Addressing Contamination Bias in Child Maltreatment Research: Innovative Methods for Enhancing the Accuracy of Causal Estimates

Chad E. Shenk, Ph.D.[1,2,*] (ces140@psu.edu)

Anneke E. Olson, M.S.[1] (aeo442@psu.edu)

Emily Dunning, B.A.[1] (ebd5322@psu.edu)

Kenneth A. Shores, Ph.D.[3] (kshores@udel.edu)

Nilam Ram, Ph.D.[4,5] (nilamram@stanford.edu)

Zachary F. Fisher, Ph.D.[1] (fisherz@psu.edu)

John F. Felt, Ph.D.[6] (jzf434@psu.edu)

Ulziimaa Chimed-Ochir, M.S.[1] (uxc17@psu.edu)

[1] = Department of Human Development and Family Studies, The Pennsylvania State University; [2] = Department of Pediatrics, The Pennsylvania State University College of Medicine; [3] = School of Education, University of Delaware; [4] = Department of Psychology, Stanford University; [5] = Department of Communications, Stanford University; [6] = Center for Healthy Aging, The Pennsylvania State University

Corresponding Author Information:

Chad Shenk, Ph.D.

115 Health and Human Development Building

University Park, PA 16802

Email: ces140@psu.edu

Biographical Statements

Chad Shenk, Ph.D.

Chad Shenk is an Associate Professor in the Department of Human Development & Family Studies and the Department of Pediatrics at Penn State. He is also a licensed clinical psychologist with specialty training in trauma exposure and pediatric psychology. Dr. Shenk's basic science research improves methods for risk estimation and target identification in prospective cohort studies of child trauma and adverse health across the lifespan. This work identifies biomarkers and putative mechanisms of adverse health conditions in the child trauma population using a multiple levels of analysis approach (e.g., biological, behavioral, environmental). His clinical trials and translational research therefore center on the optimization of behavioral interventions following exposure to child trauma by engaging identified targets and mechanisms more effectively. As Principal Investigator, Dr. Shenk's research has been funded by multiple institutes at the National Institutes of Health, the National Science Foundation, the American Psychological Association, as well as several Universities and Foundations.

Anneke E. Olson, M.S.

Anneke Olson is a doctoral student in the Department of Human Development & Family Studies at Penn State. Anneke's research interests are in the effects of child maltreatment, specifically in identifying mediators and moderators in the association between child maltreatment and later behavioral and psychiatric outcomes. She is particularly interested in individual characteristics (e.g., emotion regulation), and family relationships (e.g., the parent-child relationship) and the ways in which these factors confer risk and resilience for youth who have been maltreated. Ultimately, she is interested in the development and evaluation of prevention and intervention programs for children and families who have been impacted by child maltreatment.

Emily Dunning, B.A.

Emily Dunning is a graduate student in the Department of Human Development & Family Studies at Penn State. Emily is interested in parent-child interaction and mechanisms that impact the relationship between child maltreatment and parent and child biobehavioral outcomes. Emily is also interested in applying research findings to the evaluation and development of effective intervention programs for children and families following exposure to child maltreatment.

Kenneth A. Shores, Ph.D.

Dr. Kenneth A. Shores is an Assistant Professor specializing in education policy in the School of Education at the University of Delaware, and he is affiliated with the University of Delaware Center for Research in Education and Social Policy, the Biden School of Public Policy and Administration, and the Data Science Institute. His research is focused on educational inequality and encompasses both descriptive and causal inference. To this end, his work addresses racial/ethnic and socioeconomic inequality in test scores, school disciplinary policy, classification systems, and school resources. In addition, he has examined how improvements to school finance systems can reduce educational inequality and how vulnerabilities in school finance systems can contribute to it.

<u>Nilam Ram, Ph.D.</u>

Nilam Ram is Professor of Communication and Psychology at Stanford University. He studies the dynamic interplay of psychological and media processes and how they change from moment-to-moment and across the life span. Nilam's research grows out of a history of studying change. After completing his undergraduate study of economics, he worked as a currency trader, frantically tracking and trying to predict the movement of world markets as they jerked up, down and sideways. Later, he moved on to the study of human movement, kinesiology, and eventually psychological processes - with a specialization in longitudinal research methodology. Generally, Nilam studies how short-term changes (e.g., learning, information processing, emotion regulation, etc.) develop across the life span and how longitudinal study designs contribute to generation of new knowledge. He is developing a variety of study paradigms that use recent developments in data science and the intensive data streams arriving from social media, mobile sensors, and smartphones to study change at multiple time scales.

<u>Zachary F. Fisher, Ph.D.</u>

Zachary Fisher is an Assistant Professor in the Department of Human Development & Family at Penn State and member of the Quantitative Developmental Systems Methodology Core (QuantDev). His research interests lie at the intersection of developmental science, time-dependent processes, and statistical computing. Broadly, his research activities are focused on methods development in service of health and human development research and fall into three main areas: the modeling of complex time-dependent systems, measurement issues commonly encountered in behavioral applications, and the synthesis of multi-way data (e.g. cross-sectional and time-series). While the primary focus of his work is methodological in nature, Dr. Fisher is heavily involved in research on the bio-psycho-social consequences of childhood trauma and extreme stress.

<u>John F. Felt, Ph.D.</u>

John Felt is an Assistant Research Professor in the Center for Healthy Aging at Penn State. Dr. Felt is a health psychology methodologist with distinct, yet complementary research interests in statistical methods for evaluating change over time and biobehavioral and psychosocial processes underlying stress, early life adversity, and aging outcomes. Dr. Felt's current work uses Bayesian and frequentist approaches to understand the long-term biological and cognitive consequences of child maltreatment and implications for aging.

Ulziimaa Chimed-Ochir, M.A.

Ulziimaa Chimed-Ochir is a doctoral student in the Department of Human Development and Family Studies at PSU. Ulziimaa's research interest revolves around parenting and parent-child relationships in the context of child maltreatment and complex trauma. Informed by the developmental perspective, she is interested in studying the negative effects of trauma on stage salient tasks (e.g., formation of attachment relationships, development of autonomy, social-emotional development) during the early childhood period. She is also interested in examining the effects of trauma-focused parent-child relational treatment programs across early- to mid-childhood, to understand the developmental processes in the aftermath of trauma experience, and the ways in which negative effects of complex trauma unraveled.

Abstract

Contamination occurs when members of a control condition receive or are exposed to the treatment under scientific investigation. The presence of contamination violates assumptions within counterfactual models of causal inference and results in two systematic and sequential problems: 1) measurement error in the form of misclassification of units in a control condition, and 2) bias in statistical modeling that affects the direction, magnitude, and significance of causal estimates. Contamination has the potential to underestimate the true causal effect within an individual study while creating variation in causal estimates across studies based on different degrees of contamination present. Originally examined in experimental research, this paper introduces the concept of contamination as applied to observational research and uses the substantive area of child maltreatment as an illustrative example. The paper also offers methodological solutions to improve the detection of contamination while describing statistical approaches that demonstrate the impact of contamination bias and estimate causal effects in observational research after it is controlled. The goal of this chapter is to orient child maltreatment scientists conducting observational research to the issue of contamination bias and current approaches for addressing it.

Keywords: child maltreatment; contamination; counterfactual; observational research; dual measurement strategy; propensity score; synthetic controls; augmented synthetic controls; quasi-experimental; non-randomized; official case records; self-report; LONGSCAN; NSCAW-II

1 Background

The counterfactual model of causal inference serves as the foundation of modern scientific knowledge. The empiricist philosopher David Hume is generally credited with being the first to establish counterfactual reasoning in the case of causal inference: “We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where if the first object had not been, the second never had existed (Hume, 1748/2007, p.56).” This reasoning has been formally developed and refined since Hume proposed it (Neyman, 1923/1990; Pearl, 2000) with the counterfactual model now endemic across scientific disciplines examining cause and effect relations, including the psychological (Shadish et al., 2002), statistical (Rubin, 2005), and public health sciences (Höfler, 2005). The counterfactual model has also produced seminal experimental and observational research designs for promoting causal inference, including the randomized experiment in agriculture and biology, the natural experiment in physics and economics, the prospective cohort design in epidemiology, the case-control design in medicine, and the single-subjects design in behavioral analysis. Indeed, much of what is known about the relations among natural phenomena is the direct result of the counterfactual model of causal inference and the methods it has engendered.

