Journal of Family Theory & Review

Kristin Turney *University of California, Irvine*

Beyond Average Effects: Incorporating Heterogeneous Treatment Effects Into Family Research

Family researchers are increasingly concerned with causal inference. In this article, I urge family researchers to consider 2 types of causal inference: pretreatment heterogeneity, a consideration of nonrandom selection into a treatment (e.g., divorce), and posttreatment heterogeneity, a consideration of systematic differential responses to a treatment. I detail the heterogeneous treatment effects approach, a method designed to account for both pretreatment heterogeneity and posttreatment heterogeneity. I then review existing research that has implemented this method, paying particular attention to research on family life. Finally, I provide concrete examples of how family researchers can implement heterogeneous treatment effects to answer key research questions.

Demographic changes in family life—including the postponing of marriage, the rising number of short-term cohabiting unions, the increasing divorce rates, and the decoupling of marriage from childbearing—signify that the landscape of the American family has changed dramatically in the past 50 years (Bianchi, 2014; Cherlin, 2010; Furstenberg, 2014). These demographic changes mean that considerable numbers of adults and children experience lives characterized by family instability. Given the absolute number of individuals affected by family instability, together with the concentration of family instability among racial/ethnic minorities and those with low levels of educational attainment, family scholars have developed a heightened interest in understanding the intra- and intergenerational consequences of family instability (Cavanagh & Huston, 2006; Fomby & Cherlin, 2007; Osborne, Berger, & Magnuson, 2012; for reviews, see Amato, 2000, 2001, 2010; Amato & Keith, 1991; Demo & Fine, 2010; McLanahan, Tach, & Schneider, 2013).

Although family scholars have recently turned their attention to complex indicators of instability—including remarriage (Sweeney, 2010), multipartner fertility (Guzzo, 2014), and incarceration (Turney, 2014b)—divorce is perhaps the most commonly studied measure of instability. Although substantial variation exists in individuals' reactions to divorce, with some adjusting immediately and others adjusting over a period of time, research shows that, on average, divorce can be deleterious for both adults and children. Among adults, divorce is associated with increased mental health problems (Johnson & Wu, 2002; Wade & Pevalin, 2004), heightened economic insecurity (Osborne et al., 2012), and parenting impairments (Kalmijn, 2013). Children who experience parental divorce, compared to their counterparts who do not, have lower academic achievement and attainment (Amato & Anthony, 2014; Kim, 2011), more behavioral

University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100 (kristin.turney@uci.edu).

Key Words: Child well-being, divorce, heterogeneous treatment effects, quantitative methods.

DOI:10.1111/jftr.12114

and mental health problems (Carlson & Corcoran, 2001; Cherlin, Chase-Lansdale, & McRae, 1998; Strohschein, McDonough, Monette, & Shao, 2005), and greater family instability as adults (Hill, Yeung, & Duncan, 2001; Wolfinger, 2000). Given the steep threats to causal inference associated with nonrandom selection into divorce, research increasingly employs sophisticated analytic methods to consider the causal effects of divorce on adults and children (Amato & Anthony, 2014; Cooper, Osborne, Beck, & McLanahan, 2011; Osborne et al., 2012; for a recent review of research that relies on sophisticated techniques for establishing causal inference, see McLanahan et al., 2013).

But fully understanding the causal intra- and intergenerational consequences of divorce (and the consequences of family instability more broadly), as well as designing effective interventions that prevent the transmission of inequality both within and across generations, necessitates a comprehensive identification of both pretreatment heterogeneity, the idea that some individuals are more likely than others to divorce, and posttreatment heterogeneity, the idea that individuals may differ in their responses to divorce (Morgan & Winship, 2007). Indeed, it is implausible to assume that all adults and children have similar reactions to divorce, and considering such variation in responses, within a causal framework, is an important direction for future research. Divorce may be a deleterious life course event for some adults and children, a beneficial life course event for others, and a mostly inconsequential life course event for still others (Demo & Fine, 2010; Jaffee, Moffitt, Caspi, & Taylor, 2003).

Given the increasing attention to causal inference in family research, as well as the importance of considering both pretreatment heterogeneity and posttreatment heterogeneity in estimates of causal inference, family scholars should regularly consider implementing heterogeneous treatment effects into their research designs. Heterogeneous treatment effects allow for a consideration of the effects of a treatment (e.g., divorce) by the propensity for experiencing the treatment (e.g., characteristics associated with the likelihood of divorce, such as income, mental health, and relationship conflict) (Brand & Thomas, 2013; Xie, Brand, & Jann, 2012; see also Morgan & Winship, 2007). Conceptually, the heterogeneous treatment effects approach is similar to estimating moderating effects (e.g., including interaction terms in regression models), but the heterogeneous treatment effects approach allows researchers to simultaneously consider many moderating characteristics. This analytic approach considers whether the effects of a treatment are strongest among individuals with a low propensity for experiencing the treatment (e.g., those unlikely to divorce because of their background characteristics) or among individuals with a high propensity for experiencing the treatment (e.g., those likely to divorce because of their background characteristics).

Heterogeneous treatment effects have been recently used to study the effects of educational attainment (Bauldry, 2014; Brand, 2010; Brand & Davis, 2011; Brand & Xie, 2010; Musick, Brand, & Davis, 2012; Schafer, Wilkinson, & Ferraro, 2013), but despite their ability to document complex causal relationships and their broad applicability, they have been underutilized in family research (but see Brand & Davis, 2011; Brand & Thomas, 2014; Musick et al., 2012; Turney, 2014a). Importantly, this approach considers only observed heterogeneity, and not *unobserved* heterogeneity, a point I return to when discussing the strengths and limitations of this approach.

In this article, I argue that family research could benefit from implementing the heterogeneous treatment effects approach. To illustrate this point, I document how one research program—research on the effects of parental divorce for children's academic, behavioral, and social outcomes—can routinely incorporate heterogeneous treatment effects. The remainder of this article proceeds as follows. First, I describe the importance of causal inference in family studies research. Second, I briefly summarize existing research on the effects of parental divorce on children's outcomes, focusing on research that uses methods designed to strengthen causal inference. Third, I detail heterogeneous treatment effects, a method designed to account for both pretreatment heterogeneity and posttreatment heterogeneity that has been described elsewhere (Brand & Thomas, 2013; Xie et al., 2012). Fourth, I review existing research that has implemented this method, paying particular attention to research on family life. Finally, I provide concrete examples of how family researchers can implement heterogeneous treatment effects. In that section, I first focus on research linking divorce to child well-being. But as these methods are broadly applicable to family studies research, I provide additional examples of research areas that could benefit from implementing this method. I conclude by discussing the implications of this approach.

