Estimating the Causal Effect of Maternal Depression During Early Childhood on Child Externalizing and Internalizing Problems

by

William Pelham

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved December 2018 by the
Graduate Supervisory Committee:

Stephen G. West, Co-Chair
Keith A. Crnic, Co-Chair
Kathy Lemery-Chalfant
Sherryl H. Goodman

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

Background. Hundreds of studies have linked maternal depression to negative child outcomes. However, these studies have been correlational, so they cannot rule out alternative explanations such as that child characteristics evoke maternal depression or that confounding variables are causes of both phenomena. **Design.** I applied a propensity score approach to data from the Early Steps Multisite Trial, a sample of 731 low-income families tracked approximately annually from ages 2 through 16. Families were equated on propensity scores based on a large set of baseline characteristics, producing two groups that were similar across all measured characteristics except for the presence of clinically significant symptoms of maternal depression during toddlerhood. Children's longitudinal behavioral outcomes from parent-, teacher-, and self-report measures were compared across the equated groups in order to estimate the causal effects of maternal depression. **Results.** Both matching and weighting were successful in equating families with depressed and non-depressed mothers on a set of 89 potential confounding variables measured at child age 2. Prior to any adjustment for confounding, the effect of maternal depression was statistically significant for 41 of 48 mother-, secondary-caregiver-, and teacher-reported outcomes. Effect sizes were consistent with the larger literature and in the small to medium range. After matching or weighting to equate families with depressed versus non-depressed mothers, the effects of maternal depression at age 2 was statistically significant for 6 of 48 mother-, secondary-caregiver-, and teacher-reported outcomes. Adjusted effect sizes were in the very small to small range. **Conclusions.** Findings are consistent with the claim that there is a very small causal effect of exposure to maternal depression at child age 2 on child externalizing and internalizing behavior in

early childhood, middle childhood, and adolescence. While awaiting replication, results suggest (a) that treatment of maternal depression should not be expected to substantially reduce child externalizing and internalizing behavior problems; (b) that very large sample sizes are needed to adequately investigate causal developmental processes that link maternal depression to child behavior; and (c) that causal inference methods can be an important addition to the toolbox of developmental psychopathologists.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

PREMISE

It is widely accepted that having a depressed mother is negatively associated with children's outcomes in both the short-term and long-term. Hundreds of studies to date have indicated that the children of depressed mothers display more internalizing behavior, externalizing behavior, and negative affect (Goodman et al., 2011). However, the literature to date has been correlational, using cross-sectional, longitudinal, and mediational designs to document the presumed causal relationship between parental depression and child outcomes. Thus, the accumulated science has had limited ability to rule out the possibility that the observed relationship is causal in the opposite direction (e.g., child behavior causes maternal depression) or arises from common causes (e.g., poor marital quality, low socioeconomic status).

Statistical advances have yielded methods that can adjust for a very large number of potential confounding variables and their effects over time (Imbens & Rubin, 2015; West et al., 2014). Yet I was unable to locate any published study that has used these methodological advances to estimate the causal effect of parental depression on child outcomes. The proposed study undertakes this goal using data from the Early Steps sample, a multisite trial following 731 families assessed approximately annually between child ages 2 and 16. In short, I attempt to statistically mimic a randomized, controlled trial in which mothers were randomly assigned to levels of depression at age 2, and child outcomes were then measured repeatedly from age 3 to 16. This design permits stronger estimation of the degree to which maternal depression is damaging, with results potentially informing both developmental and intervention theory.

This introduction proceeds in three parts. First, I discuss depression in general, reviewing the most prominent theories of its origin and maintenance. Second, I discuss depression in mothers, reviewing the existing literature and identifying limitations that leave its conclusions vulnerable to alternative explanations. Third, I introduce the potential outcomes model of causal inference (Imbens & Rubin, 2015; Rubin, 1978; West & Thoemmes, 2010) and explain how it can address many of the limitations of prior research and produce causal estimates of the effects of parental depression on child outcomes.

CHAPTER 2

INTRODUCTION PART I: DEPRESSION IN GENERAL

Depression is one of the longest- and best-studied psychiatric problems, possessing a large and mature literature on its etiology, treatment, and course. An estimated 6.6% of US adults met the DSM criteria for Major Depressive Disorder (MDD) in the past year, with a lifetime prevalence of 16.2% (Kessler et al., 2003). Depressed individuals report substantial impairment in the domains of home, work, romantic relationship, and social functioning, yet only about half of those meeting diagnostic criteria receive treatment (Kessler et al., 2003). The economic burden is substantial: In the United States, depression costs an estimated $211 billion per year in direct medical expenses, suicide-related mortality, and lost work (Greenberg, Fournier, Sisitsky, Pike, & Kessler, 2015). Evidence-based treatment consists of pharmacotherapy, cognitive-behavioral therapy, or the combination thereof (Butler, Chapman, Forman, & Beck, 2006; Cuijpers et al., 2013; Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016; Fournier et al., 2010). Although this dissertation is primarily focused on depression in mothers, I begin by reviewing the most prominent theories of depression in general.

Measurement: Two conceptualizations of depression. There are two different conceptualizations that underlie the measurement of depression: (1) a continuous conception, as measured by scales that assess depressive symptomatology (e.g., BDI, Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; HDRS, Hamilton, 1960; CES-D, Radloff, 1977), and (2) a discrete conception, as measured by meeting diagnostic criteria for Major Depressive Disorder (MDD). The relation of depression with other constructs may differ (e.g., Connell & Goodman, 2002) or be similar (e.g., Goodman & Tully, 2009;

Gotlib, Lewinsohn, & Seeley, 1995) depending on which conceptualization is used. I take the position that the difference in results is an artifact of the particular symptom threshold and diagnostic criteria, such that high symptom scores and MDD reflect conceptually similar psychopathology (Boyd, Weissman, Thompson, & Myers, 1982). Differences in statistical power when predicting binary versus continuous criteria may also contribute to the reported discrepancies (Taylor, West, & Aiken, 2006). In this manuscript, I will use the term "depression" to mean depressive symptoms (in the continuous sense) and use the term "major depressive disorder" (MDD) to denote the clinical diagnosis in particular.

Theories of Depression

Dozens of models have been developed to explain the etiology and maintenance of depression, with different theories invoking cognitive, affective, behavioral, social, and biological processes in their explanations. Understanding these theories is important in designing an observational study, because we must discriminate confounding variables that need to be statistically controlled for from variables that are in fact part of the construct of the depression. I now briefly review five of the most prominent theories of depression: Beck's (1963) cognitive theory, Lewinsohn's (1974) behavioral theory, Seligman et al.'s hopelessness theory (1978), Coyne's (1976b) interactional theory, Oatley and Bolton's (1985) reaction to life events theory, and biological theories.

Cognitive theory. Beck (1963) proposed the cognitive theory of depression, under which depression is characterized by automatic, universal, and recurrent negative thoughts about the self, the environment, and the future. Depressed individuals exhibit dysfunctional cognitive patterns such as arbitrary (i.e., unsupported by evidence) negative

attributions, over-generalization of failures and setbacks, magnification of negative events, and minimization of positive events. These cognitions do not directly cause depression, but rather are "activated" by negative life experiences when the individual possesses core dysfunctional schemas (e.g., "If I'm a failure in one area of life, than I am sure to be a failure in all areas of life") (Beck, 1987; Haaga, Dyck, & Ernst, 1991). In other words, depression is caused by the interaction of core cognitive schemas and life experiences, and then maintained by depressogenic thought patterns.

Behavioral theory. Lewinsohn (1974) proposed the behavioral or learning theory of depression, under which depression is fundamentally caused by a low rate of positive reinforcement. The scarcity of positive reinforcement can arise from a combination of three different sources: (a) reduced engagement in potentially reinforcing events, (b) reduced availability of potentially reinforcing events, and (c) the person behaving in a way that reduces the amount of reinforcement obtained from his or her environment. When receiving a low rate of positive reinforcement, an individual is expected to show dysphoria, fatigue, negative thinking, and the other symptoms of depression.

Learned helplessness or hopelessness theory. Seligman, Abramson, and colleagues (Abramson et al., 1978; Alloy, Abramson, Metalsky, & Hartlage, 1988) advanced the hopelessness theory of depression, which posits that depression arises from the interaction of outcome noncontingency and individual's attributions about this noncontingency. First, the individual must notice outcome noncontingency in their environment—for example, that their work supervisor's evaluation of them is not related to the quality of work they produce. Second, the individual must attribute this

noncontingency to an uncontrollable cause—for example, that all supervisors are out to get their employees. This combination of outcome noncontingency and a recurrent pattern of making stable, global, and external attributions yields hopelessness, and thus depression.

Interactional theory. Coyne (1976b) formulated the interactional theory of depression, which suggests that depression is promoted and maintained by elicited social rejection. In this framework, depressed individuals exhibit social behaviors that are aversive and annoying to their peers, including being overly negative, inwardly-focused, and inappropriately self-disclosing. Because of these behaviors, they are more likely to experience negative social interactions in which their peers criticize them and/or withdraw from the interaction rapidly. In addition, peers are less likely to engage with the depressed individual in the future, leading to social isolation. In this way, "the depressed person and others within his social space collude to create a system in which feedback cannot be received" and depression is sustained (Coyne, 1976b).

Reaction to life events theory. Oatley and Bolton (1985) proposed the reaction to life events theory of depression, wherein depression arises from the interaction of role schemas and life events. In this theory, individuals construct a sense of self from a package of roles they hold with respect to their environment and others: for example, mother, husband, colleague, or friend. These roles can be threatened by life events, such as the incarceration of a child (for the mother role), divorce (for the husband role), or being fired (for the colleague role). When life events disrupt or remove a major role, they threaten the individual's identity and thus can result in depression.

Biological theories. Finally, there are numerous biological theories that attempt to explain depression at the level of neurochemicals, neurons, or the immune system. Prominent examples include the monoamine theory (Heninger, Delgado, & Charney, 1996), the macrophage theory (Smith, 1991), and the neurotrophic factors theory (Duman, Heninger, & Nestler, 1997). I do not review these theories further because they cannot inform the design of our observational study, which does not include biological measures among the pool of potential covariates.

CHAPTER 3

INTRODUCTION PART II: DEPRESSION IN PARENTS

General Findings on the Effects of Parental Depression

Given that these theories of the development of depression entailed frequent negative cognitions, a sense of hopelessness, and poor social functioning, it followed that the presence of depression in adults would be expected to negatively impact children in their care. The earliest published investigation of this phenomenon appears to be a monograph based on the M.D. thesis of Rutter (1966). Rutter reviewed the case files of 922 children that presented for treatment at Maudsley Hospital, a large mental health center in south London, and compared them to matched controls recruited from the pediatric and dental departments of nearby King's College Hospital. Comparison with the matched controls at King's College Hospital indicated that children with psychiatric problems were significantly more likely to have parents with psychiatric problems. Comparison within the Maudsley psychiatric sample indicated that the children of depressed parents ($N = 43$) exhibited a rate and distribution of behavioral disturbances similar to those of children of parents with schizophrenia, personality disorder, and other mental illness. Rutter believed that his findings supported a "strong association between parental [mental] illness and psychiatric disorder in children," but cautioned that "statistical associations can never prove a causal relationship" (Rutter, 1966, p. 106)—these conclusions presaged a line of research that emerged over the next half century.