A foundational assumption of the counterfactual model of causal inference is that different levels or conditions of a causal variable, such as treatment vs. control or exposure vs. comparison, are and remain mutually exclusive throughout data collection on an outcome of interest. That is, no one unit (e.g. cell, plant, person, state) assigned to one level of a causal variable participates in or receives another level of that same causal variable. While this assumption exists in multiple counterfactual frameworks (Cook & Campbell, 1979; Morgan et al., 2009), it is stated explicitly as part of the stable unit treatment value assumption (SUTVA) in the potential outcomes framework (West & Thoemmes, 2010). In this framework, it is assumed that each unit does not receive multiple levels, or different versions, of the causal variable of interest (see Imbens & Rubin, 2015, p.10). Adherence to the general mutual exclusivity assumption, and SUTVA specifically, ensures that inferences about the direction, significance, and magnitude of the causal effect are unbiased, promoting replication and reproducibility that ultimately strengthen causal inferences.

In reality, there are many occasions when SUTVA is violated in both experimental and observational research. For example, wind during an evaluation of a new, genetically-modified strain of corn designed to improve crop-yields or resistance to pest infestations may introduce cross-pollination with a non-modified strain in a nearby field (Quist & Chapela, 2001). The effects of a de-worming treatment to improve attendance in select African schools can indirectly benefit attendance in nearby control schools through a geographic reduction in overall infection rates (Davey et al., 2015; Hicks et al., 2015; Miguel & Kremer, 2004). An educational intervention for improving self-concept for certain students within a classroom can benefit other students in the same classroom who did not directly receive the intervention but who interacted with students that did about the intervention (Craven et al., 2001). Failure to detect and correct these commonly occurring SUTVA violations can have serious scientific consequences. Specifically, SUTVA violations can introduce bias into resulting causal estimates, namely, where the direction, magnitude, or statistical significance of between-level differences is attenuated because some proportion of the units in a control condition have actually received the treatment under investigation (Jo, 2002; Marfo & Okyere, 2019). This can have wide spread implications for causal estimates within and across scientific disciplines regardless of the research design or cause-effect relation under examination.

This chapter describes a specific and commonly occurring SUTVA violation, contamination, that has the potential to bias causal estimates across studies and health outcomes. While contamination has been examined extensively in experimental research, where known methodological and statistical solutions exist, it has not, to our knowledge, been directly applied to observational research when random assignment to treatment condition is impossible or unethical. This means that not only is the presence and impact of contamination in observational research relatively unknown but few options exist for controlling any resulting bias in causal estimates. As a result, this chapter aims to: 1) orient child maltreatment researchers to the presence and impact of contamination, 2) identify innovative methods for detecting and controlling contamination bias in observational research, and 3) describe the application of advanced statistical models for estimating causal effects that address other, known biases in observational research (e.g. covariate imbalance) after contamination is controlled. We define what contamination is, how it occurs in child maltreatment research, the overall prevalence of contamination, the bias it creates in causal estimates, the current approaches for controlling it, and the modeling of causal effects in its absence. We use the terms “treatment” and “control” throughout the chapter to refer to two different levels of a single causal variable of interest (e.g. child maltreatment and non-child maltreatment conditions) and to maintain consistency across terms used in experimental and observational research. We see other commonly used terms, such as “exposure” and “comparison” conditions, as equivalent for our purposes here. Also, we use the term “observational research” throughout but see other terms, such as “non-randomized” and “quasi-experimental”, as interchangeable in the current context. Our hope is that a greater awareness to the issue of contamination, and the current methods for addressing it, will provide child maltreatment researchers with the tools they need to enhance the accuracy of causal estimates and restore the benefits of the counterfactual model of causal inference in the observational case.

2 What is Contamination?

Contamination is “the use of the treatment by individuals in a control arm” (Cuzick et al., 1997, p. 1017) or “when intervention-like activities find their way into the control group” (Delgado-Rodríguez & Llorca, 2004, p. 640). Historically, the presence and impact of contamination on causal estimates has been examined in randomized controlled trials (RCT) research. For example, an RCT examining the benefits of prostate-specific antigen (PSA) screening found that a certain number of individuals assigned to a no-screening control condition ultimately received PSA screening from an independent physician (Roobol et al., 2009). Such contamination is a SUTVA violation and has the effect of creating bias in casual estimates via misclassification, where units originally assigned to the control condition, but who subsequently received PSA screening, are now misclassified as control units when in fact they received the treatment. This bias minimizes the direction, significance, and magnitude of between-group differences because the values of the outcome of interest, prostate-specific morbidity or mortality, for those misclassified units will be closer to the values observed for units in the treatment condition rather than other units of the control condition (Hirano et al., 2000; Jo, 2002; Kerkhof et al., 2010; Marfo & Okyere, 2019).

Contamination, while specific to a control condition, is similar to two other phenomena observed in RCTs: 1) non-compliance, where units assigned to a treatment or control condition fail to adhere to the prescribed protocol, including members of a control condition receiving the treatment under investigation (Angrist et al., 1996), and 2) spill-over, when units assigned to a treatment condition interfere with or disseminate information about the treatment to units in a control condition (Vanderweele et al., 2013). Because these phenomena have been well-studied in the RCT context, there are now well-known solutions available to RCT researchers. For example, treatment fidelity and adherence monitoring (Conn & Ruppar, 2017) are prospective, methodological strategies for detecting instances of contamination, non-compliance, and spill-over when they occur in an RCT. Similarly, statistical solutions that use the randomization process as an instrumental variable can estimate the complier's (Little & Rubin, 2000) or local average treatment effect (Angrist et al., 1996) that generates unbiased estimates of the causal effect.

Contamination can also occur and ultimately affect the construction and maintenance of control conditions in non-randomized, observational research. To illustrate, say a hypothetical team of investigators conducted a 10-year prospective cohort study of the effects of pediatric lead exposure on the cognitive functioning of children immigrating to the U.S. Because random assignment to lead exposure is unethical, the investigators used a confirmatory venous sample blood lead level (BLL) test with a reference value > 3.5 μg/dL to create the treatment condition. They also required the BLL test show no detectable levels of lead to create the control condition, thereby establishing a mutually-exclusive counterfactual condition at study entry. Then say, a subset of the children in the control condition relocate to an area of the U.S. that contains lead in water used for drinking, bathing, and cooking – something that is unknown or undetected by the investigative team. Such an occurrence would constitute a SUTVA violation within the prospective cohort design, as it does in an RCT when members of a control condition seek out or receive the treatment under investigation. Children in the control condition being inadvertently exposed to lead results in a misclassification of those children as controls when in fact they have received the treatment. This has the potential to create bias in causal estimates generated in observational research, like with RCTs, where the direction, significance, and magnitude of the causal estimate for lead exposure is attenuated because observed values of cognitive function for certain units in the control condition more closely approximate observed values in the treatment condition. Thus, regardless of the aims of a particular study, or whether randomization is used or not, contamination can exist in almost any research design using counterfactual conditions and constitutes a SUTVA violation that warrants correction when it occurs. Unfortunately, unlike with RCT research, there is less awareness of this issue and fewer known solutions for detecting contamination or addressing resulting bias in the observational case.

Below, we highlight how contamination occurs in observational research on child maltreatment, demonstrate the prevalence of contamination and the impact of resulting bias in causal effect estimates, and apply existing statistical methods for estimating causal effects in observational research after controlling contamination. This is an important task, as the vast majority of causal effects generated in research on child maltreatment likely contain some degree of bias due to contamination.