Causal Inference in Family Studies Research

Social science research is increasingly concerned with the counterfactual model of causality (Gangl, 2010; Morgan, 2013; Morgan & Winship, 2007; Pearl, 2000; Rubin, 1974, 1977; Winship & Morgan, 1999), and social science research on family life is no exception (McLanahan et al., 2013). Embedded in the counterfactual, or potential outcomes, model of causality is the idea that an individual can be exposed to two alternative states—a treatment state and a control state—but that an individual can be observed in only one of these states. Each individual exposed to the treatment has a potential outcome under the control state, and each individual exposed to the control has a potential outcome under the treatment state (Morgan & Winship, 2007).[1] The fundamental problem of estimating the causal effect of a treatment on an outcome is that no participant will ever simultaneously experience both the treatment (e.g., parental divorce) and the counterfactual of the treatment (e.g., no parental divorce).

In what follows, I review two threats to causal inference: pretreatment heterogeneity, the fact that some individuals will be more likely than others to experience a treatment (e.g., parental divorce), and posttreatment heterogeneity, the fact that individuals may differ in their responses to the treatment (e.g., parental divorce). Pretreatment heterogeneity is an aspect of causal inference commonly considered in research on family life, but posttreatment heterogeneity is considered less regularly.

Pretreatment Heterogeneity

Family scholars pursuing causal research questions are most often concerned with selection into experiencing the treatment, often called pretreatment heterogeneity. For example, consider research estimating the effects of parental divorce on children's academic, social, and behavioral outcomes. Parental divorce is not randomly distributed across the population of children; instead, it is more commonly experienced among certain groups, such as children of parents with low levels of educational attainment (Bramlett & Mosher, 2002) and children of parents with mental health problems (Waldron, Hughes, & Brooks, 1996). Given that the factors associated with experiencing parental divorce are often also associated with children's academic, social, and behavioral skills (for research on parental educational attainment, see Davis-Kean, 2005; for research on parental mental health, see Turney, 2011), children of divorced parents are at risk of having impairments in their academic, social, and behavioral skills long before their parents become divorced (McLanahan et al., 2013). Therefore, researchers interested in the causal intergenerational consequences of divorce must, theoretically and analytically, thoroughly consider the demographic, socioeconomic, and behavioral factors associated with selection into divorce.

Posttreatment Heterogeneity

Family scholars pursuing causal research questions are also concerned with differential responses to the treatment, also known as posttreatment heterogeneity (Brand & Thomas, 2013; Morgan & Winship, 2007; Xie et al., 2012). For example, not all children respond similarly to parental divorce; some groups of children experience more deleterious consequences than other groups of children (Demo & Fine, 2010). One way to investigate posttreatment heterogeneity is to consider whether the effect of a treatment on an outcome systematically varies by the propensity, or likelihood, for experiencing the treatment (measured by a multidimensional set of pretreatment characteristics). Heterogeneous treatment effects, which consider such systematic variation, are different from the more commonly considered method of understanding variation in a relationship—estimating an interaction term between the treatment and an individual-level variable of interest—for at least three reasons. First, they document heterogeneous relationships within a causal framework (though see Breen, Choi, & Holm, 2015). Second, because

[1] There can be more than one treatment state. However, for simplicity, the examples presented refer to only one treatment state.

considering variation by the propensity score necessitates including a host of pretreatment characteristics (as opposed to just one pretreatment characteristic), this analytic approach more accurately reflects the social realities of individuals' lives. Third, unlike when testing an array of interaction effects, considering a summary measure of pretreatment characteristics avoids exhausting model degrees of freedom (Brand & Xie, 2010).

Research Considering the Effects of Divorce on Child Well-Being

Understanding the causal effects of divorce on child well-being has become of increasing interest to family scholars (McLanahan et al., 2013). Indeed, this is a thorny issue because divorce is not a well-defined causal state (McLanahan et al., 2013; Morgan & Winship, 2007). Divorce is not a discrete event but a process that unfolds over time (Amato, 2000; Demo & Fine, 2010; Hetherington, 1979; Kim, 2011; Morrison & Cherlin, 1995), which means that the effects of divorce may be apparent before a divorce is legally finalized. Therefore, all family scholars studying the effects of divorce on child well-being must address the issue of when the treatment of divorce begins and when the treatment of divorce ends.

Despite this theoretically and analytically thorny issue, researchers have increasingly begun to employ sophisticated analytic techniques to estimate the causal effects of parental divorce on children's academic, behavioral, and social outcomes (for a review, see McLanahan et al., 2013; see also Amato, 2010). There are a variety of methodological approaches for estimating the causal effect of divorce, nearly all of which require longitudinal data. A lagged dependent variable approach is one such research design. This approach estimates an outcome (e.g., behavioral problems) as a function of parental divorce, adjusting for a variety of control variables and a lagged dependent variable (e.g., a measure of behavioral problems before parental divorce). In this case, any relationship between parental divorce and children's behavioral problems is net of prior behavioral problems (Johnson, 2005; for an example of research using this approach, see Cherlin et al., 1991). Another research design commonly used for estimating causal effects is an individual-level fixed-effects approach. This approach estimates an outcome (e.g., behavioral problems) as a function of parental divorce, adjusting for both observed and unobserved time-invariant controls (e.g., characteristics that do not change over time, such as race or gender). This approach considers how within-child changes in parental divorce are associated with changes in outcomes, with each participant essentially serving as his or her own control (Cooper et al., 2011; Ribar, 2004). For a more detailed discussion of the strengths and limitations of these and other approaches when estimating the causal effects of divorce, see McLanahan et al. (2013).

Parental divorce can have an array of deleterious, beneficial, and inconsequential effects on children. Existing research on the effects of divorce on child well-being suggests that, on average, divorce is associated with negative effects on children's academic, social, and behavioral outcomes (for reviews, see Amato, 2000, 2001, 2010; Amato & Keith, 1991; Demo & Fine, 2010; McLanahan et al., 2013). A meta-analysis (Amato, 2001) showed that children of divorce, compared to their counterparts with continuously married parents, have lower self-concept (effect size, ES = –.12), social relations (ES = –.15), academic achievement (ES = –.16), psychological adjustment (ES = –.21), and conduct (ES = –.22). An earlier meta-analysis showed similar effect sizes of between one-tenth and one-fourth of a standard deviation (Amato & Keith, 1991).