Work on this topic accumulated throughout the 1970s, with several findings arising incidentally from studies of schizophrenic mothers that included depressed mothers for comparison (e.g., Grunebaum, Cohler, Kauffman, & Gallant, 1978; Rolf,

1972). By the early 1980s there were already enough studies to be formally reviewed: Beardslee and colleagues (1983) located 11 cross-sectional studies and 13 longitudinal studies reporting on the characteristics of children of parents with depression. Most studies assessed behavioral disturbance or psychopathology, assessed these outcomes via parent questionnaire or interview, and used relatively small samples. Findings indicated that having a depressed parent was associated with elevated overall rates of psychiatric diagnoses in general and of affective disorders in particular. Children of depressed parents exhibited disturbances and impairments in diverse areas including personality disorders, adjustment reactions, hyperkinetic syndromes, cognitive and emotional delays, and school functioning. Beardslee et al (1983) concluded that the risk parental depression imparted on children was significant, but that methodological limitations (e.g., absence of control groups) limited understanding of the exact magnitude and structure of that risk.

Downey and Coyne (1990) revisited this growing literature for their 1990 review, determining that many of the methodological limitations identified by Beardslee et al. (1983) had been addressed. First, the newer studies had included control groups, providing stronger evidence that the elevated rates of child problems were due to parental depression in particular. Second, several samples had evaluated child outcomes with structured interviews, verifying that the observed problems met criteria for DSM-III or Research Domain Criteria (RDC) diagnoses. Other work had extended the findings of the Beardslee et al. (1983) review. The association of parental depression and child problems was shown to be present in infancy, to be present in child reports of child problems, and to include both social and academic difficulties. In addition, an emergent

literature had begun to consider the mechanisms by which these effects occurred, with special focus on depressed parents' parenting practices. Finally, an issue that remained salient was that of specificity—the literature had not completely addressed which negative outcomes in children were due to the effects of parental depression in particular, versus parental psychiatric problems in general.

A decade later, Connell and Goodman (2002) performed the first *quantitative* review of the effects of parental depression on child outcomes. Following the recommendation of Downey and Coyne (1990) that more attention be paid to depressed fathers, Connell and Goodman (2002) compared the effects of maternal versus paternal depression on children's externalizing and internalizing behavior. The authors identified approximately 90 studies examining these associations for mothers, and approximately 60 studies examining these associations for fathers. Weighted mean effect sizes ranged from 0.14 to 0.18 in the correlation metric, varying as a function of parent (mother vs. father) and outcome (externalizing vs. internalizing). One significant difference was observed: the weighted mean effect of mothers' depression on internalizing behavior was significantly greater than that of fathers' depression on internalizing behavior (0.18 vs. 0.14). The meta-analysis included a large number of studies and yielded weighted mean effects with narrow confidence bands (e.g., [0.17, 0.19] for mother's depression on internalizing behavior), suggesting that the effects of parental depression on child outcomes were relatively small in magnitude.

Connell and Goodman (2002) also investigated whether study attributes predicted the obtained effect sizes, finding several moderating effects of interest. First, the relationship between *maternal* depression and child outcomes was weaker in samples

with a greater mean age of children, whereas the relationship between *paternal* depression and child outcomes was stronger in samples with a greater mean age of children. Second, the effects on internalizing outcomes were larger when parental depression was measured using symptom ratings, rather than diagnosis (there was no difference with respect to externalizing outcomes). Finally, across both internalizing and externalizing outcomes, effect sizes were largest when child outcomes were reported by parents, compared to either teacher or child report.

Goodman et al. (2011) revisited the literature ten years later, this time focusing only on maternal depression and broadening the child outcomes examined to include general psychopathology, negative affect, and positive affect. They identified a total of 399 independent effect sizes from 193 studies for their meta-analysis. Comparing results to the authors' earlier review (Connell & Goodman, 2002), the addition of approximately 20-30 more studies of each outcome yielded larger weighted mean effect sizes on internalizing and externalizing problems (0.23 and 0.21, respectively, in the correlation metric). Results in the new outcome domains indicated that the effect on general psychopathology was very similar to that on the internalizing/externalizing dimensions (weighted mean effect of 0.24), and that effects on positive (-0.10) and negative affect (0.15) were statistically significant but smaller.

Goodman et al.'s (2011) moderation analyses generally replicated their earlier findings, in that effect sizes were larger in samples with lower mean age of children and when child outcomes were measured via mother report. The authors also found several new moderators of interest. First, effect sizes were larger in low-income samples than in mid-, high-, or mixed-income samples. Second, effect sizes were generally larger on

girls than on boys, with the magnitude of this difference varying by outcome. In summary, Goodman et al. (2011) (a) replicated earlier findings on the robust association between maternal depression and child internalizing and externalizing behavior, (b) extended this finding to general psychopathology, positive affect, and negative affect, and (c) identified new moderators of these effects.

Specific Findings on Effects of Parental Depression in the Early Steps Sample

In addition to the extensive literature reviewed above, the effects of parental depression on child outcomes have been examined in five previous publications using the Early Steps longitudinal sample (i.e., the dataset to be used here; see Dishion et al., 2008). Most of these studies have examined the role of maternal depression as a mediator of other developmental processes, rather than the direct and immediate cause of child outcomes. Since these findings may be particularly informative for the present study, I will now review each finding briefly.

First, Weaver et al. (2008) investigated growth in parenting self-efficacy across ages 2 to 4 years, finding that maternal depression mediated the relationship between parent self-efficacy and subsequent child externalizing behavior. Higher levels of self-efficacy at age 2 were associated with reduced maternal depression at age 3, which was in turn associated with reduced externalizing behavior at age 4. Second, Gross et al. (2008) modeled the interactive influences of child noncompliance and parental depression on subsequent child internalizing and externalizing outcomes. Results indicated that directly-observed child noncompliance at child age 2 years old (i.e., not responding to parent's request for behavior change during interaction tasks) predicted increased levels of maternal and paternal depression between ages 2 and 4, which in turn were associated

with increased internalizing problems at age 4. Third, Shaw et al. (2009) investigated maternal depression as a mediator of intervention effects, rather than of naturally occurring development processes. Findings revealed that mothers in families that received the family-based intervention at child age 2 were less depressed at age 3, and that this in turn predicted reduced growth in child externalizing and internalizing problems from child age 2 to 4. Fourth, Choe et al. (2014) examined what constructs might mediate the relationship between maternal depression and child's oppositional behavior, rather than considering maternal depression itself a mediator. Modeling revealed that maternal depression predicted subsequent levels of child oppositional behavior even after controlling for prior levels of oppositional behavior, and that this effect was mediated by reductions in the child's inhibitory control. Finally, Hails et al. (2017) modeled the transactional influences of maternal depression, parent-child coercion, and child behavior problems using assessments of each construct at child ages 2, 3, and 4. A series of cross-lagged effects emerged, showing that depression, coercion, and child behavior predicted each other over time, with depression and coercion both associated with greater externalizing and internalizing problems. Each of these five studies adjusted for different covariates and did so in a different manner; path models depicting this information for each study are reproduced in the Appendix (Table A1). To summarize, existing findings in the Early Steps sample have suggested that (a) parental depression is predictive of child externalizing behavior, internalizing behavior, and inhibitory control, and (b) that parental depression mediates the effects of intervention, parental self-efficacy, and child noncompliance on subsequent child outcomes.

Mechanisms: How Does Maternal Depression Impact Child Outcomes?

The discussion of the findings above focused on *whether* maternal depression impacts child outcomes, with results from hundreds of studies indicating that maternal depression is associated with increased child internalizing problems, externalizing problems, and negative affect. Yet the discussion of Early Steps-specific findings hinted at a second major question in this literature: *how* maternal depression impacts child outcomes. I now turn to this second aim, reviewing the most prominent theory of risk transmission, as well as the specific empirical evidence supporting each hypothesized mechanism.

Integrative Model for the Transmission of Risk to Children of Depressed Mothers. Goodman and Gotlib (1999) proposed an early, prominent model (see Figure 1) of how maternal depression impacts child outcomes (see also Goodman, 2007). The model enumerates four specific risk mechanisms that can potentially account for the observed relationship, each of which will be reviewed in turn. The first proposed risk mechanism is the heritability of depression. Children of depressed mothers are more likely to have genes that place them at risk of experiencing depression and of possessing the personality, cognitive, or behavioral tendencies that can facilitate or maintain depression. In addition to this main effect, children with depressogenic genes may be more likely to select into depressogenic environments (i.e., gene-by-environment interaction). Consistent with these hypotheses, the extant literature suggests an overall heritability of depression of approximately 40% (Polderman et al., 2015; Sullivan, Neale, & Kendler, 2000).

The second proposed risk mechanism is innate dysfunctional regulatory mechanisms. Goodman and Gotlib (1999) distinguish these "innate" problems from the

heritability mechanism described above because they may reflect prenatal exposures—environmental, rather than genetic, causes. First, while in utero, children of depressed mothers are more likely to be exposed to elevated levels (in mother) of cortisol-, beta-endorphin-, and corticotrophin-releasing hormones, as well as of catecholamines, epinephrine, and norepinephrine (Goodman & Gotlib, 1999). Second, these children are also more likely exposed to their mother's poor general health habits, such as unhealthy eating, insufficient or irregular sleep, or tobacco use (Milberger, Biederman, Faraone, Chen, & Jones, 1996). The potential role of prenatal exposures is suggested by studies finding differences in the children of depressed mothers nearly immediately after birth, including higher sensory thresholds, reduced activity, and excessive crying (e.g., Field, 1992; Zuckerman, Bauchner, Parker, & Cabral, 1990).

The third proposed risk mechanism is exposure to negative maternal cognitions, behaviors, and affect. Depressed mothers are more likely to exhibit dysfunctional cognitive styles, low self-efficacy, and generally negative affect, which may be learned by children via modeling (i.e., social learning). Depressed mothers are also more likely to exhibit annoying and aversive interpersonal behaviors (Benazon & Coyne, 2000; Coyne, 1976b, 1976a; Coyne et al., 1987), which may be learned by children via modeling or simply reduce children's interest in socialization from the mother. In terms of caregiving behavior, depressed mothers may struggle to provide warmth, nurturance, and responsiveness when the child is distressed. For example, a depressed mother may have less tolerance for a child that isn't sleeping through the night and be more likely to respond in an irritable, negative manner. Finally, depressed mothers are more likely to exhibit inconsistent and ineffective discipline behaviors (Dumas, Gibson, & Albin, 1989;

Hails et al., 2017; Kochanska, Kuczynski, Radke-Yarrow, & Welsh, 1987), which may promote the development of externalizing behaviors in particular.

The fourth proposed risk mechanism is stressful life contexts. Depressed mothers are more likely to themselves experience stress in a variety of life domains, including work, relationships, education, and finance (e.g., Hammen et al., 1987). Depression is disabling and role-impairing even before considering the care of children (Judd et al., 2000; Kessler et al., 2003), and the presence of children not only adds new daily responsibilities but makes existing responsibilities more difficult. Depressed mothers display a tendency to generate stressful life events, such as by entering into conflictual romantic relationships (Hammen, 1992) or exhibiting poor management of difficult children's behavior. Finally, depressed mothers are more likely to experience marital conflict (e.g., Bruce & Kim, 1992; du Rocher Schudlich, Papp, & Cummings, 2011; O'Leary, Christian, & Mendell, 1994).