3 How Does Contamination Occur?

There are many potential sources of contamination in child maltreatment research and the following list is not intended to be exhaustive. Instead, we highlight what we see as three sources of contamination in child maltreatment research to raise awareness on how it occurs in this area of research and to generate potential solutions for controlling its presence and impact in a given study.

3.1 Planned Matching of Control Units

Random assignment to a treatment or control condition is, on average, an effective strategy for achieving balance on a large set of confounding variables that in turn promotes causal inferences about the effects of the treatment under investigation. However, different research design strategies are needed to address concerns about balance and confounding when random assignment is not possible or is unethical. Matching, where a unit is assigned to a control condition because they did not receive the treatment and because they represent the same background strata as one or more units in the treatment condition, is one research design strategy that can control extraneous variability due to confounding and allows for a more accurate determination of the treatment effect (Rubin, 1973). For example, matching a unit in the control condition who is 10-12 years of age, female, Hispanic, has an annual family income of $60,000, and lives in a dual-caregiver home to one or more units in the treatment condition who has this same demographic background, strengthens conclusions that observed between-group differences on an outcome are due to the treatment and not these demographic confounds. Matching on a set of identified variables is a long-standing practice in child maltreatment research (Widom, 1988), as it is in a variety of cohort studies outside the substantive area of child maltreatment (Cheng et al., 2020), that attempts to mimic random assignment by balancing relevant confounders across treatment and control conditions and therefore controlling their potential impact on causal estimates.

However, the benefits of planned matching at the outset of a study should be considered along with some potential limitations of this approach. For example, several demographic variables commonly used for matching in studies of child maltreatment effects, such as age, race, income, and single-parent household, are also established risk factors for the occurrence of child maltreatment (Institute of Medicine, 2011). This means that by imposing a matching procedure based on these demographic variables, some units in a control condition may have already experienced child maltreatment or might experience maltreatment during longitudinal follow-up. If so, matching on these demographic variables has the potential to introduce contamination in a study due to the increased risk of child maltreatment occurring for units assigned to a control condition. As we illustrate below, statistically adjusting causal estimates by including matching variables into causal models does not mitigate bias attributable to contamination in these estimates (see Section 4). This requires that precise measurement of child maltreatment be employed early and repeatedly throughout a study to detect contamination when it occurs while using alternative means to control resulting bias.

3.2 Imprecision in the Measurement of Child Maltreatment

There are two primary methods for measuring child maltreatment: 1) official case records, such as the report generated from a Child Protective Services (CPS) investigation of a formal allegation of child maltreatment, and 2) self-report, including questionnaires, surveys, and interviews that determine the subjective experience of child maltreatment. Each method has unique strengths and weaknesses for measuring child maltreatment in a given study. Official case records are advantageous because they are used by the Federal government to track the incidence of child maltreatment in the U.S. They are also generated by trained professionals who are independent from a given research project and masked with respect to the research hypotheses, minimizing experimenter bias. However, an allegation of child maltreatment made to CPS is required for a record to be generated and it is very likely that not all instances of child maltreatment are reported to CPS. For example, the current incidence of child maltreatment in the U.S. ranges from 8.4 to 42.9 per 1000 children (U.S. Department of Health and Human Services, 2022), an estimate considerably lower than 152 per 1000 children estimate generated by self-report methods (Finkelhor et al., 2015). This means that official case records likely reflect “true” cases of child maltreatment when they are substantiated but miss a certain number of “true” cases of child maltreatment when they are not substantiated or reported to CPS. Hence, using only official case records to establish and maintain child maltreatment and control conditions can introduce contamination in that a certain number of units in the control condition, who do not have an official record of child maltreatment, have actually been exposed to maltreatment with this experience going unknown to investigators.

Self-report methods are often selected given their widespread availability, efficiency in determining child maltreatment status, and potential sensitivity to detecting cases of child maltreatment relative to official case records. These features are particularly advantageous for large-scale, epidemiological studies or studies employing a cross-sectional or retrospective assessment of child maltreatment. However, self-report assessments of child maltreatment rely on the content validity of the items in those assessments, something that is highly variable across different instruments and with respect to official definitions and legal standards of child maltreatment (Mathews et al., 2020). Self-report methods are also subject to recall and memory biases that can affect the reporting of child maltreatment (Baldwin et al., 2019; Hardt & Rutter, 2004), something that can be compounded by mono-method bias when a self-reported health outcome is assessed simultaneously with self-reported child maltreatment status (Green et al., 2010; Newbury et al., 2018). Finally, nearly 50% of people with an official record of child maltreatment fail to report this history during a self-report assessment (Everson et al., 2008; Widom & Shepard, 1996), suggesting that a substantial number of units in a control condition established using only a self-report assessment will actually have a history of child maltreatment that is unknown to investigators.

Unfortunately, there is no gold-standard measurement of child maltreatment and the most commonly used methods are each likely to miss cases when used in isolation, resulting in certain units within an established control condition reporting or experiencing child maltreatment at some point during a given study timeline (e.g. contamination). This problem in imprecision is made worse when one of these methods is used only once at study entry to classify individuals into treatment and control conditions. Even if a particular research project was successful in correctly classifying all cases of child maltreatment into treatment and control conditions at study entry, the risk for contamination continues should the research design be longitudinal in nature.

3.3 Cross-Sectional Assessment of a Time-Varying Phenomenon

Child maltreatment most often occurs from pregnancy throughout childhood, however, children continue to experience maltreatment up to age eighteen (Sedlak et al., 2010; U.S. Department of Health and Human Services, 2022). This means that the integrity of an established control condition needs to be continually monitored for the presence of contamination so long as data collection continues. For example, many researchers are interested in studying exposure to child maltreatment in early childhood as a sensitive period that may have lasting effects on subsequent pediatric and adulthood health (Juster et al., 2011). This type of research requires continual monitoring of contamination, as individuals who were not exposed to maltreatment in early childhood may be subsequently exposed during later childhood or adolescence, potentially biasing resulting causal estimates.

Repeatedly assessing exposure to child maltreatment is similar to treatment fidelity and adherence monitoring in RCT research, where units are continually assessed throughout data collection to ensure that those assigned to either treatment or control conditions received only the treatment to which they were assigned. In child maltreatment research, this means continually tracking whether units assigned to the control condition at study entry remain unexposed to child maltreatment for the duration of a study. The continuing risk for child maltreatment up to adulthood, particularly for those who are already at increased risk due to other demographic variables, can therefore introduce contamination into a study implementing a longitudinal, repeated measures assessment.

4 Does Contamination Bias Causal Estimates in Observational Research?

Contamination is measurement error in the form of a misclassification of units assigned to a control condition. Like in RCT's, this misclassification has the potential to produce bias that affects the direction, statistical significance, and magnitude of the treatment effect in observational research. The major concern when contamination occurs is that it is either unknown to or uncontrolled by investigators, ultimately leading to biased estimates of the causal effect of interest (see Figure 1). Below is a brief review of existing empirical studies that have demonstrated the impact of contamination bias on the significance and magnitude of causal estimates and the bias reduction that occurs in those same estimates when contamination is controlled.