However, research shows that divorce is not a uniformly negative experience for all children. Children vary greatly in their responses (Amato, 2000). The most commonly considered moderators of the relationship between divorce and children's academic, behavioral, and social outcomes include race (Broman, Li, & Reckase, 2008; Heard, 2007; Sun & Li, 2007; Wolfinger, 2003), gender (Hill et al., 2001; Painter & Levine, 2000; Sun, 2001; Sun & Li, 2002), and marital conflict (Booth & Amato, 2001; Strohschein et al., 2005). Despite growing attention to the fact that the relationship between divorce and children's academic, behavioral, and social outcomes may be heterogeneous, there exists little agreement about the individual-level characteristics that condition this relationship, perhaps resulting from different operationalization of variables and populations of interest across studies. Indeed, review articles have noted that considering moderation is an important direction for future research (Amato, 2010; McLanahan et al., 2013). A heterogeneous treatment effects approach is one way to consider moderation.

Estimating Heterogeneous Treatment Effects

When estimating causal social relationships—such as the effects of parental divorce on children's academic, behavioral, and social outcomes—an ideal research design would randomly assign children to experience parental divorce. Given that this type of experimental research design, with a treatment and control group that are similar except for their experience of the treatment, is infeasible and impractical, researchers often use observational data to estimate causal relationships between two social phenomena. Matching methods, first advanced by Rosenbaum and Rubin (1983, 1984, 1985a, 1985b), are commonly used to estimate causal effects with observational data. In this section, I first review propensity score matching, which takes into account pretreatment heterogeneity, and then I review heterogeneous treatment effects, which takes into account both pretreatment and posttreatment heterogeneity.

Propensity Score Matching

Propensity score matching, a quasi-experimental and counterfactual approach for observational data, is one method for strengthening causal inference. In the absence of a true experiment, propensity score methods approximate an experimental design by facilitating a comparison between a treatment group and a control group (Morgan & Winship, 2007; Rosenbaum & Rubin, 1983). This nonparametric approach allows researchers to consider differences in outcomes between a treatment group (e.g., children who experience divorce) and a control group (e.g., children who do not experience divorce). As in a true experimental design (Singleton & Straits, 2009), the characteristics of these two groups—at least the observed characteristics of the two groups—are comparable across demographic, socioeconomic, and behavioral characteristics and differ only in their experience of the treatment.

Propensity score matching proceeds as follows. First, the researcher estimates a logistic or probit regression model to generate a propensity score, a summary measure describing the conditional probability of experiencing the treatment that ranges from 0 to 1 for each participant, as a function of a wide array of pretreatment characteristics (Rosenbaum & Rubin, 1983). Second, the researcher uses one of the following approaches to match treatment and control observations: (a) nearest neighbor matching, which matches all treatment participants to one (or more) control participants with the closest propensity score; (b) radius matching, which matches all treatment participants to control participants within a specific radius of the propensity score (or bandwidth around a propensity score); (c) kernel matching, which matches all treatment participants to control participants by weighting control participants by their distance from treatment participants; or (d) inverse probability of treatment matching, which weights the treatment and control participants so that they are similar in terms of their observed covariates (Morgan & Todd, 2008; Morgan & Winship, 2007).[2] The researcher ensures that the means of the pretreatment characteristics are statistically indistinguishable across the treatment and control groups and removes all treatment and control participants without appropriate matches.[3] The mean values of the pretreatment characteristics in the treatment group are compared to the mean values of the characteristics in the control group. These values are compared nonparametrically with standardized differences (i.e., Cohen's d) (and not with the regression parameter). Third, the researcher estimates the effect of the treatment on the outcome (e.g., the effect of parental divorce on children's externalizing behaviors).

Although propensity score matching is increasingly used in family studies research, there is no consensus for implementing some aspects of this method, and therefore the researcher has to make several decisions. First, the researcher must decide which functional form to use to estimate the propensity score (e.g., logit, probit). This decision may not be too

[2] Because there may be subtle differences in the treatment and control groups after matching, researchers can conduct doubly robust propensity score matching analyses by further adjusting for all pretreatment characteristics (Schafer & Kang, 2008).

[3] Therefore, if it is not possible to find appropriate matches for a substantial portion of participants in the treatment group, propensity score matching is likely not the best method of analysis.

consequential because research shows that, in the case of binary treatments, the logit and probit models come to similar conclusions (Caliendo & Kopeinig, 2008). Second, the researcher must decide which covariates to include in the equation estimating the propensity score. There exists no agreed-on method for choosing the relevant pretreatment characteristics, although the goal is usually to include variables associated with the treatment (or with the treatment and outcome), to ensure that the variables explain a relatively large portion of the variance in the treatment, and to ensure that the pretreatment characteristics are similar across the treatment and control groups (Caliendo & Kopeinig, 2008).[4] Matching can be especially challenging when using multiply imputed data sets, as the same covariates may not balance across all of the imputed data sets. Third, although there is some evidence that kernel matching performs best with a well-specified propensity score—one that includes all relevant predictors of the treatment—and nearest-neighbor matching performs best with a poorly specified propensity score, there is little consensus about which matching strategy is best (Heckman, Ichimura, & Todd, 1997; Morgan & Harding, 2006). Indeed, a simulation reveals different treatment effects across matching estimators and software programs (Morgan & Harding, 2006); therefore, different matching strategies could produce different inferences about the causal effect of divorce on children's outcomes. Fourth, there is little standardization of how standard errors are estimated, which may explain why different software programs come to different conclusions (Gangl, 2010; see also Caliendo & Kopeinig, 2008).

Heterogeneous Treatment Effects

The heterogeneous treatment effect approach also operates in the counterfactual framework (Xie et al., 2012; though see Breen et al., 2015). This approach extends the propensity score model by considering how the effect of the treatment on an outcome varies by the observed propensity for experiencing the treatment. Therefore, this approach, unlike other approaches for causal inference (e.g., individual fixed effects) simultaneously allows the researcher to consider both heterogeneous pathways into experiencing the treatment (pretreatment heterogeneity) and heterogeneous effects of the treatment (posttreatment heterogeneity). For example, this approach considers the extent to which individual factors such as race/ethnicity, income, and marital conflict are associated with both the likelihood that a child experiences divorce and the child's responses to divorce.