To summarize, Goodman and Gotlib (1999) proposed a model of the mechanisms through which parental depression affects child outcomes, providing empirical support for each of four proposed pathways. They further proposed that which mechanisms affect risk (and at what magnitude they do so) varies as a function of the individual case. For example, mothers who were not depressed while pregnant would presumably not confer the prenatal risks described in the second mechanism. Goodman (2007) revisited this model in an *Annual Review* chapter, reporting the accumulation of further studies supporting the proposed mechanisms.

Evidence from intervention studies. Although Goodman and Gotlib's (1999) model is not the only model of how parental depression impacts child outcomes, nearly

all of the available empirical evidence can be subsumed within its framework. Another major line of inquiry into this relationship has come from intervention studies in which parents were provided treatment for depression. For example, results of a randomized, controlled trial of intervention to prevent depressive symptoms in children of depressed parents (Beardslee, Gladstone, Wright, & Cooper, 2003) indicated that the degree of reduction in the child's symptoms was correlated with the degree in reduction of the parent's symptoms. This finding was replicated in analysis of the STAR*D trial (Weissman et al., 2006), which showed that remission in mothers' depression was associated with reduced levels of symptoms and rates of diagnoses in their children. This mediational pathway has also been examined explicitly: Earlier I reviewed an Early Steps study (Shaw et al., 2009) in which reductions in maternal depression mediated the effect of a family-based intervention on child behavior problems. Reviewing the entire literature, Gunlicks and Weissman (2008) identified 10 studies of this phenomenon. Results indicated not only that reduction in parental depression was associated with improvements in child functioning, but that these benefits were generally maintained in follow-up. Taken together, these findings suggest a potential causal role of parental depression in determining child outcomes, and mediation studies in particular have implicated various mechanisms proposed by Goodman and Gotlib (1999).

Is This Relationship Causal? Alternative Explanations

The sizeable literature reviewed above leaves little doubt that maternal depression is associated with negative child outcomes. Most authors interpret this relationship as causal (panel A in Figure 2), whether explicitly or implicitly. For example, Goodman and Gotlib (1999) write that "there is no doubt that children are adversely affected by

their mother's depression" (explicit), and Gunlicks and Weissman (2008) interpret concomitant improvements in child behavior with treatment for parental depression as the causal outcome of depression resolution (implicit). Yet the large literature to date has only minimally addressed methodological issues that might preclude a causal interpretation of results, including failure to (a) establish directionality of association and (b) to account for confounders related to both parental depression and child mental health outcomes. Thus, it remains unclear to what extent parental depression in fact *causes* negative child outcomes, and how this effect varies across the outcome domain and across family characteristics (e.g., socioeconomic status). Obtaining accurate estimates of the magnitude and variability in this causal effect is crucial for the effective design, implementation, and allocation of finite intervention resources (Goodman & Garber, 2017; Nylen, Moran, Franklin, & O'Hara, 2006). I now consider two potential alternative, non-causal explanations for the association between maternal depression and child outcomes, before turning to methodological advances that will enable us to directly address them.

Opposite direction of causation. One alternative explanation is that child characteristics might cause an increase in maternal depression, leading to a correlation between maternal depression at time 1 and child characteristics at time 2 (e.g., Brooker et al., 2015; Nicholson, Deboeck, Farris, Boker, & Borkowski, 2011). For example, higher levels of child's defiant behavior at time 1 might both (a) lead to higher levels of maternal depression at time 1 and (b) be correlated with child's defiant behavior at time 2, thereby introducing a spurious relationship. This possibility is illustrated in panel B of Figure 2.

Indeed, experimental evidence supports the plausibility of this explanation. For example, Pelham and colleagues (1997) trained confederate children to mimic the behavior of either a "normal" child or a "deviant" child (i.e., one with Oppositional Defiant Disorder or Conduct Disorder), then had parents interact with the confederates in play and clean-up (e.g., collect all toys from floor) tasks typical of daily parent-child interactions. Parents who interacted with the confederate children who displayed deviant behavior subsequently reported significantly greater anxiety, depression, and hostility; more negative attitudes toward the child; and a reduced sense of efficacy in dealing with the child. These results illustrate a potential mechanism through which child characteristics (in this case, difficult behavior) could evoke parental depression: frequent unpleasant interactions and learned helplessness (i.e., the deviant confederates were unresponsive to parent commands throughout the interaction) could lead to parental depression.

Common cause. A second alternative explanation is that maternal depression and child outcomes arise from one or more common causes, or covariates that are themselves related to both constructs. In other words, the observed relationship could be explained by a set of confounding variables, *c*, that is related to both maternal depression and child characteristics. This possibility is illustrated in panel C of Figure 2. A large number of constructs might meet this criterion: marital quality, negative life events, presence of a second parent, socioeconomic status, comorbid psychiatric problems (e.g., substance use), and genes, for example. Any subset of these variables may be sufficient to produce a spurious relationship between maternal depression and child outcomes, but

the methodologies used in the literature reviewed above did not fully address this possibility.

This potential explanation can be concretely illustrated using data from the Early Steps sample. Consider the association between maternal depression and child externalizing behavior. In addition to these two constructs, 204 other "scores" (i.e., scale scores, composites) were collected at the baseline assessment (age 2). Of these 204, 67 different variables were correlated greater than 0.10 with both parental depression and child externalizing problems at this wave. Prominent examples include poverty status, financial stress, home "chaos", daily hassles, neighborhood danger, partner's emotional dysregulation, and parental marijuana and tobacco use. Any of these confounding variables might be partially responsible for the observed relationship between maternal depression and child externalizing behavior, providing a viable alternative explanation. In summary, despite many studies documenting association, *the degree to which maternal depression causes poor child outcomes remains unclear.*

CHAPTER 4

INTRODUCTION PART III: MODERN METHODOLOGY FOR CAUSAL INFERENCE

The present study will revisit this causal question using the potential outcomes (PO) model of causal inference (Holland, 1986; Imbens & Rubin, 2015; Rubin, 1978, 2005). This framework arose from a series of papers published by Rubin in the 1970s (Rubin, 1974, 1977, 1978), which built on the foundation earlier provided by Neyman (1923, 1990). Rubin's approach provides a mathematical framework that clarifies the conditions necessary for inferring causality and facilitates careful reasoning about challenging statistical issues such as missing data, non-random assignment, and time-varying treatments. Rubin begins by framing causal inference as the explicit comparison between two "potential outcomes" for a single unit—the outcome under treatment regime X_0 versus the outcome under treatment regime X_1. The "unit" (*i*) might be an individual child, and the "treatment" might be a counseling intervention: the two potential outcomes would then be (a) the child's outcome if he did *not* receive the counseling intervention ($Y_i|X_{0i}$) and (b) the *same* child's outcome if he did receive the counseling intervention ($Y_i|X_{1i}$). Rubin would then define the causal effect of the counseling intervention as the difference between $Y_i|X_{0i}$ and $Y_i|X_{1i}$.

It is impossible to observe both of these potential outcomes simultaneously—this is the so-called "fundamental problem of causal inference" (Holland, 1986). A single participant cannot simultaneously participate in both the control group *and* the intervention group. This problem can be addressed by analyzing the potential outcomes at the group rather than individual level. If it is random *which* potential outcome is

observed for each participant in a sample, then the group-level estimate of the causal effect (i.e., the difference in mean outcomes between the group that receives X_0 and the group that receives X_1) will correspond in expectation to the mean of the participant-level causal effects. This reasoning motivates the use of randomized, controlled trials (RCTs) to answer causal questions, such as the efficacy of an intervention or medication.

Outside of the RCT context, which potential outcome is observed for each participant is typically not determined by random assignment. For example, in the current context, mothers' levels of depression are not experimentally controlled, but rather mothers' level of depression is the result of a non-random assignment mechanism. Each mother's level of depression is not random, but instead dependent on their income, their romantic relationships, their family history, and many other factors (Kendler & Gardner, 2011). The potential outcomes perspective clarifies the steps needed to address this situation. Rubin considers the RCT as a special case of an observational study in which the researcher knows exactly how cases were assigned to levels of treatment, permitting clean causal inference. This framing suggests an approach to designing observational studies (like the one currently proposed) in which the analyst tries to statistically mimic an RCT of the treatment regime under study, modeling the mechanism by which units (i.e., mothers) were assigned to levels of treatment (i.e., depression) (Rubin, 2007, 2008).

Rosenbaum and Rubin (1983, 1984) proposed what has become an increasingly popular method of modeling the assignment mechanism: the propensity score (Thoemmes & Kim, 2011; West et al., 2014). In the context of a binary treatment, the propensity score is the "conditional probability of assignment to a particular treatment

given a vector of observed covariates" (Rosenbaum & Rubin, 1983). In other words, it is an estimate of the probability that a particular child will be exposed to a "treatment" (e.g., maternal depression) given the values of covariates for that particular child (e.g., impulsivity, neighborhood, parents' marital status). When two groups of units are balanced on the propensity score, then they are in expectation balanced on all the covariates included in the propensity score model (Rosenbaum & Rubin, 1983). Thus, propensity scores provide an approach to difficult casual inference problems by separating the problem into two modeling steps: first, creating groups that are equivalent on all measured characteristics *except* the treatment, and then second, comparing these equated groups to make inferences about the causal effect of the treatment. Other approaches that attempt to model both the assignment mechanism and the outcome distribution simultaneously (e.g., analysis of covariance [ANCOVA]) may be computationally infeasible (e.g., the model may be non-identified due to too many covariates) or provide nonsensical estimates (e.g., estimated causal effects for females, when no females were present in the treatment condition) (Imbens & Rubin, 2015).

The potential outcomes model (and the propensity score approach in particular) relies on several carefully stated assumptions, which are elaborated upon elsewhere (Holland, 1986; Imbens & Rubin, 2015; West et al., 2014) and will be discussed in the context of the design of the current study. Active research topics include how best to model the propensity function (Lockwood & McCaffrey, 2016; McCaffrey, Ridgeway, & Morral, 2004), to handle missing data on the covariates (Cham & West, 2016), to achieve balance on the propensity score (Austin, 2014), and to examine outcomes after completing matching (Kang & Schafer, 2007; Rubin & Thomas, 2000). The important

takeaway for present purposes is that the potential outcomes model has motivated methodological advances that will enable us to address many of the limitations of previous research on the relationship between parental depression and child outcomes.

CHAPTER 5

CURRENT STUDY

In many ways, we have made remarkable progress in the study of the effects of maternal depression on child outcomes. Since Rutter's (1966) initial report of this association, we have replicated this association in hundreds of studies, developed models of how this risk is transmitted, and begun delineating the conditions under which this association is magnified or attenuated. Yet in other ways, the field has made less progress. Compare the conclusions of the first review of the effects of parental depression on child outcomes, published in 1983, and the most recent review, published in 2011:

> "Assuming that there is a difference between children of affectively ill parents and controls, as the weight of the evidence indicates, there is the question of etiology." (Beardslee et al., 1983, p. 830, emphasis added).

> "Although the weight of evidence supports maternal depression being associations [sic] with children's emotional and behavioral problems, *causation, and the direction of association are not established.*" (Goodman et al., 2011, emphasis added).

Thus, although the weight of the evidence for association has increased dramatically in the intervening 30 years, basic questions of causation have still not been addressed. Each of the 200+ studies to date has suffered from the existence of many confounding variables, and none has attempted to estimate a causal effect.