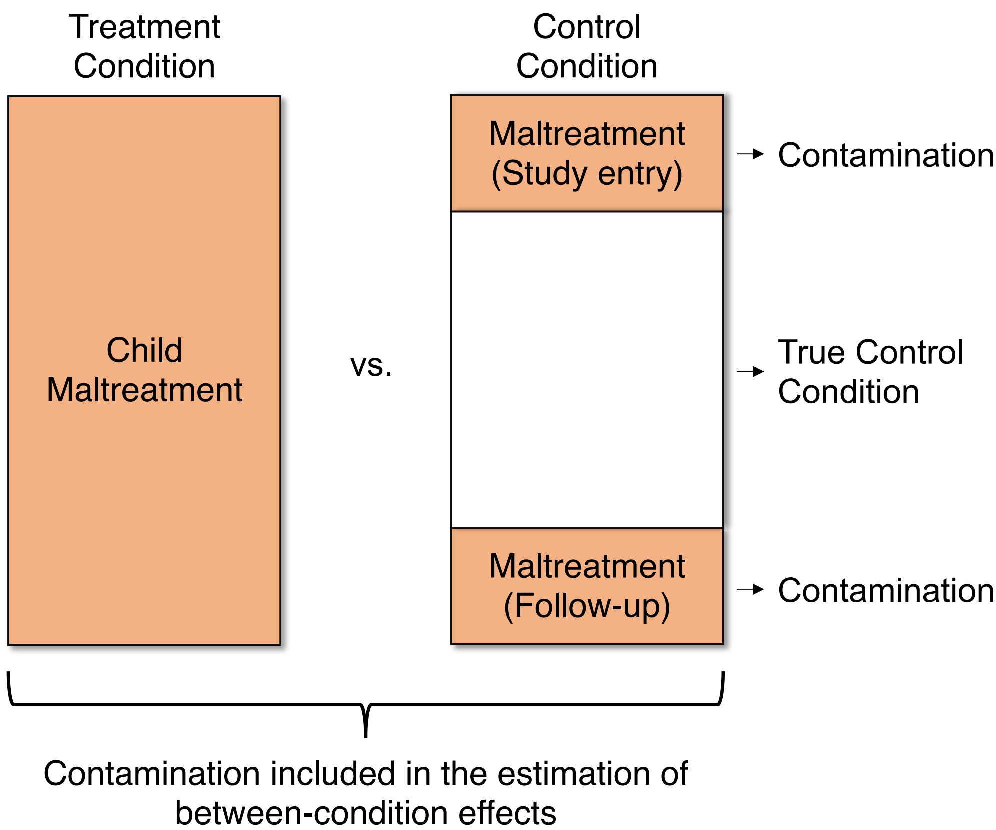

Figure 1. Traditional Modeling of Child Maltreatment Effects

Scott and colleagues (2010) examined the risk for psychiatric disorders in young adulthood following exposure to child maltreatment in a nationally-representative cohort in New Zealand (N = 2144). In this study, the investigators established child maltreatment and control conditions using official case records, where child maltreatment status was determined by an allegation of child maltreatment that was reported to and investigated by CPS. Interestingly, the investigators administered a self-report assessment of child maltreatment in this study, which indicated that 15.4% of units in the control condition reported exposure to child maltreatment. Using traditional estimation procedures that retained contamination in the statistical model, results indicated that child maltreatment significantly increased the risk for past year (OR = 2.32) and lifetime (OR = 2.12) occurrence of a psychiatric disorder. However, when contamination was controlled by removing the 15.4% of control units who self-reported child maltreatment from the model and re-estimating risks for these same outcomes, the effect size magnitudes for both past-year (OR = 2.83) and lifetime (OR = 2.80) risk of a psychiatric disorder increased by 22%-32%, respectively. Similar trends in effect size magnitude were observed for individual psychiatric disorders, including several that achieved statistical significance only after contamination was controlled.

In a multi-wave, prospective cohort study in the U.S. (N = 514), Shenk and colleagues (2016) attempted to replicate prior research establishing child maltreatment effects on several indicators of female health at the transition to adulthood: teenage births, past-month cigarette use, obesity status, and clinical levels of major depressive disorder symptoms. Child maltreatment status was determined using official case records, where an allegation of child maltreatment was made, investigated, and substantiated by CPS. Control units were demographically-matched to units in the maltreatment condition on age, race, family income, and single-parent household. Using traditional estimation procedures that adjusted estimates based on the inclusion of matched demographic variables, this study failed to replicate statistically significant child maltreatment effects for obesity and major depressive disorder symptoms. However, Shenk and colleagues then screened for contamination at each wave of data collection using both official case records and self-reports of child maltreatment. This multi-method screen identified 44.8% of control units who experienced child maltreatment. When this contamination was controlled by removing these units from the statistical model, child maltreatment significantly predicted all four female adolescent health outcomes (RR = 1.47-2.95), replicating prior research (e.g. Danese & Tan, 2014; Widom et al., 2007). Moreover, effect size magnitudes for these four outcomes increased by 24%-130% once contamination was controlled, again providing an indication of the degree of contamination bias in the causal estimate.

Shenk and colleagues (2021) investigated the impact of contamination on causal estimates when examining the effect of child maltreatment on age-heterogeneous trajectories of internalizing and externalizing behaviors across childhood and adolescence. Using existing data from a national, multi-site, and multi-wave prospective cohort in the U.S. (N = 1354), this study established child maltreatment and control conditions using official case records where independent raters confirmed exposure to child maltreatment based on details obtained in the official records. Shenk and colleagues then screened the no confirmed child maltreatment control condition for contamination using a repeatedly administered self-report assessment of child maltreatment. The authors identified a contamination prevalence estimate of 65.1% in this sample, meaning nearly two-thirds of units assigned to a no confirmed child maltreatment control condition reported maltreatment and were ultimately misclassified. Traditional estimation of confirmed child maltreatment effects revealed statistically significant risks for both internalizing and externalizing behavior trajectories with effect size magnitudes ranging from d’s = .19-.40. Contamination was then controlled by modeling the 65.1% of individuals who were misclassified as control units as a third condition (self-reported maltreatment without a confirmed case record) and deriving contrasts between the confirmed child maltreatment condition and the resulting no confirmed, no self-reported child maltreatment control condition. When models were re-estimated, statistically significant risks for internalizing and externalizing behavior trajectories were again observed for confirmed child maltreatment but with effect size magnitude increases of 27.5%-52.6% (d's = .29-.51).

These three studies illustrate several important aspects pertaining to misclassification of control units in the form of contamination in child maltreatment research. One, contamination exists in child maltreatment research with current prevalence estimates ranging from 15.4% to 65.1%. The existence of contamination constitutes a SUTVA violation and requires detection and control to generate accurate estimates of child maltreatment effects. Two, failing to control contamination can bias the significance and magnitude of causal estimates toward the null, making it harder to detect an effect, as well as the precise degree of that effect, for child maltreatment when it exists. This has obvious implications for the replication and reproducibility of effects observed in child maltreatment research, particularly when contamination is more or less prevalent across independent studies. Finally, traditional modeling of child maltreatment effects likely contains some degree of contamination bias, an approach that can underestimate the true causal effect of child maltreatment. Identifying ways to detect and control contamination prior to modeling child maltreatment effects holds considerable promise for improving the accuracy of causal estimates in this area of research.

5 How to Detect and Control Contamination Bias?

Ways for detecting and controlling contamination bias is an ongoing area of research. So far, a dual measurement strategy (Brenner & Blettner, 1993; Marshall & Graham, 1984), one that capitalizes on the different strengths of existing methods for determining exposure to child maltreatment, offers some degree of control of contamination while also demonstrating the bias in causal effects that contamination creates. For example, each of the three studies reviewed in Section 4 applied a dual measurement strategy that used official case records to establish treatment and control conditions and self-report methods to detect and control contamination. This approach appears advantageous for two reasons. One, an indication of child maltreatment using official case records is most likely to result in lower false positive rates, where if maltreatment is determined to have occurred it is most likely to be a true instance of child maltreatment. Two, it is not appropriate to conclude that a negative indication of child maltreatment based on official case records means that maltreatment did not occur, only that the investigation did not produce enough evidence to confirm, indicate, or substantiate the allegation. As a result, using a second measure that is likely more sensitive to detecting child maltreatment, such as self-reports of child maltreatment, can detect cases of child maltreatment in the control condition and offer one way to identify and control contamination.