The heterogeneous treatment effect approach takes three steps (and can be estimated with the –hte– Stata module; Jann, Brand, & Xie, 2007; see also Becker & Ichino, 2002). First, similar to propensity score matching, for each participant, a researcher estimates a propensity score, which can be considered a multidimensional measure of advantages and disadvantages in social background (Brand & Thomas, 2013; Xie et al., 2012). An additional related step in estimating heterogeneous treatment effects is to group participants into strata based on their propensity score. Conceptually, creating the strata allows the researcher to systematically consider differences in the effect of a treatment (e.g., parental divorce) based on the likelihood of experiencing the treatment. This makes it possible to compare participants who have a low propensity for experiencing the treatment (e.g., those children unlikely to experience parental divorce) and participants who have a high propensity for experiencing the treatment (e.g., those children likely to experience parental divorce). Although there may be a theoretical reason for creating a certain number of strata, the final number of strata usually depends on the number of participants in each stratum and the natural cut points of the propensity scores (Brand & Xie, 2010; Xie et al., 2012). Importantly, the researcher must ensure that, within each stratum, the means of the pretreatment characteristics are statistically indistinguishable between the treatment and control groups; the treatment and control groups must differ only in their experience of the treatment. This can be an especially difficult step, as it is not always possible to achieve balance with the same set of covariates used to achieve balancing in propensity score matching. In this case, the researcher may need to exclude some covariates from the propensity

[4]In estimating the effects of divorce on children's academic, behavioral, and social outcomes, one would want to include characteristics associated with the probability of divorce such as race/ethnicity, income, and marital quality.

score matching equation or include squared or quadratic terms in the equation (Morgan & Winship, 2007). Participants who do not have a proper match are excluded from the analyses.[5]

Second, the researcher uses kernel matching to estimate the effect of the treatment on the outcome (e.g., the effect of parental divorce on children's externalizing behaviors) within each stratum (Brand & Thomas, 2013; Xie et al., 2012). This is the Level 1 effect.[6]

Third, the researcher estimates the trend in the variation of effects across propensity score strata with a variance weighted-least-squares regression (Brand & Thomas, 2013; Xie et al., 2012). This is the Level 2 effect. The Level 2 effect tests the direction, magnitude, and statistical significance of variation in the effect of the treatment across propensity score strata. A positive, significant Level 2 slope means that, for each unit change in strata (i.e., moving up or down a strata), there is an increase in the effect of the treatment on the outcome. A negative, significant coefficient means that there is a decrease in the effect of the treatment on the outcome for each unit change in strata. Therefore, it is possible to examine whether the effect of the treatment is different among participants with a relatively low propensity of experiencing the treatment and participants with a relatively high propensity of experiencing the treatment.[7]

Therefore, although there are several similarities between propensity score matching and heterogeneous treatment effects, there are a number of differences between the two analytic strategies. First, and perhaps obviously, propensity score matching is used to estimate average treatment effects, and heterogeneous treatment effects are used to estimate heterogeneous treatment effects. Second, and relatedly, the models differ in their comparison groups. In propensity score matching, a comparison is made between the treatment group and the control group across a population. In heterogeneous treatment effects (sometimes called multi-level propensity score matching), a comparison is first made between the treatment group and the control group within relatively homogenous strata and then considered across the various strata (Xie et al., 2012). Therefore, employing these two analytic strategies concurrently can yield nuanced information about both the average and heterogeneous effects of a treatment (Brand & Thomas, 2014).

Advantages and Disadvantages of Matching Methods

Both propensity score matching and heterogeneous treatment effects have several advantages. First, the ability to match participants on observed covariates is especially important when there is substantial pretreatment heterogeneity. This advantage is further strengthened when researchers have a wide array of covariates—including demographic, socioeconomic, and behavioral characteristics—at their disposal. Second, unlike traditional regression models, this method allows researchers to exclude participants without an appropriate match from the analyses. Third, unlike traditional regression models, propensity score matching does not impose a functional form on how the control variables are associated with the treatment. Finally, the heterogeneous treatment effects are especially advantageous because they allow for an estimation of effect heterogeneity in the counterfactual framework.

[5] The following equation estimates the propensity score:

$$P_i = p\left(d_i = 1|\mathrm{X}\right) = log\ \tfrac{d_i}{1-d_i} = \left(\sum_{k=0}^{K} \beta_k \chi_{ik}\right).$$

This equation is the familiar log odds, where P_i is the propensity score for the ith individual (e.g., the log odds of experiencing the treatment), which is modeled as a linear function of the X variables (e.g., pretreatment covariates) and an estimated set of regression weights.

[6] The following equation estimates the Level 1 (stratum-specific) effects:

$$\widehat{TT} = {}^{1}\!/_{n_i} \sum_{i}^{n_i} \left(y_{i,d=1} - \sum_{i(j)}^{iJ} w_{i(j)} y_{i(j),d=0} \right).$$

This equation estimates the effect of the treatment on the outcome, where n_i is the number of treatment cases, i is the index over treatment cases, $i(j)$ is the index over control cases for treatment case i, and $w_{i(j)}$ is the scaled weight (with sum of 1) measuring the importance of each control case. Therefore, the summation $\sum$ adjusts for the probability of being a treatment case. This formula is applied sequentially to each stratum.

[7] The following equation estimates the Level 2 effect: $\delta_s = \hat{\delta}_1 + \gamma S + \epsilon_s$. In this equation, Level 1 slopes δ_s are regressed on propensity score rank indexed by S, $\hat{\delta}_1$ represents the Level 2 intercept (i.e., the predicted value of the effect of the treatment for the lowest propensity score stratum), and γ represents the Level 2 slope (i.e., change in the effect of the treatment on the outcome based on change in propensity score stratum).

However, these methods are not without disadvantages. First, propensity score matching can be implemented only when the key explanatory variable is dichotomous (although it is possible to estimate multiple dichotomous treatments). Additionally, because the analyses rest on balancing participants in the treatment and control groups, the two groups must have considerable overlap in their observed characteristics. Finally, and perhaps most important, both regular propensity score matching and heterogeneous treatment effects proceed under ignorability, the assumption that all relevant determinants of the treatment are used to generate the propensity score (Morgan & Harding, 2006; Shadish, 2013). This means that these methods, unlike some other strategies for dealing with causal inference (e.g., individual-level fixed effects), do not account for unobserved characteristics of individuals. It is quite possible, then, that unobserved characteristics could render the relationship between the treatment and outcome spurious. However, there are several ways researchers can address unobserved heterogeneity. For one, researchers can rely on data sources that include rich and detailed demographic, socioeconomic, and behavioral measures (Shadish, 2013). Also, although it is not possible to directly test the plausibility of the ignorability assumption, it is possible to examine just how large an unobserved variable would have to be to confound the relationship between the treatment and the outcome (Rosenbaum, 2002).