The present study is an attempt to address this causal question. I apply a propensity score method to data from the Early Steps Multisite Trial, a high-risk sample of 731 low-income families tracked approximately annually from ages 2 through 16.

Families will be equated on a large set of baseline characteristics, producing two groups that are similar across all characteristics except the presence of clinically significant maternal depression during toddlerhood. Children's longitudinal psychopathology outcomes assessed by parent-, teacher-, and self-report will then be compared across the matched groups in order to estimate the causal effects of parental depression.

Hypotheses. I hypothesize that the magnitude of the relationship between maternal depression at age 2 and later child externalizing and internalizing behavior will be reduced by approximately half after matching or weighting with propensity scores. In other words, I hypothesize that half the *prima facie* effect will be eliminated by proper adjustment for confounding variables, and half the observed effect will remain. I do not expect results to differ by child outcome, as there is both theoretical and empirical support for a causal effect on both externalizing and internalizing behavior.

CHAPTER 6

GENERAL METHODS

Sample

The Early Steps trial has tracked 731 at-risk families recruited from the Women, Infants, and Children (WIC) Nutritional Supplement program beginning when children were 2 years old. Families were distributed across three sites—Eugene, OR; Charlottesville, VA; and Pittsburgh, PA—and were identified as being at-risk because they possessed multiple risk factors for the development of child conduct problems. One such risk factor was the presence of family problems like parental depression, contributing to the high rates of parental depression for the present study. 37% of primary caregivers were married and 60% reported having a live-in partner. 24% of primary caregivers did not have a high school degree; only 3% had a degree from a four-year college. Two-thirds of families reported income below $20,000 annually. 50% of children were male, 50% were European American, 28% were African American, and 13% were Hispanic American. Data were collected from primary caregivers, 97% of whom were mothers and 3% of whom were fathers. Data was also collected from secondary caregivers when available. The present analyses used data from only 707 of the 731 families. 21 cases were dropped because the primary caregiver was not the mother, 2 cases were dropped due to missing value for maternal depression at age 2, and 1 case was dropped due to missing value for whether mother had a live-in partner at age 2.

Design of Early Steps Trial

The primary purpose of the Early Steps trial was to investigate whether providing an annual family-based intervention (the Family Check-Up; Dishion & Kavanagh, 2003) during early childhood could prevent the development of child conduct problems and subsequent early-onset substance use. Families were randomly assigned to either intervention or control conditions (probability = 0.50) when children were 2 years old, then tracked prospectively. Families have been assessed at child age 3, 4, 5, 7, 8, 9, 10, 12, 14, and 16 years, with retention exceeding 75% at all waves. Intervention effects have been examined in other publications (Dishion et al., 2014, 2008; Shaw et al., 2009), but here I utilize the developmental, longitudinal aspects of the study. Intervention status is regarded as a covariate to be controlled for in estimating the causal effect of exposure to parental depression, since intervention group may be related to subsequent outcomes.

Defining the Treatment: Maternal Depression at Age 2

Maternal depression at baseline was assessed via the Center for Epidemiological Studies on Depression Scale (CES-D; Radloff, 1977), a reliable and valid measure of depressive symptoms. The "treatment" (i.e., having a depressed mother) was defined using the primary caregiver's self-report of symptoms. Twenty items measure the degree to which the respondent exhibited depressive mood (e.g., "felt lonely," "felt sad"), thoughts (e.g., "thought my life had been a failure"), and behaviors (e.g., "did not feel like eating", "felt that everything I did was an effort") during the past week (see Table 1 for list of items and their endorsement rates). The response scale for these items includes four values: *rarely or none* (0), *some or a little* (1), *occasionally* (2), or *most or all of the time* (3). A total score ranging from zero to 60 was calculated as the sum of the 20 items ($\alpha = 0.88$), with scores of 16 or greater taken to indicate significant risk of Major

Depressive Disorder (Lewinsohn, Seeley, Roberts, & Allen, 1997). According to this definition, 45% of primary caregivers met criteria for maternal depression at child age 2 (treatment = 1) and 55% of primary caregivers did not (treatment = 0).

The potential outcomes model emphasizes that the "treatment" must be explicitly defined (Holland, 1986; Rubin, 2010), typically by the comparison of two treatment groups. In this study, I define "treatment" as follows:

- X_0 – having a mother with CES-D total score < 16 at age 2 years old
- X_1 – having a mother with CES-D total score ≥ 16 at age 2 years old

The cutoff at a total CES-D score of 16 was chosen based on established norms (Lewinsohn et al., 1997). A histogram of the continuous scores is shown in Figure A1. Given that the nature of the CES-D items by necessity define our causal effect, they merit careful inspection and are reproduced in Table 1.

Computing

All data management and statistical analysis were conducted in the R statistical software environment (R Core Team, 2018) unless otherwise specified.

Organization of Remainder of Text

This design consists of a two-step procedure. In the first step, families with depressed mothers and families with non-depressed mothers were equated using propensity scores. In the second step, the longitudinal child outcomes of each group of families were compared. To improve readability, I present Methods and Results for the first step, followed by Methods and Results for the second step. Collapsing both steps would have produced a very long Methods section that would be difficult to map to the relevant parts of the Results.

CHAPTER 7

METHODS FOR STEP 1: EQUATING FAMILIES WITH DEPRESSED AND NON-DEPRESSED MOTHERS AT CHILD AGE 2

In the first step of this design I equated families with depressed and non-depressed mothers at age 2. Only data collected at age 2 were used, and child outcomes were never considered. My goal was to produce two groups of families that differed in maternal depression, but not on any other confounding variables (e.g., income, family structure, child aggression, mother substance use).

Measurement of Covariates at Age 2 Baseline

Nearly all covariates for balancing were drawn from the measures available at baseline, when the child was 2 years old. The mother completed 18 questionnaires, or self-report survey instruments. When available, the mother's live-in partner completed 8 questionnaires. Study staff completed home environment inventories. The only covariates that were not collected at child age 2 were scores indexing genetic risk for aggression based on the EAGLE Consortium's genome-wide association analyses (Pappa et al., 2016). Children's genes were sequenced at age 14. Although this was after potential exposure to maternal depression (i.e., age 2), these risk scores were deemed acceptable for inclusion in the balancing pool since they reflect inborn characteristics that could not be changed by maternal depression.

Table A2 presents a list of available measures and their corresponding citations. Drawing from these sources, I considered a total of approximately 350 variables for potential inclusion in the estimation of the propensity score (i.e., for equating families with depressed and non-depressed mothers). Covariates were eliminated due to

insufficient variance (e.g., only 1% of the sample endorsed drug use by the child's sibling), high rates of missing data, or lack of theoretical and empirical relation with the treatment variable. Variables were recoded as necessary to be suitable for inclusion in a regression model (e.g., categorical variables were converted into a set of dummy variables). Any variables that were reported by a live-in partner were recoded to zero (rather than missing) when a live-in partner did not exist, following the procedure for structurally-missing variables described by Dziak and Henry (2017). An indicator of whether the live-in partner existed was then included in the model.

Final covariates. This procedure resulted in the identification of 89 total covariates that were included in estimation of the propensity score. A table of these variables and their descriptive statistics is provided in the Appendix (Table A3). Covariates spanned the domains of demographics (e.g., sex, race, ethnicity, income, marital status), areas of family strength (e.g., support from extended family), negative impact factors (e.g., drug use by parent, death in the family), child behavior (e.g., aggression, noncompliance, anxiety, sleep), neighborhood factors (e.g., danger, cohesion), parent functioning (e.g., substance use, frequency of contact with friends), and factors related to live-in partners (e.g., relationship satisfaction of mother, live-in partner's substance use).

Handling Missing Data in Baseline Covariates

33 of the 89 covariates had no missing values, and 82 of the 89 covariates were missing fewer than 5% of values. Covariates with higher rates of missing data were measures of parent substance use (6 to 18% missing) and genetic risk scores (30% missing). There were 102 different patterns of missingness, and the majority of these

patterns applied to only one case in the dataset. Accordingly, I used the "imputation with constant plus missingness indicators" method to address missing data in the covariates (Cham & West, 2016). This method requires creating dummy variables indicating the presence or absence of a value for each covariate and including these indicators in the model for the propensity score.

Estimating the Propensity Score

Propensity scores were initially estimated using logistic regression, modeling the odds of the mother being depressed at age 2 as a function of the first order effects of 105 variables. 89 variables were measurements of a covariate (i.e., those listed in Table A3), and 16 variables were indicators of missingness. Each case's propensity score was then defined as the predicted probability of the mother being depressed, conditional on that case's values on all covariate terms.

After estimation of the propensity score, balance in both the matched and weighted datasets was checked. Terms that remained imbalanced (e.g., specific second-order term) were added to the model specification, and the propensity scores were re-estimated. This procedure was repeated until adequate balance was manifest in both the matched and weighted samples. Results presented below describe the balance achieved in this final model specification.

Equating Families Using the Propensity Score

Once a propensity score has been estimated for each family in the sample, the next step is to use this score to equate the families with depressed and non-depressed mothers. There are several approaches to doing so, including matching, subclassification, and weighting (Austin, 2014; Imbens & Rubin, 2015; Stuart, 2010). Here, I used both

matching and weighting procedures, which have different strengths and weaknesses (Austin, 2011; West et al., 2014).

Matching. In the matching procedure, I equated the two groups by forming a series of 1-to-1 matches of a depressed to a non-depressed mother using the *matchIt* package (Ho, Imai, King, & Stuart, 2011) in R. The distance metric (Stuart, 2010) was the logit of the propensity score (i.e., the predicted log-odds of exposure to parental depression), and a caliper of 0.10 SD units was enforced (i.e., matches must be within 0.10 SD of each other in the logit of the propensity score). Matching has the benefit of ensuring common overlap in the propensity score distribution (i.e., region of support) because untreated cases with extreme propensity scores will only be included if a treated case with similarly extreme propensity score can be found. Matching has the downside of reducing sample size, because some cases will go unmatched and so be excluded from further analysis. Moreover, because matching excludes untreated cases that cannot be matched, the resulting estimates of treatment effects are only applicable to the treated portion of the sample (Average Treatment Effect for the Treated; ATT).

Weighting. In the weighting procedure, I equated the two groups by weighting the data to be representative of a sample in which maternal depression was randomly assigned. Every case received a survey weight in subsequent analyses equal to the inverse of the estimated probability of having a depressed mother at age 2. Weighting has the benefit of retaining the full sample size, since matching is unnecessary. As a result, the treatment effects estimated using weighting are applicable to the entire sample (Average Treatment Effect; ATT). Weighting has the downside of potentially including cases from regions of the propensity score in which there is not in fact common support

in the treated or untreated cases. For example, there may be a group of depressed mothers with propensity scores of 0.90+, but no corresponding group of non-depressed mothers with propensity scores of 0.90+ with which to compare them. In this case, we cannot credibly estimate the effect of maternal depression for mothers in this region of propensity score, and results should not be trusted.