This dual measurement strategy has proved effective in prior child maltreatment research that demonstrated enhanced sensitivity for detecting cases of child maltreatment (Swahn et al., 2006) as well as stronger effects for child maltreatment on subsequent psychiatric disorders (Shaffer et al., 2008). Furthermore, unpublished results from a prior examination of contamination bias (Shenk et al., 2016) highlight the impact of using a dual measurement strategy to detect contamination on the significance and magnitude of resulting effect size estimates relative to any single measure making up that dual strategy (see Table 1). Thus, for now, a dual measurement strategy to detect and control contamination in observational child maltreatment research appears to be the best method for approximating the benefits of treatment fidelity and adherence monitoring used to detect contamination in RCT research.

	No Screening	**Self-report Only**	**Case Records Only**	**Dual Measurement**
Major Depression	1.28 (0.79-2.08)	1.70 (0.94-3.08)	**1.79** (1.05-3.02)	**2.95** (1.22-7.16)
Teen Births	**1.66** (1.06-2.61)	1.27 (0.93-1.74)	1.29 (0.96-1.72)	**2.21** (1.06-4.63)
Obesity	1.16 (0.90-1.50)	**2.84** (1.24-6.51)	1.22 (0.72-2.05)	**1.47** (1.03-2.08)
Cigarette Use	**1.36** (1.06-1.74)	**1.56** (1.14-2.14)	**1.46** (1.13-1.87)	**1.68** (1.21-2.35)

Table 1. Dual Measurement Strategy for Detecting Contamination. Effect size estimates are relative risks and corresponding 95% confidence intervals after accounting for demographic covariates. Bolded numbers are statistically significant estimates ($p < .05$).

However, once contamination is detected using this dual measurement strategy, it remains unclear whether completely removing identified control units from a statistical model (i.e. reduced overall sample size), examining a “contamination” condition as its own condition in a statistical model (i.e. retains overall sample size), or some other approach provides the most accurate estimator of child maltreatment effects. Two obvious implications for controlling contamination based on the existing research are the potential impact on: 1) statistical power via reductions in sample size combined with expected increases in effect magnitude, features that are likely to vary based on different degrees of contamination present across studies, and 2) the external validity of results based on how similar the analytic sample matches the overall sample and overall population (Degtiar & Rose, 2021). One study did demonstrate that a dual measurement strategy to control contamination by removing units from the statistical model did result in a revised control condition that had prevalence estimates for all outcomes that more closely approximated U.S. population prevalence estimates, aiding generalizability (see Shenk et al., 2016; Figure 1). Future research evaluating these approaches to controlling contamination should also evaluate resulting implications on statistical power and generalizability, implications that are likely to offer child maltreatment researchers several empirically-driven strategies in which to approach the issue of detecting and controlling contamination bias.

Two other potential ways for addressing contamination in child maltreatment research have been suggested during peer review of our prior research, which we do not recommend at this point. The first is to move misclassified control units who were identified as experiencing maltreatment using one method (e.g. self-report) into a child maltreatment condition established using a second method (e.g. official case records). There are several reasons for not following this approach. One, there is low agreement between self-report and official case records as indicators of child maltreatment status (Baldwin et al., 2019; Everson et al., 2008), demonstrating that these two methods are not interchangeable. Two, prospective methods of assessing child maltreatment, typically represented by official case records, and retrospective methods of assessing child maltreatment, typically represented by self-report methods, are associated with differential risks for physical and behavioral health outcomes (Newbury et al., 2018) - particularly when those outcomes are also assessed using self-report methods (mono-method bias; see review by Coleman & Baldwin in this book). Identifying child maltreatment in a control condition using one method and then adding those units to a child maltreatment condition established using another method adds unnecessary heterogeneity into the treatment condition and the evaluation of treatment effects that is attributable to the measurement of child maltreatment (see Figures 1 and 2 in Shenk et al., 2021). Three, detecting and controlling contamination using a dual measurement strategy means this – that identified control units are misclassified and should no longer be regarded as control units. It does not mean, at least given the differences in self-report and official case records for determining child maltreatment, that those misclassified units are now equivalent to those units in the treatment condition.

The second way is to capitalize on existing approaches for addressing misclassification bias that have been developed and examined in the epidemiological and econometric literatures (Keogh et al., 2020). In these approaches, the potential misclassification of units is examined not just in a control condition but also in the treatment condition. This examination then determines whether the misclassification is differential or non-differential across the treatment and control conditions, which ultimately determines how treatment effects will be estimated. We view contamination in child maltreatment research to be distinct from the application of the existing approaches to addressing misclassification bias for several reasons. One, as described above, there is no gold-standard measure of child maltreatment nor are the existing methods non-differential in their sensitivity and specificity for detecting maltreatment, assumptions that are often required when addressing misclassification bias in this framework (Flanders et al., 1995; Johnson et al., 2014). Two, contamination refers to a phenomenon that only affects a control condition, not a treatment condition, and examining differential or non-differential misclassification across both of these conditions does not fully address the issue of contamination in promoting the accuracy of causal estimates. Instead, a dual measurement strategy specific for a control condition capitalizes on the differential psychometric strengths of both measures to add precision to the measurement of child maltreatment. Three, official case records yield low false positive rates, meaning that when an allegation of child maltreatment is made, investigated, and determined to have occurred, it is likely true. In other words, there is little chance that a unit in a treatment condition created using official case records can "not-comply" or "non-adhere" to this treatment condition or be misclassified. A positive indication of exposure to child maltreatment using official case records will also always be true regardless of whether or not an individual recalls it on a self-report instrument at a later date, something that could be affected by an individual simply not having a memory of the maltreatment event because it happened early in infancy or childhood. As noted earlier, child maltreatment most often occurs early in the life course and many child maltreatment researchers are interested in how this event early in life creates lasting neurobiological change that results in later health risks (Juster et al., 2011). This can be measured using official case records regardless of whether a person can or does recall this event on a self-report instrument or not.

Below are active projects in which we are evaluating different research design, measurement strategies, and statistical approaches that not only attempt to detect and control contamination but also generate causal estimates of the effect of child maltreatment on subsequent health. Each uses a unique dual measurement strategy for detecting and controlling contamination. The latter two, given data are already collected, are also applying advanced statistical methods for estimating causal effects in observational research and demonstrating the degree of bias in causal effect estimates due to contamination. We have been fortunate to receive two federally-funded grants, one from the National Institutes of Health (NIH; R03HD104739) and the other from the National Science Foundation (BCS-2041333), allowing us to formally examine ways for evaluating the impact of contamination bias and generating causal estimates of the effects of child maltreatment on subsequent behavior problems once it is controlled.

5.1 The Translational Center for Child Maltreatment Studies (TCCMS)

A dual measurement strategy to detect and control contamination is designed to capitalize on the individual strengths of each method and is enhanced by an early, repeated, and prospective administration for determining exposure to child maltreatment. The TCCMS at Penn State, the first NIH Capstone Center for research on child maltreatment, has implemented a dual measurement strategy for detecting contamination within a prospective cohort study of the long-term effects of child maltreatment on pediatric health. The two methods constituting the dual measurement strategy in this study are: 1) official case records, and 2) self-reported instances of child maltreated obtained from research interview. The TCCMS is applying several innovative methods for establishing exposure to child maltreatment from official case records. For example, eligibility for inclusion in the study requires that units in the child maltreatment condition were the subject of a child maltreatment report, that is, an allegation of child maltreatment that was investigated by CPS, within the year prior to study entry. This report could have been substantiated or not, providing the opportunity to examine three different child maltreatment conditions: investigated but not substantiated, investigated and substantiated, and neither investigated nor substantiated (e.g. control condition). This approach to characterizing child maltreatment status has shown to enhance the sensitivity of official case reports by increasing overall prevalence with few differences in health outcomes observed between investigated but not substantiated vs. investigated and substantiated conditions (Kugler et al., 2019). The TCCMS is also employing the Maltreatment Classification System (MCS; Barnett et al., 1993) to review and code official case records for varying child maltreatment dimensions, including type, duration, age at onset, and severity, spanning the entire lifetime of units. MCS codes are determined by independent raters unaware of the research hypotheses who review the official case records generated by CPS to make their codes regarding various child maltreatment dimensions. The MCS has enhanced sensitivity to detecting child maltreatment (Runyan et al., 2005), a feature that can strengthen the assessment of maltreatment using information from official case records.