Finally, there are some disadvantages specific to heterogeneous treatment effects. For one, although the propensity matching framework does allow researchers to consider how large an unobserved variable would have to be to suggest a spurious relationship, it is not possible to determine whether an unobserved factor would render the trend line (Level 2 effect) statistically insignificant. Second, heterogeneous treatment effects are limited to considering variation across stratum and do not readily allow for variation within stratum. Third, the trend line (Level 2 effect) cannot be estimated across multiply imputed data sets, so analyses are limited to a single imputed data set, participants with no missing values, or a manual calculation of the coefficients and standard errors across multiply imputed data sets. Finally, as a practical matter, there needs to be a large enough number of participants experiencing the treatment to allow for at least 20 treatment cases in each stratum (Xie et al., 2012).

Examples of Research Considering Heterogeneous Treatment Effects

Heterogeneous treatment effects have been most commonly applied to research considering the effects of educational attainment on later life course outcomes. Therefore, in this section, I first review examples of research that estimates the heterogeneous effects of educational attainment. I then review the smaller body of research in family studies that have applied these methods.

Heterogeneous Treatment Effects of Educational Attainment

Heterogeneous treatment effects have been developed recently and have been most commonly used to examine the effect of educational attainment on a variety of outcomes, including wages (Brand & Xie, 2010; Hu & Hibel, 2013), civic participation (Brand, 2010), family formation (Brand & Davis, 2011), fertility (Musick et al., 2012), and health (Bauldry, 2014; Hu, 2014; Schafer et al., 2013). Broadly speaking, this burgeoning literature adjudicates between two hypotheses: (a) the negative selection hypothesis, which posits that the positive effects of educational attainment are strongest for individuals least likely to attain that educational attainment (e.g., those with a low propensity, or likelihood, for educational attainment), and (b) the positive selection hypothesis, which posits that the positive effects of educational attainment are strongest for individuals most likely to attain that educational attainment (e.g., those with a high propensity, or likelihood, for educational attainment).

With some exceptions (Bauldry, 2014; Hu & Hibel, 2014), this literature finds that college attendance (Brand & Davis, 2011; Musick et al., 2012) or college completion (Brand, 2010; Brand & Xie, 2010; Hu, 2014; Schafer et al., 2013; for research on the heterogeneous effects of military service on wages, see DellaPosta, 2013) is more beneficial for individuals whose demographic and socioeconomic background characteristics make them unlikely to attend or complete college (in support of the negative selection hypothesis). For example, in perhaps the first empirical consideration of heterogeneous treatment effects, at least within the counterfactual framework, Brand and Xie (2010) used data from the Wisconsin Longitudinal Study (WLS) and the National Longitudinal Study of Youth 1979 (NLSY79) to examine the heterogeneous effects of college completion on earnings throughout the life course. They find that, across both cohorts of respondents and for both men and women, individuals least likely to obtain a college degree—based on a host of pretreatment characteristics (e.g., parental income, high school class rank, encouragement from teachers)—are those most likely to experience earnings benefits from college completion (Brand & Xie, 2010). These findings, by showing that not all individuals benefit equally from completing college, emphasize that the processes underlying social stratification are quite complicated.

Heterogeneous Treatment Effects in Family Research

Although estimating the heterogeneous effects of educational attainment is the most commonly considered application of heterogeneous treatment effects, recent research has explicitly focused on the heterogeneous treatment effects of various aspects of family life. In this section, I review two examples (for other family research considering heterogeneity outside of the heterogeneous treatment effect framework, see Amato & Anthony, 2014; Augustine & Raley, 2013; Ryan, 2012; for research on the heterogeneous effects of educational attainment on family life, see Brand & Davis, 2011; Musick et al., 2102; for research on the heterogeneous effects of program participation on marriage, see Harknett, 2006). In one recent example, Brand and Thomas (2014) use data from the NLSY79 and the National Longitudinal Survey's Child–Mother File (NLSCM) to estimate the heterogeneous effects of single mother's job displacement on children's high school completion, college attendance, college completion, and depressive symptoms in young adulthood. They find that, by and large, the deleterious effects of maternal job displacement decrease as the propensity for experiencing maternal job displacement increases. The authors suggest that stigma—which may be greatest among children unlikely to experience a single mother's job displacement—may be one mechanism linking single mothers' job displacement to deleterious outcomes in young adulthood.

In another example, Turney (2014a) used longitudinal data from the Fragile Families and Child Wellbeing Study (FFCWB) to examine the heterogeneous effects of paternal incarceration on maternal parenting (measured as neglect, psychological aggression, and physical aggression). In this research, mothers were separated into three strata: those with a low propensity (0%–5% chance) for experiencing incarceration of the focal child's father, those with a moderate propensity (5%–15% chance), and those with a high propensity (15%–79% chance). As expected, descriptive statistics show that mothers with a low propensity for experiencing paternal incarceration are generally more advantaged than those with a high propensity for experiencing paternal incarceration. For example, having postsecondary education is more common among low-propensity mothers than among moderate- or high-propensity mothers. Conversely, living in poverty is more common among high-propensity mothers. The results show that the effect of paternal incarceration on maternal neglect is heterogeneous, with the strongest effects found among mothers with a low propensity for experiencing paternal incarceration. On the implications of these findings, Turney (2014a) wrote: "These results suggest that incarceration—given its concentration among disadvantaged families and, at least in one domain, its most consequential effects for the most advantaged of these disadvantaged families—has complicated and countervailing implications for inequalities in family life" (p. 1607).

Incorporating Heterogeneous Treatment Effects Into Family Research

There are many possibilities for incorporating heterogeneous treatment effects into family research. In this section, I first describe how research on the effects of paternal divorce on children's academic, behavioral, and social outcomes, which increasingly considers pretreatment and posttreatment heterogeneity (see especially McLanahan et al., 2013), could benefit from this approach. I then describe several other substantive areas that may also consider this approach.

Heterogeneity in the Effects of Parental Divorce

Although most research on the intergenerational consequences of divorce statistically controls for elements of the social context (e.g., children's developmental stage, race/ethnicity, poverty), very little research considers the complex and multidimensional ways elements of the social context interact with divorce to influence children's academic, behavioral, and social outcomes. Theoretically, there are important reasons to consider the comprehensive set of factors that influence children's risk of—or propensity for—experiencing divorce (i.e., pretreatment heterogeneity) and how that risk systematically and differentially shapes children's academic, behavioral, and social outcomes (i.e., posttreatment heterogeneity).