Verifying that Families Have Been Successfully Equated

We can infer the assignment mechanism has been modeled correctly whenever the resulting matched groups are well-balanced (the so-called "propensity score tautology"; Ho, Imai, King, & Stuart, 2007). Following standard practice, I focused on the balance of treated versus untreated participants across the full set of covariates (Rubin, 2001). Following guidelines suggested by Rubin (2001), I verified there were no standardized mean differences (SMDs) greater than 0.20 standard deviations and there are no variance ratios outside of 0.5 to 2. For categorical variables, I verified that the proportion of cases in each category differed by no more than 5%. Variables with remaining non-negligible differences (e.g., SMD > 0.10) were included as covariates in all outcome models (Rubin & Thomas, 2000; Schafer & Kang, 2008), since they would otherwise remain as alternative explanations for differences in child outcome.

Before checking balance, both the truly missing and structurally missing values that were modified for entry into the propensity score model were reset to missing. Squared terms for all covariates were added to the dataset. After these modifications, balance statistics for matching were calculated using only those families that were successfully matched. Balance statistics for the weighted sample were calculated using all 707 families and using weighted means and weighted variances.

CHAPTER 8

RESULTS FOR STEP 1: EQUATING FAMILIES WITH DEPRESSED AND NON-DEPRESSED MOTHERS AT CHILD AGE 2

Matching

179 depressed mothers were matched with 179 non-depressed mothers (from a pool of 323 depressed and 385 non-depressed mothers). Figure 3 shows the distribution of estimated propensity scores in the depressed and non-depressed groups before and after matching. Although the propensity scores of depressed and non-depressed mothers differed significantly prior to matching (left panel), they exhibited a similar distribution after matching (right panel). The matched groups covered the same span on the probability metric, indicating a common region of support for estimates based on the matched cases.

Metric variables. Table 1 compares the depressed and non-depressed groups' means and variances on each covariate after matching. After matching, the main effects of all 35 non-binary covariates exhibited SMDs of less than 0.20 SD (Figure 4). Only two variables exhibited SMDs exceeding 0.10 SD. Depressed mothers reported a greater frequency of daily hassles (SMD = 0.12) and children of depressed mothers had higher GWAS scores for middle childhood aggression (SMD = 0.11). These two variables were included as covariates in subsequent analyses using the matched dataset. Finally, all covariates exhibited a variance ratio of between 0.5 and 2.0, with nearly all falling between 0.8 and 1.4 (Figure 5). Balance on squared terms was very similar to that on main effects.

Binary variables. Table 1 also compares the proportions of respondents endorsing each binary covariate in the depressed and non-depressed groups after matching. After matching, the main effects of 67 of 68 binary covariates exhibited differences in prevalence of less than 5% (Figure 6). The remaining covariate was whether live-in partner currently smokes cigarettes, with a difference in prevalence of 6%. Because the difference in prevalence was small, based on a small number of cases (n = 54 per group), and presents difficulty for the imputation model due to structural missingness, this variable was *not* included in subsequent analyses using the matched dataset.

Weighting

Inverse probability of treatment weights were created using the estimated propensity scores. Weights ranged from 1.00 to 24.73, with a median of 1.39 and an interquartile range of 1.15 to 1.95. The top 10 weights were 24.73, 21.93, 18.88, 18.90, 14.93, 14.20, 13.17, 12.64, 12.10, and 11.81.

Metric variables. Table 2 compares the depressed and non-depressed groups' means and variances on each covariate after weighting. After matching, the main effects of all 35 non-binary covariates exhibited SMDs of less than 0.20 SD (Figure 7). Only four variables exhibited SMDs exceeding 0.10 SD. Families with a depressed mother had older children at study entry (SMD = 0.15), had more people living in the home (SMD = 0.10), had more adults living in the home (SMD = 0.12), and received higher parent involvement ratings from home visitors (SMD = 0.14). These four variables were included as covariates in subsequent analyses using the weighted dataset. Finally, all covariates exhibited a variance ratio of between 0.5 and 2.0, with most falling between

0.9 and 1.3 (Figure 8). Balance on squared terms was very similar to that on main effects.

Binary variables. Table 2 compares the prevalence of each binary covariate in the depressed and non-depressed groups after weighting. After weighting, the main effects of 68 of 70 binary covariates exhibited differences in prevalence of less than 5% (Figure 9). Depressed mothers were less likely to have been a teen parent (-5%), and children of depressed mothers were more likely to be male (+5%). Both variables were complete and could be included as covariates in subsequent analyses using the weighted dataset.

CHAPTER 9

METHODS FOR STEP 2: CHILD OUTCOME ANALYSES

In step 1, families with depressed and non-depressed mothers were successfully equated at age 2 on a broad set of confounding variables, using both matching and weighting methods. At no point in this process were subsequent child outcomes considered, preserving the integrity of the hypothetical randomized, controlled trial I am seeking to emulate (Rubin, 2008). I now proceed to compare the longitudinal child outcomes of the equated groups.

Measurement of Child Outcomes at Ages 3 to 14

Child outcomes across ages 3 to 14 were assessed using multiple measures via mother, secondary caregiver, teacher, and child report. I focus first on the immediate outcome at age 3, before proceeding to examine outcomes across the full age range that was assessed. A table of descriptive statistics for outcome variables is provided in the Appendix (Table A4). Outcomes can be grouped into the following three categories.

Broadband behavioral ratings at ages 3 to 14. The Child Behavior Checklist (CBC; Achenbach & Rescorla, 2001) was completed by both the mother and secondary caregiver (when available) at ages 3, 4, 5, 7, 8, 9, 10, and 14. Secondary caregivers were primarily biological father (43%), grandmother (15%), mother's male boyfriend (13%), stepfather (9%), or aunt (5%), with all remaining categories each comprising less than 5% of reports (Figure A2). At ages 7, 8, 9, 10, and 14, between 80 and 84% of secondary caregivers lived with the child (this information was not collected at earlier waves). The Teacher Report Form (TRF; Achenbach & Rescorla, 2001) was completed by the teacher at ages 7, 8, 9, 10. The externalizing and internalizing subscales of the CBC and TRF

were analyzed as outcomes. Total scores on each subscale (e.g., the sum of all responses to items measuring externalizing problems) were divided by the number of constituent items to reflect mean response per item (range = 0 – 2), since the number of items on each scale differed across forms of the CBCL. The preschool form of the CBCL (designed for children ages 1.5 to 5) was administered in this study at child ages 2, 3, and 4. The school-age form of the CBCL (designed for children ages 6 to 18) was administered in this study at child ages 5, 7, 8, 9, 10, and 14. Different items contribute to the externalizing and internalizing subscales on these two forms, and this must be kept in mind when comparing effects across ages.

Child-report measures at age 14. Three child-report measures were administered at age 14: a short form of the Child Depression Inventory (CDI) (Saylor, Finch, Spirito, & Bennett, 1984), a short form of the Multidimensional Anxiety Scale for Children (MASC) (March & Sullivan, 1999), and a study-specific scale assessing adolescents' deviant behavior (ASRD) (see Appendix for details). Total scores from each of these three measures at all available ages were analyzed as outcomes.

DSM symptoms at age 10. DSM symptoms were assessed at age 10. Both mother and child were administered a structured diagnostic interview with the Diagnostic Interview Schedule for Children (DISC-IV) (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000). Mother and child report were analyzed separately. The total number of symptoms endorsed for the following disorders were analyzed as outcomes: Social Phobia Disorder, Separation Anxiety Disorder, Generalized Anxiety Disorder, Obsessive-Compulsive Disorder, and Major Depressive Disorder. Note that DISC symptom counts

can be larger than the number of DSM symptoms, because a single DSM symptom might be queried with multiple DISC questions.

In addition, mothers completed the Disruptive Behavior Disorders Rating Scale (DBD-RS) (Pelham, Gnagy, & Milich, 1992). The DBD-RS asks the informant to rate each DSM symptom of a disruptive behavior disorder on a scale from 0 (*not at all*) to 3 (*very much*). The mean symptom rating for the following three disorders were analyzed as outcomes: Attention Deficit Hyperactivity Disorder, Oppositional Defiant Disorder, and Conduct Disorder.

Handling Missing Data in Child Outcomes

The rate of missing data on the outcome variables varied by informant (see Table A4). Parent-reported outcomes range from 10 to 27% missing, teacher-reported outcomes range from 48 to 61% missing, and child-reported outcomes range from 24 to 35% missing. Missing data were addressed using multiple imputation by chained equations (MICE) (Raghunathan, Lepowski, van Hoewyk, & Solenberger, 2001; van Buuren, 2012; White, Royston, & Wood, 2011) with the *mice* package (van Buuren & Groothuis-Oudshoorn, 2011) in R. All 707 families were included in the same imputation model.

MICE assumes that data are missing at random conditional on all variables included in the imputation model (MAR; Rubin, 1976). This assumption is more plausible when the imputation model is comprehensive, including not only all variables used in subsequent analyses but also auxiliary variables that may predict both missingness in and values of the outcome variable (Graham, 2009). Screening of the age 2 variables identified several covariates that were significantly correlated with both

outcomes and outcome missingness. The imputation model included the following variables:

(a) all outcome variables

(b) binary indicator of maternal depression at age 2

(c) logit of the propensity score

(d) dummy variables indicating site

(e) dummy variable indicating randomized to intervention group (in RCT)

(f) dummy variable indicating child being male

(g) dummy variable indicating child being African American

(h) dummy variable indicating mother married or living with significant other at age 2

(i) dummy variable indicating family being below poverty line at age 2

(j) mother report indicating externalizing behavior at age 2

(k) mother report of internalizing behavior at age 2

(l) the two variables identified as having residual imbalance in matching (mother report of frequency of total daily hassles, GWAS score for middle childhood aggression)

(m) the six variables identified as having residual imbalance in weighting (child age, child is male, mother was a teen parent, number of people living in home, number of adults living in home, home visitor rating of parent involvement)

Logistic regression was used to impute missing values for all binary variables; predictive mean matching was used to impute missing values for all other variables. Variables in

(b) through (m) were specified to predict all variables in (a). Variables in (a) were specified to predict all variables in (b) through (m).

Five imputed datasets were created and convergence of the MICE algorithm was verified by visual inspection of trace plots and a maximum Gelman-Rubin proportional scale reduction statistic of 1.1 for all means and variances (Gelman & Rubin, 1992). Plots of the observed versus imputed data for each variable did not reveal any major discrepancies or indicate potential computational issues (Abayomi, Gelman, & Levy, 2008). Given that the imputation model passed these diagnostics, 500 imputed datasets were created to minimize simulation error and maximize statistical power (Graham, 2009). All outcome analyses were repeated in each of the 500 imputed datasets, and the resulting estimates were combined using Rubin's rules (Rubin, 1987).

Analytic Plan

As described above, outcome measures were drawn from eight different waves spanning ages 3 to 14, from four different informants, and from multiple domains of child behavior. Thus, I faced a tradeoff between dramatically restricting the scope of outcome analyses or having a multiplicity of statistical comparisons. I chose to err on the side of a comprehensive analysis, wherein the consistency of results across ages, across informants, and across methods (e.g., matching vs. weighting) can be evaluated. However, the multiplicity of comparisons implies that the reported *p*-values and confidence intervals should be treated with caution.