The second method used in the dual measurement strategy for the TCCMS is the Trauma History Profile (THP) of the UCLA PTSD Reaction Index for DSM-5 (RI-5;

Pynoos & Steinberg, 2015), which is administered prospectively at each wave of data collection via a semi-structured interview with trained and reliable interviewers. The RI-5 is a psychometrically-sound instrument (Kaplow et al., 2020; Steinberg et al., 2013) that determines self-reported exposure to the most common types of child maltreatment, including physical abuse, sexual abuse, psychological abuse, and neglect. Once a control condition is established using official case records, information from the THP will be used to detect instances of self-reported child maltreatment in that control condition, that is, contamination. This allows for an estimation of the prevalence of contamination in the TCCMS study, identifying a subgroup of units that report experiencing child maltreatment but who have no official case record. It can also inform ways in which to control contamination based on the potential impact on statistical power and external validity as well as various statistical modeling strategies (see below). This innovation in research design via a dual measurement strategy for determining child maltreatment at study entry and at each subsequent wave of data collection is important because data generated from the TCCMS study will be made available to scientists across the nation and world to study the public health impact of child maltreatment. Thus, a more accurate understanding of the effects of child maltreatment is likely to be gained from this study when contamination is detected and controlled in subsequent modeling.

5.2 The Longitudinal Studies of Child Abuse and Neglect (LONGSCAN)

LONGSCAN (Runyan et al., 1998) is a multi-wave, multi-site, prospective cohort study of child maltreatment in the U.S. occurring from birth through age eighteen. The LONGSCAN database provides a unique opportunity to examine the issue of contamination given the repeated measurement of child maltreatment and behavioral health outcomes throughout the entire pediatric period. The dual measurement strategy we are evaluating with the LONGSCAN cohort is: 1) official case records where a modified version of the MCS was developed, applied, and evaluated, and 2) self-report methods, where one set of measures was administered prospectively from 12-16 years of age and another set was administered retrospectively at age 18. This dual measurement strategy affords several opportunities to examine contamination. One, it capitalizes on the strengths of official case records and enhanced sensitivity of the MCS to flexibly characterize different child maltreatment and control conditions. Two, once a control condition is established using official case records, the dual measurement strategy then allows for an estimation of contamination prevalence and bias using both prospective and retrospective self-report methods – the most commonly used approaches for estimating the effects of child maltreatment. Examining contamination prevalence and bias in LONGSCAN across prospective and retrospective self-report measures can inform researchers on whether the bias stemming from contamination varies depending on which of these assessment methods are used. These efforts will provide further insights into the prevalence and impact of contamination in the LONGSCAN cohort and the effectiveness of the proposed dual measurement strategy.

The nature of the LONGSCAN data also provides opportunities for innovative econometric modeling. For example, synthetic and augmented synthetic control modeling (SCM) approaches for estimating causal effects in observational research can be leveraged using the LONGSCAN cohort design. Briefly, SCMs estimate causal effects by matching a unit in the treatment condition to one or more units in the control condition who are similar in their weighted levels of an outcome measured prior to the treatment of interest occurring. For example, we are using repeated measurements of child behavior problems in LONGSCAN to find a matched control condition similar to the maltreatment condition in their pre-maltreatment observations of these behavior problems. Because LONGSCAN repeatedly measured child behavior problems at each wave of data collection from ages 4 to 16, we are able to use the observations on this outcome across ages 4 to 8 to identify control units who were never maltreated throughout LONGSCAN and match them to units who were not maltreated from 4 to 8 years of age but who were maltreated at age 10 or later. The research design and measurement strategy of LONGSCAN, coupled with the innovative application of SCMs to these data, will allow for the first ever causal estimation of child maltreatment effects on child behavior problems throughout adolescence once contamination bias is controlled.

5.3 The National Survey of Child and Adolescent Well-being - II (NSCAW-II)

NSCAW-II is a multi-wave, prospective cohort study of child maltreatment (N = 5872; Dolan et al., 2011) comprised of children between birth and 17.5 years of age at study enrollment who were recruited from 83 counties across 30 states in the U.S. All participants were recruited from child welfare investigations closed between February 2008 and April 2009. The first wave of NSCAW-II data collection occurred between March 2008 and September 2009, following the closure of the NSCAW-II index case investigation. The second wave of data collection occurred approximately 18 months thereafter and the final or third wave occurred 18 months following the second wave or approximately 36 months from baseline data collection. Importantly, the NSCAW-II index case, that is the child welfare investigation closed just prior to baseline data collection, includes both substantiated and unsubstantiated cases. NSCAW-II also oversampled for infants, families receiving services, and children in out-of-home placements.

Similar to LONGSCAN, NSCAW-II utilized a multi-informant, dual measurement strategy that affords the opportunity to examine the impact of contamination on child maltreatment effect estimates. To assess contamination in NSCAW-II, we are utilizing: 1) an official substantiation indicator, provided by the NSCAW-II team, that designates whether the index child maltreatment investigation that brought participants into the NSCAW-II cohort was or was not substantiated, and 2) caregiver reports, administered at each of the three waves of NSCAW-II data collection via the Parent-Child Conflict Tactics Scale (Straus et al., 1998), which provides data on caregiver-reported psychological aggression, physical assault, child neglect, and sexual maltreatment. Thus, as described above with LONGSCAN, using a substantiation indicator based on official case records, we are able to establish initial treatment and control conditions. In turn, the prevalence of contamination and resulting bias can be estimated using the seemingly more sensitive caregiver-reports of child maltreatment.

Using NSCAW-II, we are applying propensity score methods as a means to examine the impact of contamination on causal estimates of child maltreatment effects. Propensity score methods are a well-established approach to achieving covariate balance in observational research when randomization to treatment condition is not feasible or is unethical, such as in the case of child maltreatment. Specifically, propensity score methods aim to reduce bias in the estimation of a treatment effect in observational research, that is, the difference between the treatment and control conditions, when it is possible that there are systematic differences between these conditions on confounding variables. A propensity score refers to “the conditional probability of assignment to a particular treatment given a vector of observed covariates” (Rosenbaum & Rubin, 1983, p. 41). In the case of child maltreatment, therefore, a propensity score refers to the conditional probability of experiencing child maltreatment given particular covariates (e.g., income, child gender, caregiver education, race, and ethnicity) and is typically calculated using logistic regression, wherein the treatment is the dependent variable and covariates are predictors. In turn, using this propensity score, different methods can be applied such as propensity score matching or inverse probability of treatment weighting (IPTW) to create covariate balance across treatment and control groups and ultimately more accurately estimate the effect of a given treatment (Austin, 2011). Furthermore, NSCAW-II not only provides dual-measurement of child maltreatment but also longitudinal assessments of child behavior problems and ample covariates related to child maltreatment status, lending itself to propensity score estimation. In sum, using NSCAW-II, we will apply propensity score methods to estimate the casual impact of child maltreatment on child behavior problems both when contamination is and is not controlled.

6 Conclusion

It is very likely the case that contamination is a feature, not a bug, of conducting child maltreatment research. Attention to this issue is important, as contamination results in a bias that minimizes the significance and magnitude of causal effects regardless of the outcome examined. One possible implication of this phenomenon is that existing causal estimates are likely underestimating the true public health impact of child maltreatment. This has the potential to affect scientific, policy, and provider decisions about how to inform a statistical power analysis for future research, whether to devote intervention resources to child welfare or another family serving agency, or whether to consider the potential etiological role of child maltreatment for a presenting complaint in a pediatrician's office. Furthermore, the prevalence of contamination is likely to vary across sampling strategies used in different observational designs within child maltreatment research (e.g. prospective vs. retrospective). This means that the degree of bias in causal estimates, and therefore their impact on the direction, significance, and magnitude of causal effects, is likely to vary across studies proportional to the degree of contamination present. Variation in contamination bias across studies, combined with a lack of control for this bias, could be one explanation for prior failures in replication and reproducibility within child maltreatment research. This is likely one contributor to the overall replication crisis in the behavioral sciences at large. The implications of addressing contamination bias in child maltreatment research, or failing to, are indeed considerable.