On the one hand, the negative intergenerational consequences of parental divorce may be strongest among children with a relatively low propensity for experiencing divorce. Children with a low propensity for experiencing parental divorce likely endure an otherwise relatively advantaged social context prior to divorce. They generally have stable home environments with low parental conflict, are shielded from severe economic deprivation, and live in high-resource neighborhoods. They are also unlikely to have parents who experience mental health impairments (e.g., depression, alcoholism) or engage in domestic violence. It is possible that, for these children, parental divorce may be an event stressor, an unexpected life event that is especially detrimental to well-being (Eaton, 1978; Wheaton, 1982; see also Wheaton, 1990). On the other hand, the negative intergenerational consequences of parental divorce may be strongest among children with a high propensity for experiencing divorce. Children who are especially vulnerable to experiencing divorce likely do not experience divorce in isolation. Instead, prior to divorce, these children endure a complex array of disadvantages. Their vulnerable social contexts are fraught with poor quality relationship between parents, poverty, and disadvantaged neighborhood environments. Their parents have a greater than average likelihood of mental health impairments or domestic violence. Divorce may be one of many chronic stressors that emerge gradually from their social environments that can have deleterious effects on children's outcomes (Pearlin, 1989). It is also possible that this heterogeneity by children's risk of experiencing parental divorce differentially exists across developmental stages (Bronfenbrenner & Morris, 1998; Elder, 1998). Therefore, research on the effects of divorce on children's academic, behavioral, and social outcomes could benefit from a heterogeneous treatment effect approach.

Other Opportunities for Implementing Heterogeneous Treatment Effects

There are many other possibilities for rigorously implementing heterogeneous treatment effects into family research. Here I provide two additional examples of research topics that may especially benefit from a consideration of heterogeneous effects. First, consider the growing literature on the intergenerational consequences of paternal incarceration (Eddy & Poehlmann, 2010; Johnson & Easterling, 2012). This literature increasingly takes into account pretreatment heterogeneity (e.g., for individual-level fixed effects, see Geller, Cooper, Garfinkel, Schwartz-Soicher, & Mincy, 2012; Wildeman, 2010; for propensity score matching, see Turney & Haskins, 2014) but, by and large, often fails to take into account posttreatment heterogeneity (for research that considers individual-level moderators such and race/ethnicity and gender, see, e.g., Haskins, 2014; Wildeman, 2010). Theoretically, there are good reasons to expect that paternal incarceration has deleterious consequences for children's well-being, but there are also good reasons the removal of fathers from households via incarceration—particularly if the father is violent or abusing substances—may be beneficial for children (Wildeman, 2010). Implementing heterogeneous treatment effects would allow researchers to simultaneously consider the possibilities of negative, positive, or null effects (Sampson, 2011), as well as reconcile conflicting findings from qualitative research on families more broadly (Braman, 2004; Comfort, 2008; Edin, Nelson, & Paranal, 2004; Giordano, 2010).

Another literature for which heterogeneous treatment effects may be particularly instructive is research that considers the relationship between grandparent coresidence and child well-being (Dunifon, 2013; for an alternative way of considering heterogeneity, see Augustine & Raley, 2013). This research often comes to inconsistent conclusions, with some research suggesting beneficial effects of grandparent coresidence (DeLeire & Kalil, 2002) and other research suggesting detrimental effects of grandparent coresidence (Foster & Kalil, 2007). It may be that the relationship between grandparent coresidence and child well-being varies by the social factors that are associated with selection into grandparent coresidence. Indeed, the circumstances that precede grandparent coresidence are quite varied. Some grandparents move in with their adult children because of their own failing health or an inability to economically support themselves, which may lead to negative outcomes for their grandchildren. Other grandparents move in with their adult children because of their children's needs for financial, emotional, or instrumental support, and in these cases, grandparent coresidence may be beneficial for grandchildren (Dunifon, 2013). These research programs—the effects of paternal incarceration on child well-being and the effects of grandparent coresidence on child well-being—are two examples of research that may benefit from incorporating heterogeneous treatment effects into analytic strategies. But there are many other applications of this method in family research, including research on the effects of teenage childbearing on adult and child outcomes, the effects of marital status on health outcomes, and the effects of father involvement on children.

Conclusion

Family researchers are increasingly concerned with asking and answering causal questions with observational data. Does divorce cause children to have academic difficulties? Does paternal incarceration lead to behavioral problems in children? Do children who live with their grandparents have better outcomes than children who do not live with their grandparents? Does marriage lead to better health? Family researchers have increasing become concerned with pretreatment heterogeneity, the idea that the individual background characteristics associated with experiencing the treatment are also likely associated with outcomes. But family researchers pay much less attention to posttreatment heterogeneity, the idea that individuals often have varied responses to treatments.

In this article, I have made the case for incorporating heterogeneous treatment effects, an analytic approach that takes into account both pretreatment and posttreatment heterogeneity, when answering causal questions about family life. By moving beyond an examination of the average causal effects of measures of family life (e.g., parental divorce), and by considering the nuances inherent in the consequences of these measures, this research will have implications for social policies and interventions. For example, understanding heterogeneity in the consequences of paternal divorce across population subgroups will provide an understanding about which children most need and will most benefit from interventions and, therefore, provide guidance about how to allocate resources. The result will be findings that document a more complex—and potentially countervailing—set of relationships regarding inequality in family life.

References

Amato, P. R. (2000). The consequences of divorce for adults and children. *Journal of Marriage and the Family, 62*(4), 1269–1287.

Amato, P. R. (2001). Children of divorce in the 1990s: An update of the Amato and Keith (1991) meta-analysis. *Journal of Family Psychology*, *15*(3), 355–370.

Amato, P. R. (2010). Research on divorce: Continuing trends and new developments. *Journal of Marriage and Family*, *72*(3), 650–666.

Amato, P. R., & Anthony, C. J. (2014). Estimating the effects of parental divorce and death with fixed effects models. *Journal of Marriage and Family*, *76*(2), 370–386.

Amato, P. R., & Keith, B. (1991). Parental divorce and the well-being of children: A meta-analysis. *Psychological Bulletin*, *110*(1), 26–46.

Augustine, J. M., & Raley, K. R. (2013). Multigenerational households and the school readiness of children born to unmarried mothers. *Journal of Family Issues*, *34*(4), 431–459.

Bauldry, S. (2014). Conditional health-related benefits of higher education: An assessment of compensatory versus accumulative mechanisms. *Social Science & Medicine*, *111*, 94–100.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal*, *2*(4), 358–377.

Bianchi, S. M. (2014). A demographic perspective on family change. *Journal of Family Theory & Review*, *6*(1), 35–44.

Booth, A., & Amato, P. R. (2001). Parental predivorce relations and offspring postdivorce well-being. *Journal of Marriage and Family*, *63*(1), 197–212.

Braman, D. (2004). *Doing time on the outside: Incarceration and family life in urban America*. Ann Arbor, MI: University of Michigan Press.

Bramlett, M. D., & Mosher, W. D. (2002). Cohabitation, marriage, divorce, and remarriage in the United States. *National Center for Health Statistics: Vital Health Statistics*, *23*(22), 1–32.