Calculation of effect on each outcome at each age. To estimate unadjusted *prima facie* effects, I regressed the outcome variable on a dummy variable indicating that mother was non-depressed (coded as 0) versus depressed (coded as 1) at age 2. All 707

families were included. To estimate effects based on propensity score matching, I regressed the outcome variable on the maternal depression dummy variable as well as the two covariates identified as having remaining imbalance (both of which were centered). Data from only the 358 families that were successfully matched were included. To estimate effects based on propensity score weighting, I regressed the outcome variable on the maternal depression variable as well as the six covariates identified as having remaining imbalance (all of which were centered). Data from all 707 families were included, and each case was weighted by the inverse probability of treatment weight using the *survey* package in R (Lumley, 2003). Families in the depressed group were weighted by [1 / *pscore*]; families in the non-depressed group were weighted by [1 / (1 - *pscore*)], where *pscore* is the propensity score.

Conversion to standardized mean difference. For ease of interpretation, all effects were converted to the Cohen's *d* metric. For each measure (e.g., CBCL), the standard deviation was calculated using data from all 707 families at the earliest administration of that measure (e.g., for the CBCL internalizing subscale, age 3). This same standard deviation was used when calculating effects on the outcomes at each age, so that effect sizes across ages would be directly comparable (Greenland, 1987). For each outcome, the unadjusted, matched, and weighted effect coefficients were divided by the sample-wide standard deviation. Confidence intervals were obtained in the original metric of the unstandardized coefficient before conversion to the metric of Cohen's *d*.

Mixed models for outcomes measured repeatedly. Several outcomes were assessed at multiple ages (e.g., teacher rated externalizing behavior when child was 7, 8, 9, and 10). Linear mixed models were fit to investigate whether the effect of maternal

depression at age 2 varied as a function of the age at which the child outcome was assessed. For each outcome with repeated measurements, a series of four models were fit:

M1. A random intercept model (i.e., random family effect).

M2. A random intercept model with fixed effects for age of assessment.

M3. A random intercept model with fixed effects for age of assessment, plus fixed effect for maternal depression at child age 2 (i.e., depressed vs. non-depressed).

M4. A random intercept model with fixed effects for age of assessment, plus fixed effect for maternal depression at child age 2 (i.e., depressed vs. non-depressed), plus fixed effects for the interaction of age of assessment and maternal depression at child age 2.

Models fit to the propensity-score-matched dataset also included frequency of daily hassles and GWAS scores for middle childhood aggression as covariates. Each of models M1 through M4 is nested within the subsequent one, so their fit can be compared using the likelihood ratio test. Test statistics were pooled across imputations using the method of Meng and Rubin (1992), as implemented in the *mitml* package (Grund, Robitzsch, & Luedtke, 2018). The test comparing Model 2 to Model 1 tested the null hypothesis that accounting for age of assessment does not improve fit beyond accounting for random family effects. The test comparing Model 3 to Model 2 tested the null hypothesis that accounting for maternal depression group at child age 2 does not improve model fit after accounting for random family effects and fixed age of assessment effects. Finally, the test comparing Model 4 to Model 3 tested the null hypothesis that permitting

the effect of maternal depression at child age 2 to vary by age of assessment does not improve model fit after accounting for random family effects, fixed age of assessment effects, and fixed effect of maternal depression at child age 2. In the event of a statistically significant test statistic for comparing Model 4 to Model 3, plots of effects over time were visually inspected to identify trends.

Sensitivity analysis. Finally, sensitivity analyses were conducted to probe the extent to which the results are robust to the potential existence of unmeasured confounding variables that were not included when equating the treated and nontreated groups (Rosenbaum, 1986; West et al., 2014). Let U be a hypothetical, unmeasured confounding variable, r_{uy} be the correlation of this variable with an outcome of interest, and smd_u be the difference between the non-depressed and depressed groups on this variable at baseline. Following Hong (2004), I used the available data to estimate plausible maximum values of r_{uy} and smd_u (higher values indicate greater degree of confounding). I then evaluated how the observed effects of maternal depression at age 2 would change if an unmeasured confounder with these maximum values were properly accounted for. Otherwise stated, if a confounder existed with the maximum r_{uy} and smd_u observed in the data set (a kind of worst case scenario), how much would the effect size of mother's depression be attenuated and would the effect still be statistically significant?

CHAPTER 10

RESULTS FOR STEP 2: CHILD OUTCOME ANALYSES

All effects are reported as standardized mean differences in the Cohen's *d* metric. "Unadjusted" *prima facie* effects refer to simple comparisons of families with depressed vs. non-depressed mothers, without consideration of potential confounding variables. "Adjusted" effects refer to comparisons that adjust for confounding variables using either (a) *matching* on the propensity score or (b) *weighting* by the inverse probability of treatment. Estimated effect sizes and confidence intervals for all outcomes are reported in Table A5. For outcomes measured at multiple ages (i.e., repeated measures), Table 3 reports tests of (a) the effect of age of assessment, (b) the effect of maternal depression at child age 2, and (c) their interaction.

The results are also depicted graphically. Figure 10 shows mean effects on externalizing and internalizing behavior collapsed across all available ages. This figure provides an overall picture of the results. Figure 11 shows the proportion of the *prima facie* effect size at each age was eliminated by adjustment for confounding. This figure shows the impact of adjusting for the effects of potential confounding variables. Figures 12, 13, and 14 show effects of maternal depression at age 2 on all outcomes reported by mother, secondary caregiver, and teacher, respectively. Figure 15 shows effects on DSM symptoms at age 10, per both mother and child report. Finally, Figure 16 shows the effects over time for each informant/measure combination.

Age 3 Outcomes (CBCL)

Following the suggestion of my committee, I first focus on outcomes at age 3, before considering outcomes spanning the full range of ages 3 to 14. Age 3 is the first

available follow-up, and these results reflect children's outcomes 1 year after maternal depression was measured. Based on mother's report, the unadjusted *prima facie* effect of maternal depression at age 2 on externalizing behavior at age 3 was $d = 0.51$ ($p < .05$, 95% CI = [0.36, 0.66]). After adjustment, the effect was reduced to $d = 0.16$ (*ns*) when using propensity score matching and $d = 0.21$ ($p < .05$) when using propensity score weighting. These results are depicted in Figure 12, Panel A. The unadjusted *prima facie* effect of maternal depression at age 2 on internalizing behavior at age 3 was $d = 0.51$ ($p < .05$, 95% CI = [0.36, 0.66]). After adjustment, the effect was reduced to $d = 0.13$ (*ns*) when using matching and $d = 0.21$ ($p < .05$) when using weighting. These results are depicted in Figure 12, Panel I.

Based on secondary caregiver's report, the unadjusted *prima facie* effect of maternal depression at age 2 on externalizing behavior at age 3 was $d = 0.32$ ($p < .05$, 95% CI = [0.13, 0.51]). After adjustment, the effect was reduced to $d = 0.07$ (*ns*) when using matching and $d = 0.05$ (*ns*) when using weighting. These results are depicted in Figure 13, Panel A. The unadjusted *prima facie* effect of maternal depression at age 2 on internalizing behavior at age 3 was $d = 0.19$ ($p < .05$, 95% CI = [0.00, 0.38]). After adjustment, the effect was reduced to $d = -0.03$ (*ns*) when using matching and $d = -0.01$ (*ns*) when using weighting. These results are depicted in Figure 13, Panel I. Having examined the immediate outcomes at age 3, I now consider the outcomes at all available ages.

Mother-Reported Outcomes (CBCL)

Figure 12 shows the effects of maternal depression at age 2 on mother-reported outcomes between ages 3 and 14. Table 3 reports the results of mixed models.

Externalizing behavior. In the unadjusted mixed model analysis, there was a significant main effect of maternal depression at child age 2 ($F(1, 175826.5) = 50.25$, $p < .05$) and a significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 50262.4) = 3.41$, $p < .05$). The unadjusted *prima facie* effect of maternal depression was smaller at child ages 5 and 14 than at the other ages (Figure 16, Panel A). In the mixed model analysis after matching, there was no significant main effect of maternal depression at child age 2 ($F(1, 223247.0) = 1.79$, *ns*) and no significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 50703.5) = 1.24$, *ns*).

Considering effect sizes at single timepoints, the unadjusted *prima facie* effects of maternal depression at child age 2 on mother report of child externalizing behavior ranged from $d = 0.34$ to 0.53 (all ps $< .05$), with a mean value across ages of 0.45. After matching, effects ranged from $d = -0.06$ to 0.30, with a mean value of $d = 0.11$. Using weighting, effects ranged from $d = -0.04$ to 0.28, with a mean value of $d = 0.14$. Only 3 of the 16 matched or weighted effects were statistically significant (Figure 12).

Internalizing behavior. In the unadjusted mixed model analysis, there was a significant main effect of maternal depression at child age 2 ($F(1, 115296.2) = 44.87$, $p < .05$) but no significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 45.394.7) = 0.97$, *ns*). In the mixed model analysis after matching, there was no significant main effect of maternal depression at child age 2 ($F(1, 127810.6) = 0.63$, *ns*) and no significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 48465.4) = 0.69$, *ns*).

Considering effect sizes at single timepoints, the unadjusted *prima facie* effects of maternal depression at child age 2 on mother report of child internalizing behavior ranged from $d = 0.25$ to 0.51 (all ps < .05), with a mean value across ages of $d = 0.35$. After matching, effects ranged from $d = -0.03$ to 0.18, with a mean value of $d = 0.05$. Using weighting, effects ranged from $d = 0.04$ to 0.24, with a mean value of $d = 0.14$. Only 2 of 16 matched or weighted effects were statistically significant (Figure 12).

Secondary-Caregiver-Reported Outcomes (CBCL)

Figure 13 shows the effects of maternal depression at age 2 on secondary-caregiver-reported outcomes between ages 3 and 14. Table 3 reports the results of mixed models.

Externalizing behavior. In the unadjusted mixed model analysis, there was a significant main effect of maternal depression at child age 2 ($F(1, 7987.2) = 34.69$, $p < .05$) but no significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 13265.2) = 0.75$, *ns*). In the mixed model analysis after matching, there was no significant main effect of maternal depression at child age 2 ($F(1, 11514.3) = 0.73$, *ns*) and no significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 14522.0) = 0.23$, *ns*).

Considering effect sizes at single timepoints, the unadjusted *prima facie* effects of maternal depression at child age 2 on secondary caregiver report of child externalizing behavior ranged from $d = 0.12$ to 0.38 ($p < .05$ at seven of eight ages), with a mean value across ages of 0.30. After matching, effects ranged from $d = -0.08$ to 0.16, with a mean value of $d = 0.06$. Using weighting, effects ranged from $d = -0.05$ to 0.15, with a mean value of $d = 0.06$. None of the matched or weighted effects was statistically significant.

Internalizing behavior. In the unadjusted mixed model analysis, there was a significant main effect of maternal depression at child age 2 ($F(1, 5068.9) = 25.63$, $p < .05$) but no significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 12501.4) = 0.58$, *ns*). In the mixed model analysis after matching, there was no significant main effect of maternal depression at child age 2 ($F(1, 6121.6) = 0.33$, *ns*) and no significant interaction of maternal depression at child age 2 and age of assessment ($F(7, 15810.4) = 0.74$, *ns*).

Considering effect sizes at single timepoints, the unadjusted *prima facie* effects of maternal depression at child age 2 on secondary caregiver report of child internalizing behavior ranged from $d = 0.11$ to 0.30 ($p < .05$ at six of eight ages), with a mean value across ages of $d = 0.22$. After matching, effects ranged from $d = -0.18$ to 0.14, with a mean value of $d = 0.03$. Using weighting, effects ranged from $d = -0.01$ to 0.15, with a mean value of $d = 0.07$. None of the matched or weighted effects was statistically significant.