This chapter highlights a program of research that is applying innovative methods for addressing the presence of contamination and controlling resulting bias in casual estimates within the substantive area of child maltreatment. While the implications for child maltreatment researchers are clear, we expect that contamination is a general methodological phenomenon in observational research even outside this substantive area. The methods and procedures identified here are likely highly relevant across scientific disciplines employing observational research to examine cause and effect relations between phenomena of interest. Our hope is that greater attention to addressing contamination will restore the benefits of the counterfactual model of causal inference for child maltreatment and other scientists so that statistical models can more readily produce accurate casual estimates.

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion and rejoinder). *Journal of the American Statistical Association*, *91*, 444-472.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399-424. https://doi.org/https://doi.org/10.1080/00273171.2011.568786

Baldwin, J. R., Reuben, A., Newbury, J. B., & Danese, A. (2019). Agreement Between Prospective and Retrospective Measures of Childhood Maltreatment: A Systematic Review and Meta-analysis. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2019.0097

Barnett, D., Manly, J. T., & Cicchetti, D. (1993). Defining child maltreatment: The interface between policy and research. In D. Cicchetti & S. L. Toth (Eds.), *Advances in Applied Developmental Psychology: Child Abuse, Child Development and Social Policy* (pp. 7-74). Ablex.

Brenner, H., & Blettner, M. (1993). Misclassification bias arising from random error in exposure measurement: implications for dual measurement strategies. *Am J Epidemiol*, *138*(6), 453-461. https://doi.org/10.1093/oxfordjournals.aje.a116877

Cheng, J., Gregorich, S. E., Gansky, S. A., Fisher-Owens, S. A., Kottek, A. M., White, J. M., & Mertz, E. A. (2020). Constructing Matched Groups in Dental Observational Health Disparity Studies for Causal Effects. *JDR Clin Trans Res*, *5*(1), 82-91. https://doi.org/10.1177/2380084419830655

Conn, V. S., & Ruppar, T. M. (2017). Medication adherence outcomes of 771 intervention trials: Systematic review and meta-analysis. *Prev Med*, *99*, 269-276. https://doi.org/10.1016/j.ypmed.2017.03.008

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Rand McNally.

Craven, R. G., Marsh, H. W., Debus, R. L., & Jayasinghe, U. (2001). Diffusion effects: Control group contamination threats to the validity of teacher-administered interventions. *Journal of Educational Psychology*, *93*(3), 639-645. https://doi.org/http://dx.doi.org/10.1037/0022-0663.93.3.639

Cuzick, J., Edwards, R., & Segnan, N. (1997). Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med*, *16*(9), 1017-1029.

Danese, A., & Tan, M. (2014). Childhood maltreatment and obesity: systematic review and meta-analysis. *Mol Psychiatry*, *19*(5), 544-554. https://doi.org/10.1038/mp.2013.54

Davey, C., Aiken, A. M., Hayes, R. J., & Hargreaves, J. R. (2015). Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial. *Int J Epidemiol*, *44*(5), 1581-1592. https://doi.org/10.1093/ije/dyv128

Degtiar, I., & Rose, S. (2021). *A Review of Generalizability and Transportability [PREPRINT]*. Harvard T. H. Chan School of Public Health & Stanford University. https://arxiv.org/abs/2102.11904v1 [stat.ME]

Delgado-Rodríguez, M., & Llorca, J. (2004). Bias. *J Epidemiol Community Health*, *58*(8), 635-641. https://doi.org/10.1136/jech.2003.008466

Dolan, M., Smith, K., Casanueva, C., Ringeisen, H., & RTI International. (2011). *NSCAW II Baseline Report: Introduction to NSCAW II*. R. a. E. Office of Planning, Adminstration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/documents/opre/nscaw2_intro.pdf

Everson, M. D., Smith, J. B., Hussey, J. M., English, D., Litrownik, A. J., Dubowitz, H., Thompson, R., Dawes Knight, E., & Runyan, D. K. (2008). Concordance between adolescent reports of childhood abuse and Child Protective Service determinations in an at-risk sample of young adolescents. *Child Maltreat*, *13*(1), 14-26. https://doi.org/10.1177/1077559507307837

Finkelhor, D., Turner, H. A., Shattuck, A., & Hamby, S. L. (2015). Prevalence of Childhood Exposure to Violence, Crime, and Abuse: Results From the National Survey of Children's Exposure to Violence. *JAMA Pediatr*, *169*(8), 746-754. https://doi.org/10.1001/jamapediatrics.2015.0676

Flanders, W. D., Drews, C. D., & Kosinski, A. S. (1995). Methodology to correct for differential misclassification. *Epidemiology*, *6*(2), 152-156. https://doi.org/10.1097/00001648-199503000-00011

Green, J. G., McLaughlin, K. A., Berglund, P. A., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2010). Childhood adversities and adult psychiatric disorders in the National Comorbidity Survey Replication I: Associations with first onset of DSM-IV disorders. *Archives of General Psychiatry*, *67*(2), 113-123. https://doi.org/10.1001/archgenpsychiatry.2009.186

Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology & Psychiatry*, *45*(2), 260-273.

Hicks, J. H., Kremer, M., & Miguel, E. (2015). Commentary: Deworming externalities and schooling impacts in Kenya: a comment on Aiken et al. (2015) and Davey et al. (2015). *Int J Epidemiol*, *44*(5), 1593-1596. https://doi.org/10.1093/ije/dyv129

Hirano, K., Imbens, G. W., Rubin, D. B., & Zhou, X. H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, *1*(1), 69-88. https://doi.org/10.1093/biostatistics/1.1.69

Höfler, M. (2005). Causal inference based on counterfactuals. *BMC Med Res Methodol*, *5*, 28. https://doi.org/10.1186/1471-2288-5-28

Hume, D. (1748/2007). *An Enquiry Concerning Human Understanding* (P. Millican, Ed.). Oxford University Press.

Imbens, G., & Rubin, D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Institute of Medicine. (2011). Causality. In *New Directions in Child Abuse and Neglect Research*. National Academies Press.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, *7*(2), 178-193. https://doi.org/http://dx.doi.org/10.1037/1082-989X.7.2.178

Johnson, C. Y., Flanders, W. D., Strickland, M. J., Honein, M. A., & Howards, P. P. (2014). Potential sensitivity of bias analysis results to incorrect assumptions of nondifferential or differential binary exposure misclassification. *Epidemiology*, *25*(6), 902-909. https://doi.org/10.1097/ede.0000000000000166

Juster, R. P., Bizik, G., Picard, M., Arsenault-Lapierre, G., Sindi, S., Trepanier, L., Marin, M. F., Wan, N., Sekerovic, Z., Lord, C., Fiocco, A. J., Plusquellec, P., McEwen, B. S., & Lupien, S. J. (2011). A transdisciplinary perspective of chronic stress in relation to psychopathology throughout life span development. *Development and Psychopathology*, *23*(3), 725-776. https://doi.org/10.1017/s0954579411000289

Kaplow, J. B., Rolon-Arroyo, B., Layne, C. M., Rooney, E., Oosterhoff, B., Hill, R., Steinberg, A. M., Lotterman, J., Gallagher, K. A. S., & Pynoos, R. S. (2020). Validation of the UCLA PTSD Reaction Index for DSM-5: A Developmentally Informed Assessment Tool for Youth. *J Am Acad Child Adolesc Psychiatry*, *59*(1), 186-194. https://doi.org/10.1016/j.jaac.2018.10.019