Brand, J. E. (2010). Civic returns to higher education: A note on heterogeneous effects. *Social Forces*, *89*(2), 417–433.

Brand, J. E., & Davis, D. (2011). The impact of college education on fertility: Evidence for heterogeneous effects. *Demography*, *48*(3), 863–887.

Brand, J. E., & Thomas, J. S. (2013). Causal effect heterogeneity. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 189–213). New York, NY: Springer.

Brand, J. E., & Thomas, J. S. (2014). Job displacement among single mothers: Effects on children's outcomes in young adulthood. *American Journal of Sociology*, *119*(4), 955–1001.

Brand, J. E., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, *75*(2), 273–302.

Breen, R., Choi, S., & Holm, A. (2015). Heterogeneous causal effects and sample selection bias. *Sociological Science*, *2*, 351–369.

Broman, C. L., Li, X., & Reckase, M. (2008). Family structure and mediators of adolescent drug use. *Journal of Family Issues*, *29*(12), 1625–1649.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Volume 1. Theoretical models of human development* (pp. 993–1028). New York, NY: Wiley.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72.

Carlson, M. J., & Corcoran, M. E. (2001). Family structure and children's behavioral and cognitive outcomes. *Journal of Marriage and Family*, *63*(3), 779–792.

Cavanagh, S. E., & Huston, A. C. (2006). Family instability and children's early problem behavior. *Social Forces*, *85*(1), 551–581.

Cherlin, A. J. (2010). Demographic trends in the United States: A review of research in the 2000s. *Journal of Marriage and Family*, *72*(3), 403–419.

Cherlin, A. J., Chase-Lansdale, P. L., & McRae, C. (1998). Effects of parental divorce on mental health throughout the life course. *American Sociological Review*, *63*(2), 239–249.

Cherlin, A. J., Furstenberg, F. F., Chase-Lansdale, L., Kiernan, K. E., Robins, P. K., Morrison, D. R., & Teitler, J. O. (1991). Longitudinal studies of effects of divorce on children in Great Britain and the United States. *Science*, *252*(5011), 1386–1389.

Comfort, M. (2008). *Doing time together: Love and family in the shadow of the prison*. Chicago, IL: University of Chicago Press.

Cooper, C. E., Osborne, C., Beck, A. N., & McLanahan, S. S. (2011). Partnership instability, school readiness, and gender disparities. *Sociology of Education*, *84*(3), 246–259.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304.

DeLeire, T., & Kalil, A. (2002). Good things come in threes: Single-parent multigenerational family structure and adolescent adjustment. *Demography*, *39*(2), 393–413.

DellaPosta, D. J. (2013). The heterogeneous economic returns to military service: Evidence from the Wisconsin longitudinal study. *Research in Social Stratification and Mobility*, *34*, 73–95.

Demo, D. H., & Fine, M. A. (2010). *Beyond the average divorce*. Thousand Oaks, CA: Sage.

Dunifon, R. (2013). The influence of grandparents on the lives of children and adolescents. *Child Development Perspectives*, *7*, 55–60.

Eaton, W. W. (1978). Life events, social supports, and psychiatric symptoms: A re-analysis of the New Haven data. *Journal of Health and Social Behavior*, *19*(2), 230–234.

Eddy, J. M., & Poehlmann, J. (2010). *Children of incarcerated parents: A handbook for researchers and practitioners*. Washington, DC: Urban Institute Press.

Edin, K., Nelson, T. J., & Paranal, R. (2004). Fatherhood and incarceration as potential turning points in the criminal careers of unskilled men. In M. Patillo, D. Weiman, & B. Western (Eds.), *Imprisoning America: The social effects of mass incarceration* (pp. 46–75). New York, NY: Russell Sage Foundation.

Elder, G. H. (1998). The life course as developmental theory. *Child Development*, *69*(1), 1–12.

Fomby, P., & Cherlin, A. (2007). Family instability and child well-being. *American Sociological Review*, *72*, 181–204.

Foster, E. M., & Kalil, A. (2007). Living arrangements and children's development in low-income White, Black, and Latino families. *Child Development*, *78*(6), 1657–1674.

Furstenberg, F. F. (2014). Fifty years of family change: From consensus to complexity. *Annals of the American Academy of Political and Social Science*, *654*(1), 12–30.

Gangl, M. (2010). Causal inference in sociological research. *Annual Review of Sociology*, *36*, 21–47.

Geller, A., Cooper, C. E., Garfinkel, I., Schwartz-Soicher, O., & Mincy, R. B. (2012). Beyond absenteeism: Father incarceration and child development. *Demography*, *49*(1), 49–76.

Giordano, P. C. (2010). *Legacies of crime: A follow-up of the children of highly delinquent girls and boys*. New York, NY: Cambridge University Press.

Guzzo, K. B. (2014). New partners, more kids: Multiple-partner fertility in the United States. *Annals of the American Academy of Political and Social Science*, *654*(1), 66–86.

Harknett, K. (2006). Does receiving an earnings supplement affect union formation? Estimating effects for program participants using propensity score matching. *Evaluation Review*, *30*(6), 741–778.

Haskins, A. R. (2014). Unintended consequences: Effects of paternal incarceration on child school readiness and later special education placement. *Sociological Science*, *1*, 141–158.

Heard, H. E. (2007). The family structure trajectory and adolescent school performance: Differential effects by race and ethnicity. *Journal of Family Issues*, *28*(3), 319–356.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, *64*(4), 6059–654.

Hetherington, E. M. (1979). Divorce: A child's perspective. *American Psychologist*, *34*(10), 851–858.

Hill, M. S., Yeung, W. J. J., & Duncan, G. J. (2001). Childhood family structure and young adult behaviors. *Journal of Population Economics*, *14*(2), 271–299.

Hu, A., & Hibel, J. (2014). Changes in college attainment and the economic returns to a college degree in urban China, 2003–2010: Implications for social equality. *Social Science Research*, *44*, 173–186.

Hu, A. (2014). The health benefits of college education in urban China: Selection bias and heterogeneity. *Social Indicators Research*, *115*(3), 1101–1121.

Jaffee, S. R., Moffitt, T. E., Caspi, A., & Taylor, A. (2003). Life with (or without) father: The benefits of living with two biological parents depend on the father's antisocial behavior. *Child Development*, *74*(1), 109–126.

Jann, B., Brand, J. E., & Xie, Y. (2007). *HTE: Stata module to perform heterogeneous treatment effect analysis*. Boston, MA: Boston College Department of Economics.

Johnson, D. (2005). Two-wave panel analysis: Comparing statistical methods for studying the effects of transitions. *Journal of Marriage and Family*, *67*(4), 1061–1075.