Teacher-Reported Outcomes (TRF)

Figure 14 shows the effects of maternal depression at age 2 on teacher-reported outcomes between ages 7 and 10. Table 3 reports the results of mixed models.

Externalizing behavior. In the unadjusted mixed model analysis, there was a significant main effect of maternal depression at child age 2 ($F(1, 4594.4) = 17.14$, $p < .05$) but no significant interaction of maternal depression at child age 2 and age of assessment ($F(3, 4857.6) = 0.13$, *ns*). In the mixed model analysis after matching, there was a significant main effect of maternal depression at child age 2 ($F(1, 4939.4) = 5.61$, *p*

< .05) but no significant interaction of maternal depression at child age 2 and age of assessment ($F(3, 5548.3) = 0.15$, *ns*).

Considering effect sizes at single timepoints, the unadjusted *prima facie* effects of maternal depression at child age 2 on teacher report of child externalizing behavior ranged from $d = 0.21$ to 0.29 (all *p*s < .05), with a mean value across ages of $d = 0.25$. After matching, effects ranged from $d = 0.14$ to 0.25, with a mean value of $d = 0.20$. Using weighting, effects ranged from $d = 0.08$ to 0.24, with a mean value of $d = 0.18$. None of the matched or weighted effects was statistically significant.

Internalizing behavior. In the unadjusted mixed model analysis, there was a significant main effect of maternal depression at child age 2 ($F(1, 2069.4) = 7.29$, $p <$ .05) but no significant interaction of maternal depression at child age 2 and age of assessment ($F(3, 4469.9) = 0.86$, *ns*). In the mixed model analysis after matching, there was no significant main effect of maternal depression at child age 2 ($F(1, 2511.6) = 2.25$, *ns*) and no significant interaction of maternal depression at child age 2 and age of assessment ($F(3, 5376.1) = 0.69$, *ns*).

Considering effect sizes at single timepoints, the unadjusted *prima facie* effects of maternal depression at child age 2 on teacher report of child internalizing behavior ranged from $d = 0.07$ to 0.30 ($p < .05$ only at age 10), with a mean value across ages of $d = 0.17$. After matching effects ranged from $d = 0.03$ to 0.29, with a mean value of $d = 0.12$. Using weighting, effects ranged from $d = -0.03$ to 0.23, with a mean value of $d = 0.08$. None of the matched or weighted effects was statistically significant.

Child-Reported Outcomes (MASC, CDI, and ASRD)

Unadjusted *prima facie* effects of maternal depression at age 2 on child-reported depression, anxiety, and deviancy at age 14 were all negligible ($d < 0.10$, *ns*; see Table A5 for estimates and confidence intervals). Effects after matching and weighting were also negligible.

DSM Symptoms at Age 10 (DBD-RS and DISC)

Figure 15 shows the effects of maternal depression at age 2 on DSM symptoms at age 10, per both parent and child report. Exact values are reported in Table A5. For DSM symptom ratings of externalizing disorders, unadjusted *prima facie* effects on mother-reported symptoms ranged from $d = 0.23$ to 0.35 (all *p*s < .05), with a mean value across symptom clusters of $d = 0.29$. After matching, effects ranged from $d = 0.07$ to 0.16, with a mean value of $d = 0.11$. Using weighting, effects ranged from $d = 0.10$ to $d = 0.13$, with a mean value of $d = 0.12$. None of the matched or weighted effects was statistically significant.

For DSM symptom counts of internalizing disorders, unadjusted effects on mother-reported symptoms ranged from $d = 0.15$ to 0.38 ($p < .05$ except for Social Phobia Disorder), with a mean value across symptom clusters of $d = 0.27$. The largest effect size ($d = 0.38$) was observed for symptoms of Major Depressive Disorder. After matching, effects ranged from $d = 0.01$ to 0.22, with a mean value of $d = 0.11$. Using weighting, effects ranged from $d = 0.05$ to 0.28, with a mean value of $d = 0.16$. Only 2 of the 10 matched or weighted effects were statistically significant. Per child report, the unadjusted effects ranged from $d = -0.07$ and 0.05 (all *ns*), with a mean value of $d = -0.01$. Matched and weighted effects were also negligible.

Sensitivity Analysis

The largest standardized difference at baseline between families with depressed versus non-depressed mothers at age 2 (i.e., smd_U) was on mother perception of total daily hassles ($d = 0.69$). The largest correlation observed between any baseline covariate and any of the outcomes (i.e., r_{UY}) was between mother ratings of adult-child relationship conflict at age 2 (Pianta, 1995) and mother report of child externalizing behavior at age 3 ($r = 0.41$). These values were plugged in to the formulas in West et al. (2014). An unmeasured confounder with a standardized difference and correlation at these values would eliminate an observed effect of 0.20 (*d*) or lower, and attenuate an observed effect of 0.30 down to 0.10 (*d*). Thus, an unmeasured confounder corresponding to the largest standardized mean difference and largest baseline covariate-outcome correlation would theoretically be expected to explain nearly all the remaining effect of maternal depression at age 2 on later child outcomes (see magnitude of adjusted *d*s in Figure 10, which are almost entirely below 0.20). This sensitivity analysis reflects a kind of worst case scenario: It is quite difficult to imagine an unmeasured variable that is would be this strongly related to both maternal depression and to child outcomes following adjustment through matching or weighting for the 89 confounding variables listed in Table A3.

CHAPTER 11

DISCUSSION

I used a propensity-score-based design to estimate the causal effect of having a mother with CES-D total score above 15 at child age 2 years on child externalizing and internalizing behavior spanning ages 3 to 14, per multiple informants. Both matching and weighting approaches were successful in equating families with depressed and non-depressed mothers on a set of 89 potentially confounding variables measured at child age 2, including demographics, mother characteristics, and child characteristics. Prior to adjustment for confounding, small- to medium-sized negative effects of maternal depression at age 2 were consistently observed on children's externalizing and internalizing behavior between ages 3 and 14, per mother, secondary caregiver, and teacher reports. After adjustment for confounding variables, these effects were in the very small to small range and few were statistically significant. The children of depressed vs. non-depressed mothers at age 2 reported no differences in externalizing or internalizing outcomes at age 10 or 14, with or without adjustment for confounding variables. Taken together, results suggest that the *prima facie* relation of maternal depression at child age 2 to subsequent behavioral outcomes is in large part non-causal.

Effects of Maternal Depression *Before* Adjusting for Confounding Variables

This study replicated past literature (Goodman et al., 2011) in finding a robust positive association of maternal depression with subsequent child behavior problems. Before adjusting for confounding variables, maternal depression at child age 2 had a small to medium effect on child behavior across ages 3 to 14. Mean effect sizes were as follows: mother-reported externalizing behavior ($d = 0.45$), mother-reported internalizing

behavior (d = 0.35), secondary-caregiver-reported externalizing behavior (d = 0.30) and internalizing behavior (d = 0.22), and teacher-reported externalizing behavior (d = 0.25) and internalizing behavior (d = 0.17). 34 of 40 effects were statistically significant. Before proceeding to consider adjusted effects, I compare these unadjusted effect sizes to the existing literature.

Goodman et al.'s (2011) meta-analysis includes 193 studies of the effects of maternal depression on child externalizing and internalizing behavior.[1] Design issues complicate the comparison of my overall results (e.g., Figure 10) with those of Goodman et al. (2011). On the one hand, this study's length of follow-up is probably longer than most of the studies included in the meta-analysis, which would presumably attenuate the observed effects of maternal depression because more time elapsed between this exposure and the measurement of child outcomes. On the other hand, the Early Steps sample had several attributes that were associated with stronger effects of maternal depression in the meta-analysis, including lower family income, younger child age, higher proportion of unmarried mothers, and higher proportion of ethnic minority mothers.

To facilitate the clearest comparison to Goodman et al. (2011), I used the effect of maternal depression observed at the first available wave of follow-up for each outcome (i.e., child age 3 for mother report, child age 7 for teacher report, and child age 14 for child report). For child externalizing problems, the effect per mother report was d = 0.51

[1] Goodman et al. (2011) used the correlation metric, r, to report all results. To facilitate comparison to our results, their reported correlations were converted to the standardized difference metric, d, using equation 7.4 of Borenstein et al. (2009).

(cf. $d = 0.52$ in Goodman et al., 2011), the effect per teacher report was $d = 0.26$ (cf. $d = 0.28$), and the effect per child report was $d = 0.06$ (cf. $d = 0.22$). For child internalizing problems, the effect per mother report was $d = 0.51$ (cf. $d = 0.52$), the effect per teacher report was $d = 0.10$ (cf. $d = 0.30$), and the effect per child report was $d = 0.06$ (cf. $d = 0.35$). Thus, the unadjusted effects of maternal depression on mother- and teacher-reported child externalizing and internalizing behavior in this study were quite similar to those found in the Goodman et al.'s (2011) meta-analysis of past literature. However, the unadjusted effects on *child*-reported externalizing and internalizing symptoms were considerably smaller than those reported in Goodman et al. (2011). This may be explained by this study having a longer interval (i.e., 8 to 12 years) between measurement of maternal depression and child report of own behavior than did previous studies.

Effects of Maternal Depression *After* Adjusting for Confounding Variables

I adjusted for the effects of confounding variables with two methods: *matching* on the propensity score and *weighting* by the inverse probability of treatment. Results were similar with both methods, suggesting that our conclusions are robust to the mechanism by which I made the adjustment. Both methods suggested that a substantial portion of the association of maternal depression and subsequent child behavior is explained by confounding variables (e.g., socioeconomic status), rather than a causal process.

Proportion of effect eliminated. One way to formulate the effect of adjustment for confounding variables is the proportion of the original effect that is eliminated after the adjustment (Figure 11). Since the results after matching and weighting were similar, I defined the "adjusted" effect on each outcome as the average of the effect per matching and the effect per weighting. Across timepoints, between 45% and 100% (mean = 73%)

of the effect of maternal depression on mother-reported child behavior was eliminated by adjustment for confounding variables; between 46 and 100% (mean = 77%) of the effect on secondary-caregiver-reported child behavior was eliminated; and between 2 and 100% (mean = 37%) of the effect on teacher-reported child behavior was eliminated. As indicated by those ranges, there was considerable variability in proportion eliminated across timepoints. The proportion eliminated was generally similar for child externalizing and internalizing behavior.

As described above, effects on measures at child age 3 may be most directly comparable to the existing literature. At child age 3, adjustment for confounding variables eliminated 64% of the effect of maternal depression on mother-reported externalizing problems, 67% of the effect on mother-reported internalizing problems, 81% of the effect on secondary-caregiver-reported externalizing problems, and 100% of the effect on secondary-caregiver-reported internalizing problems.

These proportions must be interpreted cautiously. For example, in Panel A of Figure 16, the vertical distance (i.e., difference in *d*) between the unadjusted and adjusted lines is stable over time, although the proportion of effect eliminated varies as both lines shift upward or downward over time. Comparison of two specific instances illustrates this point. 100% of the effect of maternal depression on teacher-reported internalizing behavior at child age 8 was eliminated, as the effect dropped from $d = 0.07$ to $d = 0$. Only 64% of the effect of maternal depression on mother-reported internalizing behavior at child age 3 was eliminated, but this reflected a substantially larger drop in effect size from $d = 0.51$ to $d = 0.19$.

