. R., McKenna, M. J., Meydan, C., Mishra, T., Nasrini, J., et al. (2019). The NASA twins study: A multidimensional analysis of a year-long human spaceflight. *Science*, *364*(6436), eaau8650. https://doi.org/10.1126/science.aau8650

Greeson, J. K. P., Garcia, A. R., Kim, M., & Courtney, M. E. (2015). Foster youth and social support: The first RCT of independent living services. *Research on Social Work Practice*, *25*(3), 349–357. https://doi.org/10.1177/1049731514534900

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One*, *11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

Hart, S. A., Little, C., & van Bergen, E. (2021). Nurture might be nature: Cautionary tales and proposed solutions. *Npj Science of Learning*, *6*(1), 1–12. https://doi.org/10.1038/s41539-020-00079-z

Hart, S. A., Little, C. W., & van Bergen, E. (2020). *Logic behind twin research*. https://doi.org/10.6084/m9.figshare.12751004.v1

Hart, S. A., Martinez, K., Kennedy, P. C., Ganley, C. M., & Taylor, J. (2019). The national project on achievement in twins. *Twin Research and Human Genetics*, *22*(6), 761–764. https://doi.org/10.1017/thg.2019.86

Hart, S. A., Taylor, J., & Schatschneider, C. (2013). There is a world outside of experimental designs: Using twins to investigate causation. *Assessment for Effective Intervention*, *38*(2), 117–126. https://doi.org/10.1177/1534508412451490

Huggins-Manley, A. C., Beal, C. R., D'Mello, S. K., Leite, W. L., Cetin-Berber, D. D., Kim, D., & McNamara, D. S. (2019). A commentary on construct validity when using operational virtual learning environment data in effectiveness studies. *Journal of Research on Educational Effectiveness*, *12*(4), 750–759. https://doi.org/10.1080/19345747.2019.1639869

Hur, Y.-M., Bogl, L. H., Ordoñana, J. R., Taylor, J., Hart, S. A., Tuvblad, C., Ystrom, E., Dalgård, C., Skytthe, A., & Willemsen, G. (Eds.). (2019). Twin family registries worldwide: An important resource for scientific research. *Twin Research and Human Genetics*, *22*(6), 427–437. https://doi.org/10.1017/thg.2019.121

Kaidesoja, M., Aaltonen, S., Bogl, L. H., Heikkilä, K., Kaartinen, S., Kujala, U. M., Kärkkäinen, U., Masip, G., Mustelin, L., Palviainen, T., et al. (2019). Finntwin16: A longitudinal study from age 16 of a population-based Finnish twin cohort. *Twin Research and Human Genetics*, *22*(6), 530–539. https://doi.org/10.1017/thg.2019.106

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics*. Worth Publishers, Macmillan Learning.

Koenen, K. C., Lyons, M. J., Goldberg, J., Simpson, J., Williams, W. M., Toomey, R., Eisen, S. A., True, W., & Tsuang, M. T. (2003). Co-twin control study of relationships among combat exposure, combat-related PTSD, and other mental disorders. *Journal of Traumatic Stress*, *16*(5), 433–438. https://doi.org/10.1023/A:1025786925483

Leite, W. (2016). *Practical propensity score methods using R*. Sage Publications.

Lilley, E. C. H., Morris, A.-T., & Silberg, J. L. (2019). The mid-Atlantic twin registry of Virginia Commonwealth University. *Twin Research and Human Genetics*, *22*(6), 753–756. https://doi.org/10.1017/thg.2019.87

Lim, K. X., Liu, C.-Y., Schoeler, T., Cecil, C. A. M., Barker, E. D., Viding, E., Greven, C. U., & Pingault, J.-B. (2018). The role of birth weight on the causal pathway to child and adolescent ADHD symptomatology: A population-based twin differences longitudinal design. *Journal of Child Psychology and Psychiatry*, *59*(10), 1036–1043. https://doi.org/10.1111/jcpp.12949

McGue, M., Osler, M., & Christensen, K. (2010). Causal inference and observational research: The utility of twins. *Perspectives on Psychological Science*, *5*(5), 546–556. https://doi.org/10.1177/1745691610383511

Mosteller, F., & Boruch, R. F. (2002). *Evidence matters: Randomized trials in education research*. Brookings Institution Press.

Ordoñana, J. R., Carrillo, E., Colodro-Conde, L., García-Palomo, F. J., González-Javier, F., Madrid-Valero, J. J., Martínez Selva, J. M., Monteagudo, O., Morosoli, J. J., Pérez-Riquelme, F., et al. (2019). An update of twin research in Spain: The Murcia twin registry. *Twin Research and Human Genetics*, *22*(6), 667–671. https://doi.org/10.1017/thg.2019.60

Prescott, C. A., Walters, E. E., Arpawong, T. E., Zavala, C., Gruenewald, T. L., & Gatz, M. (2019). The project talent twin and sibling study: Zygosity and new data collection. *Twin Research and Human Genetics*, *22*(6), 769–778. https://doi.org/10.1017/thg.2019.117

Rubin, D. B. (2004). Teaching statistical inference for causal effects in experiments and observational studies. *Journal of Educational and Behavioral Statistics*, *29*(3), 343–367. https://doi.org/10.3102/10769986029003343

Scaini, S., Centorame, G., Lissandrello, F., Sardella, S., Stazi, M. A., Fagnani, C., Brombin, C., & Battaglia, M. (2021). The role of genetic and environmental factors in covariation between anxiety and anger in childhood. *European Child & Adolescent Psychiatry*, *30*(4), 607–617. https://doi.org/10.1007/s00787-020-01543-2

Segal, N. L. (2019). Co-twin control studies: Natural events, experimental interventions and rare happenings/twin research: Cancer risk in

overweight twins; prognosis after fetal loss of one twin; twin concordance for Parkinson's disease; neuroanatomy of musically discordant MZ twins/news articles: Twin birth with two wombs; twins' prenatal interactions; switched-at-birth twins; fetus-in-fetu; unsolved paternity. *Twin Research and Human Genetics*, *22*(4), 272–276. https://doi.org/10.1017/thg.2019.29

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Taylor, J., Martinez, K., & Hart, S. A. (2019). The Florida state twin registry. *Twin Research and Human Genetics*, *22*(6), 728–730. https://doi.org/10.1017/thg.2019.102

Thompson, R. J., Napoli, A. R., & Purpura, D. J. (2017). Age-related differences in the relation between the home numeracy environment and numeracy skills. *Infant and Child Development*, *26*(5), e2019. https://doi.org/10.1002/icd.2019

Thyer, B. A. (2012). *Quasi-experimental research designs*. Oxford University Press.

Thyer, B. A. (2015). A bibliography of randomized controlled experiments in social work (1949–2013): Solvitur Ambulando. *Research on Social Work Practice*, *25*(7), 753–793. https://doi.org/10.1177/1049731515599174

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *Journal of Learning Disabilities*, *46*(6), 500–516. https://doi.org/10.1177/0022219413477476

Wilson, A. B., Mahadevan, T., Villodas, M., Rodriguez, M., Bailliard, A., & Cuddeback, G. (2020). Tiny homes are huge for people living with serious mental illnesses. *Research on Social Work Practice*, 1049731520961448. https://doi.org/10.1177/1049731520961448

Zlotnik, J. L., & Solt, B. E. (2006). The institute for the advancement of social work research: Working to increase our practice and policy evidence base. *Research on Social Work Practice*, *16*(5), 534–539. https://doi.org/10.1177/1087057106289725