Keogh, R. H., Shaw, P. A., Gustafson, P., Carroll, R. J., Deffner, V., Dodd, K. W., Küchenhoff, H., Tooze, J. A., Wallace, M. P., Kipnis, V., & Freedman, L. S. (2020). STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1-Basic theory and simple methods of adjustment. *Stat Med*, *39*(16), 2197-2231. https://doi.org/10.1002/sim.8532

Kerkhof, M., Roobol, M. J., Cuzick, J., Sasieni, P., Roemeling, S., Schröder, F. H., & Steyerberg, E. W. (2010). Effect of the correction for noncompliance and contamination on the estimated reduction of metastatic prostate cancer within a randomized screening trial (ERSPC section Rotterdam). *Int J Cancer*, *127*(11), 2639-2644. https://doi.org/10.1002/ijc.25278

Kugler, K. C., Guastaferro, K., Shenk, C. E., Beal, S. J., Zadzora, K. M., & Noll, J. G. (2019). The effect of substantiated and unsubstantiated investigations of child maltreatment and subsequent adolescent health. *Child Abuse Negl*, *87*, 112-119. https://doi.org/10.1016/j.chiabu.2018.06.005

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*, *21*, 121-145. https://doi.org/10.1146/annurev.publhealth.21.1.121

Marfo, P., & Okyere, G. A. (2019). The accuracy of effect-size estimates under normals and contaminated normals in meta-analysis. *Heliyon*, *5*(6), e01838. https://doi.org/10.1016/j.heliyon.2019.e01838

Marshall, J. R., & Graham, S. (1984). Use of dual responses to increase validity of case-control studies. *J Chronic Dis*, *37*(2), 125-136. https://doi.org/10.1016/0021-9681(84)90054-7

Mathews, B., Pacella, R., Dunne, M. P., Simunovic, M., & Marston, C. (2020). Improving measurement of child abuse and neglect: A systematic review and analysis of national prevalence studies. *PLoS One*, *15*(1), e0227884. https://doi.org/10.1371/journal.pone.0227884

Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*(1), 159-217.

Morgan, S. L., Winship, C., & Vanderweele, T. J. (2009). Counterfactuals and causal inference: methods and principles for social research. *Journal of the American Statistical Association*, *104*(488), 1715-1716.

Newbury, J. B., Arseneault, L., Moffitt, T. E., Caspi, A., Danese, A., Baldwin, J. R., & Fisher, H. L. (2018). Measuring childhood maltreatment to predict early-adult psychopathology: Comparison of prospective informant-reports and retrospective self-reports. *J Psychiatr Res*, *96*, 57-64. https://doi.org/10.1016/j.jpsychires.2017.09.020

Neyman, J. (1923/1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, *5*(4), 465-472.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press, New York, NY.

Pynoos, R. S., & Steinberg, A. M. (2015). *The UCLA PTSD Reaction Index for Children and Adolescents - DSM-5*. Behavioral Health Innovations, LLC.

Quist, D., & Chapela, I. H. (2001). Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. *Nature*, *414*(6863), 541-543. https://doi.org/10.1038/35107068

Roobol, M. J., Kerkhof, M., Schröder, F. H., Cuzick, J., Sasieni, P., Hakama, M., Stenman, U. H., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., Denis, L., Recker, F., Berenguer, A., Ruutu, M., Kujala, P., Bangma, C. H., . . . Auvinen, A. (2009). Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). *Eur Urol*, *56*(4), 584-591. https://doi.org/10.1016/j.eururo.2009.07.018

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, *29*, 159-183.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*, 322-331.

Runyan, D. K., Cox, C. E., Dubowitz, H., Newton, R. R., Upadhyaya, M., Kotch, J. B., Leeb, R. T., Everson, M. D., & Knight, E. D. (2005). Describing maltreatment: Do child protective service reports and research definitions agree? *Child Abuse & Neglect*, *29*(5), 461-477. https://doi.org/10.1016/j.chiabu.2004.06.015

Runyan, D. K., Curtis, P. A., Hunter, W. M., Black, M. M., Kotch, J. B., Bangdiwala, S., Dubowitz, H., English, D., Everson, M. D., & Landsverk, J. (1998). LONGSCAN: A consortium for longitudinal studies of maltreatment and the life course of children. *Aggression and Violent Behavior*, *3*(3), 275-285. https://doi.org/http://dx.doi.org/10.1016/S1359-1789(96)00027-4

Scott, K. M., Smith, D. R., & Ellis, P. M. (2010). Prospectively ascertained child maltreatment and its association with DSM-IV mental disorders in young adults. *Archives of General Psychiatry*, *67*(7), 712-719. https://doi.org/10.1001/archgenpsychiatry.2010.71

Sedlak, A. J., Mettenburg, J., Basena, M., Petta, I., McPherson, K., & Greene, A. (2010). *Fourth National Incidence Study of Child Abuse and Neglect (NIS-4):*

Report to Congress. US Dept. of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, National Center on Child Abuse and Neglect.

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

Shaffer, A., Huston, L., & Egeland, B. (2008). Identification of child maltreatment using prospective and self-report methodologies: A comparison of maltreatment incidence and relation to later psychopathology. *Child Abuse & Neglect*, *32*(7), 682-692. https://doi.org/10.1016/j.chiabu.2007.09.010

Shenk, C. E., Noll, J. G., Peugh, J. L., Griffin, A. M., & Bensman, H. E. (2016). Contamination in the prospective study of child maltreatment and female adolescent health. *Journal of Pediatric Psychology*, *41*, 37-45. https://doi.org/10.1093/jpepsy/jsv017

Shenk, C. E., Rausch, J. R., Shores, K. A., Allen, E. K., & Olson, A. E. (2021). Controlling contamination in child maltreatment research: Impact on effect size estimates for child behavior problems measured throughout childhood and adolescence. *Dev Psychopathol*, 1-13. https://doi.org/10.1017/s0954579420002242

Steinberg, A. M., Brymer, M. J., Kim, S., Briggs, E. C., Ippen, C. G., Ostrowski, S. A., Gully, K. J., & Pynoos, R. S. (2013). Psychometric properties of the UCLA PTSD Reaction Index: Part I. *Journal of Traumatic Stress*, *26*(1), 1-9. https://doi.org/10.1002/jts.21780

Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of child maltreatment with the Parent–Child Conflict Tactics Scales: Development and psychometric data for a national sample of American parents. *Child Abuse & Neglect*, *22*(4), 249-270. https://doi.org/http://dx.doi.org/10.1016/S0145-2134(97)00174-9

Swahn, M. H., Whitaker, D. J., Pippen, C. B., Leeb, R. T., Teplin, L. A., Abram, K. M., & McClelland, G. M. (2006). Concordance between self-reported maltreatment and court records of abuse or neglect among high-risk youths. *Am J Public Health*, *96*(10), 1849-1853. https://doi.org/10.2105/ajph.2004.058230

U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth, and Families, Children's Bureau. (2022). *Child Maltreatment 2020* Retrieved from https://www.acf.hhs.gov/cb/report/child-maltreatment-2020

Vanderweele, T. J., Hong, G., Jones, S. M., & Brown, J. L. (2013). Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. *J Am Stat Assoc*, *108*(502), 469-482. https://doi.org/10.1080/01621459.2013.779832

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*(1), 18-37. https://doi.org/10.1037/a0015917

Widom, C. S. (1988). Sampling biases and implications for child abuse research. *American Journal of Orthopsychiatry*, *58*(2), 260-270.

Widom, C. S., DuMont, K., & Czaja, S. J. (2007). A prospective investigation of major depressive disorder and comorbidity in abused and neglected children grown up. *Archives of General Psychiatry*, *64*(1), 49-56.

Widom, C. S., & Shepard, R. L. (1996). Accuracy of adult recollections of childhood victimization: Part I: Childhood physical abuse. *Psychological Assessment*, *8*, 412-421.