Johnson, D. R., & Wu, J. (2002). An empirical test of crisis, social selection, and role explanations of the relationship between marital disruption and psychological distress: A pooled time-series analysis of four-wave panel data. *Journal of Marriage and Family*, *64*(1), 211–224.

Johnson, E. I., & Easterling, B. (2012). Understanding unique effects of parental incarceration on children: Challenges, progress, and recommendations. *Journal of Marriage and Family* *74*(2), 342–356.

Kalmijn, M. (2013). Adult children's relationships with married parents, divorced parents, and stepparents: Biology, marriage, or residence? *Journal of Marriage and Family*, *75*(5), 1181–1193.

Kim, H. S. (2011). Consequences of parental divorce for child development. *American Sociological Review*, *76*(3), 487–511.

McLanahan, S., Tach, L., & Schneider, D. (2013). The causal effects of father absence. *Annual Review of Sociology*, *39*, 399–427.

Morgan, S. (2013). *Handbook of causal analysis for social research*. New York, NY: Springer.

Morgan, S. L., & Harding, D. J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods & Research*, *35*(1), 3–60.

Morgan, S. L., & Todd, J. J. (2008). A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociological Methodology*, *38*(1), 231–281.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York, NY: Cambridge University Press.

Morrison, D. R., & Cherlin, A. J. (1995). The divorce process and young children's well-being: A prospective analysis. *Journal of Marriage and the Family*, *57*(3), 800–812.

Musick, K., Brand, J. E., & Davis, D. (2012). Variation in the relationship between education and marriage: Marriage market mismatch? *Journal of Marriage and Family*, *74*(1), 53–69.

Osborne, C., Berger, L. M., & Magnuson, K. (2012). Family structure transitions and changes in maternal resources and well-being. *Demography*, *49*(1), 23–47.

Painter, G., & Levine, D. I. (2000). Family structure and youths' outcomes: Which correlations are causal? *Journal of Human Resources*, *35*(3), 524–549.

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

Pearlin, L. I. (1989). The sociological study of stress. *Journal of Health and Social Behavior*, *30*(3), 241–256.

Ribar, D. C. (2004). What do social scientists know about the benefits of marriage? A review of quantitative methodologies *(Discussion Paper No. 998).* Bonn, Germany: Institute for the Study of Labor (IZA).

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York, NY: Springer.

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.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*(387), 516–524.

Rosenbaum, P. R., & Rubin, D. B. (1985a). Constructing a control group using multivariate matched sampling methods. *American Statistician*, *39*(1), 33–38.

Rosenbaum, P. R., & Rubin, D. B. (1985b). The bias due to incomplete matching. *Biometrics*, *41*(1), 103–116.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational and Behavioral Statistics*, *2*(1), 1–26.

Ryan, R. M. (2012). Marital birth and early child outcomes: The moderating influence of marriage propensity. *Child Development*, *83*(3), 1085–1101.

Sampson, R. J. (2011). The incarceration ledger: Toward a new era in assessing societal consequences. *Criminology & Public Policy*, *10*, 819–828.

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*(4), 279–313.

Schafer, M. H., Wilkinson, L. R., & Ferraro, K. F. (2013). Childhood (mis)fortune, educational attainment, and adult health: Contingent benefits of a college degree? *Social Forces*, *91*(3), 1007–1034.

Shadish, W. R. (2013). Propensity score analysis: Promise, reality and irrational exuberance. *Journal of Experimental Criminology*, *9*(2), 129–144.

Singleton, R. A., & Straits, B. C. (2009). *Approaches to social research (*fifth edition*)*. New York, NY: Oxford University Press.

Strohschein, L., McDonough, P., Monette, G., & Shao, Q. (2005). Marital transitions and mental health: Are there gender differences in the short-term effects of marital status change? *Social Science and Medicine*, *61*(11), 2293–2303.

Sun, Y. M. (2001). Family environment and adolescents' well-being before and after parents' marital disruption: A longitudinal analysis. *Journal of Marriage and Family*, *63*(3), 697–713.

Sun, Y. M., & Li, Y. Z. (2002). Children's well-being during parents' marital disruption process: A pooled time-series analysis. *Journal of Marriage and Family*, *64*(2), 472–488.

Sun, Y., & Li, Y. (2007). Racial and ethnic differences in experiencing parents' marital disruption during late adolescence. *Journal of Marriage and Family*, *69*(3), 742–762.

Sweeney, M. M. (2010). Remarriage and stepfamilies: Strategic sites for family scholarship in the 21st century. *Journal of Marriage and Family*, *72*(3), 667–684.

Turney, K. (2011). Chronic and proximate depression among mothers: Implications for child well-being. *Journal of Marriage and Family*, *73*(1), 149–163.

Turney, K. (2014a). The consequences of paternal incarceration for maternal neglect and harsh parenting. *Social Forces*, *92*(4), 1607–1636.

Turney, K. (2014b). The intergenerational consequences of mass incarceration: Implications for children's co-residence and contact with grandparents. *Social Forces*, *93*(1), 299–327.

Turney, K., & Haskins, A. R. (2014). Falling behind? Children's early grade retention after paternal incarceration. *Sociology of Education*, *87*(4), 241–258.

Wade, T. J., & Pevalin, D. J. (2004). Marital transitions and mental health. *Journal of Health and Social Behavior*, *45*(3), 155–170.

Waldron, I., Hughes, M. E., & Brooks, T. L. (1996). Marriage protection and marriage selection: Prospective evidence for reciprocal effects of marital status and health. *Social Science and Medicine*, *43*(1), 113–123.

Wheaton, B. (1982). A comparison of the moderating effects of personal coping resources on the impact of exposure to stress in two groups. *Journal of Community Psychology*, *10*(4), 293–311.

Wheaton, B. (1990). Life transitions, role histories, and mental health. *American Sociological Review*, *55*(2), 209–223.

Wildeman, C. (2010). Paternal incarceration and children's physically aggressive behaviors: Evidence from the Fragile Families and Child Wellbeing Study. *Social Forces*, *89*(1), 285–309.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659–706.

Wolfinger, N. H. (2000). Beyond the intergenerational transmission of divorce: Do people replicate the patterns of marital instability they grew up with? *Journal of Family Issues*, *21*(8), 1061–1086.

Wolfinger, N. H. (2003). Parental divorce and offspring marriage: Early or late? *Social Forces*, *82*(1), 337–353.

Xie, Y., Brand, J. E., & Jann, B. (2012). Estimating heterogeneous treatment effects with observational data. *Sociological Methodology*, *42*(1), 314–347.