Magnitude of causal effects. A more important metric is the magnitude of the remaining causal effect (Figure 10). Even if the apparent effect of a risk factor were reduced by half from $d = 2.00$ (correlational) to $d = 1.00$ (causal), this risk factor would still reflect a major developmental influence. Mean causal effects of maternal depression on child behavior were as follows: mother-reported externalizing behavior ($d = 0.13$), mother-reported internalizing behavior ($d = 0.10$), secondary-caregiver-reported externalizing behavior ($d = 0.06$) and internalizing behavior ($d = 0.05$), teacher-reported externalizing behavior ($d = 0.19$) and internalizing behavior ($d = 0.10$). There was some apparent variability in the causal effect size over time (Figure 16). However, results of mixed modeling indicated that the interaction of maternal depression and child age was not statistically significant and near zero in magnitude, failing to support the notion that the causal effect varied materially over time (Table 3).

Again, the effects on measures at child age 3 may be those most directly comparable to the existing literature. The causal effect of maternal depression on child behavior at age 3 was as follows: mother-reported externalizing behavior ($d = 0.18$) and internalizing behavior ($d = 0.17$), secondary-caregiver-reported externalizing behavior ($d = 0.06$) and internalizing behavior ($d = 0.02$).

Overall, the results are consistent with a very small to small causal effect of having a mother with CES-D total score greater than 15 at child age 2 years on children's later broadband externalizing and internalizing behavior, as rated by mothers, secondary caregivers, and teachers. 66 of 80 estimated adjusted (i.e., causal) effects had positive sign, indicating higher levels of behavior problems among children with depressed mothers. However, only 5 of 80 reached statistical significance, suggesting that larger

sample sizes may be needed to reliably estimate them. Indeed, if I assume that the true effect size is equal to $d = 0.10$, more than 3,000 cases would needed to have 80% power to detect this effect using a two-sample *t*-test (Cohen, 1988).

There was no relation of maternal depression at child age 2 with later child-reported externalizing and internalizing problems, either before ($d = 0.02$) or after ($d = 0.03$) adjustment for confounding. This finding diverges from the existing literature, in which maternal depression was found to have a small to medium effect on child reported outcomes prior to adjustment. As such, it is difficult to interpret the causal effects, since I did not find a *prima facie* effect. One possible explanation ofr this result is the long interval between exposure to maternal depression (age 2) and child report (age 10 or 14). However, maternal depression was related to both caregiver and teacher reports across this same interval (i.e., at ages 10 and 14). A second possibility is that exposure to a depressed mother fundamentally changes the nature of children's self-report, creating a measurement invariance problem (De Los Reyes et al., 2015). For example, perhaps teens who have been exposed to a highly depressed mother for the duration of their childhood implicitly compares their own moods to those of their mothers when completing rating scales, thereby responding with lower symptom ratings than they would have had they not had a depressed mother. This bias might offset an increase in true symptom levels, thereby producing a null effect of maternal depression.

Comparison to past literature with experimental designs. This is the first study of which I am aware in which the causal effect of maternal depression on child behavior has been estimated. There are two literatures which may appear at first glance to have estimated this causal effect but which have not. In order to understand why the

constituent results are not directly comparable to my own, I will describe each literature in turn.

The first literature consists of studies that evaluate whether randomizing mothers to treatment for depression (e.g., CBT) affects children's behavioral outcomes. A meta-analysis of five studies found that randomization of mothers to treatment for depression has significant beneficial effects on child behavior, with a pooled $g = 0.40$ (Cuijpers, Weitz, Karyotaki, Garber, & Andersson, 2015).[2] Effects ranged from $g = 0.07$ to $g = 0.81$, perhaps reflecting differences among the included studies in intervention protocol (e.g., cognitive-behavioral therapy vs. interpersonal therapy), intervention format (e.g., group vs. individual), and age of child (e.g., infants, children, adolescents). Another meta-analysis examined whether randomization of pregnant women to (preventive) intervention for depression affected child behavior (Goodman, Cullum, Dimidjian, River, & Kim, 2018). Two studies contributed to a pooled effect on behavioral/emotional problems of $g = 0.11$, and nine studies contributed to a pooled effect on dysregulated behavior of $g = 0.19$. Thus, studies in this first literature suggest that treatment of maternal depression produces small to moderate improvements in child behavior.

However, these studies estimate the causal effect of *therapy* on child outcomes, not the causal effect of *depression* on child outcomes. If therapy for the mother has any impact on child behavior that is *not* mediated by reductions in maternal depression, these two effects will not be identical. For example, mothers receiving cognitive-behavioral

[2] g is a version of the standardized mean difference that adjusts for its positive bias in small samples sizes (Borenstein et al., 2009). Functionally, these values of g can be compared to the values of d used in the rest of this manuscript.

therapy may reduce their use of harsh parenting, which in turn improves child behavior (even in the absence of any change in maternal depression). In this case, the *prima facie* effect of therapy for mother on child behavior would reflect both the effect of maternal depression and effect of harsh parenting. Tests of whether reductions in depression mediate the effects of therapy for mother on child behavior have yielded inconsistent results (Goodman et al., 2018; Gunlicks & Weissman, 2008), suggesting there may indeed be other mediators involved in this process.

The second literature consists of studies that evaluate whether the effect of treatment on child outcomes is mediated by changes in maternal depression in the context of a randomized, controlled trial. For example, several studies have shown that randomization to a parenting intervention (*X*) reduces mother's depressive symptoms (*M*), which, in turn, reduces children's behavior problems (*Y*) (e.g., Reuben, Shaw, Brennan, Dishion, & Wilson, 2015; Shaw et al., 2009). Since the effect on child behavior is mediated by maternal depression, this design appears to indicate a causal effect of maternal depression on child behavior. However, the effect of maternal depression is correlational rather than causal (Imai, Keele, & Tingley, 2010; Valente, Pelham III, Smyth, & MacKinnon, 2017; Valeri & VanderWeele, 2013).

Since participants are randomized to the parenting intervention (*X*), the effect of the parenting intervention on maternal depression (*M*) and the effect of the parenting intervention on child behavior (*Y*) can both be given causal interpretation. Since participants are *not* randomized to level of maternal depression, the effect of maternal depression on child behavior might still be explained by confounders that are correlated with both variables (e.g., socioeconomic status, poor marital quality). In fact, the

mediated effect will only be causal when the mediator and outcome are uncorrelated *within* each of the control and intervention groups. Since maternal depression and child behavior exhibit substantial correlation ($r > 0.20$) absent the presence of any intervention (Connell & Goodman, 2002; Goodman et al., 2011), designs that test maternal depression as a mediator of intervention effects on child behavior do not produce evidence of a direct causal effect of maternal depression on child behavior.

On Child Behavior Causing Maternal Depression

One of the alternative explanations for the *prima facie* effect of maternal depression on child behavior is that causation may flow in the opposite direction: child behavior causes maternal depression (see panel B of Figure 2). However, none of my analyses explicitly tested whether child behavior causes maternal depression. To estimate the causal effect of child behavior on maternal depression, I would invert the design, creating groups of families at child age 2 that differed on the child's level of behavior problems (e.g., none, low, medium, and high) but were equivalent on other characteristics related to both child behavior and maternal depression (e.g., marital quality, socioeconomic status). I would then compare the depression of mothers in these groups over time, at child ages 3, 4, 5, and so on. Such an analysis would be straightforward to do in this dataset and may be a valuable topic for future work.

Although I did not estimate the causal effect of child behavior on maternal depression, I did attempt to rule out child behavior as an explanation for the adjusted effect of maternal depression on later child behavior. I equated the depressed and non-depressed groups on a broad array of child characteristics, including age, sex, race, and several mother-rated domains of current behavior problems (see Table 1, Table 2). Thus,

differences in child behavior at baseline are not plausible explanations for the causal effects of maternal depression on later child behavior that were estimated in my design.

Developmental Considerations

A developmental psychopathology perspective suggests that the timing of exposure to maternal depression and of measurement of child behavior is an important consideration when interpreting these results (Goodman, 2007). Maternal depression was measured when the child was approximately 2 years old (mean = 28.2 months, *SD* = 3.29 months). Goodman and Gotlib (1999) outlined several reasons to expect that exposure to maternal depression in this age range will produce larger effects than exposure when children are older. Mothers play a larger role in regulating their children's emotions when they are younger; negative effects incurred at a younger age may have more time to "snowball" into cumulative deficits; and toddlers have less ability than older children to escape exposure to depressed mothers' affect and behavior and seek out alternative sources of support. Meta-analyses confirm that the association of maternal depression with child behavior is stronger in younger children (Connell & Goodman, 2002; Goodman et al., 2011). CBT for maternal depression during pregnancy has stronger impact on children's behavior when outcome variables are measured at a younger age (Goodman et al., 2018). Thus, the developmental timing of exposure to maternal depression in this study would be expected to increase the effect sizes observed.[3]

Child behavior was measured at child ages 3, 4, 5, 7, 8, 9, 10, and 14, spanning multiple developmental epochs. Although there were clear age trends in the level of

[3] I return to the measurement of maternal depression in later sections.

child problem behavior, I did not find evidence for age trends in the effect of maternal depression on child problem behavior (Table 3). The effects of maternal depression did not consistently vary depending on whether children were evaluated in early childhood (i.e., ages 3, 4, 5), middle childhood (i.e., ages 7, 8, 9, 10), or adolescence (i.e., age 14) (Figure 16). The meta-analyses described above did not report on the lag between measurement of maternal depression and child behavior, so it is difficult to contextualize these findings within the previous literature.

There were no measurements of child behavior between age 2 and age 3, an interval in which development is occurring rapidly. Measurements in this interval would have enabled me to track the evolution of effects of maternal depression during a time more proximal to when it was measured at age 2. Maternal depression at age 2 was significantly related to measurements of externalizing behavior as late as age 14 and internalizing behavior as late as age 10, suggesting that the effect did not disappear before outcomes were measured.

Remaining Confounders of the Causal Effect of Maternal Depression on Child Behavior

Although I adjusted for 89 measured variables that could potentially confound the relation of maternal depression to child behavior, I was not able to adjust for all possible confounders. I discuss three key examples below. In all cases, properly accounting for these remaining confounders would be expected to further attenuate the causal effect of maternal depression at child age 2.

Genetic influences. I adjusted for GWAS scores indexing risk for aggression in early and middle childhood. However, GWAS scores account for only a small portion of

the total genetic heritability that is revealed by twin designs (Manolio et al., 2009), and I did not include a GWAS score measuring risk for internalizing behavior.[4] Thus, even after adjustment through matching or weighting, the children of depressed mothers probably had higher genetic risk for externalizing and particularly internalizing behavior than did the children of non-depressed mothers.

However, to the extent that genetic influences on child behavior are mediated by variables I adjusted for at child age 2, their confounding influence will be accounted for. For example, I equated the depressed and non-depressed groups on child externalizing behavior at age 2. The portion of genetic risk that is already manifest in the measure of child externalizing behavior at age 2 will thus be accounted for, whereas latent risk that has not yet manifest in the covariate pool will not be accounted for.

Mother's history of maternal depression prior to child age 2. Mothers did not report on their depression earlier in the child's life (e.g., at age 1) or when the child was *in utero*. Both perinatal and postnatal exposure to depression are associated with increased risk of behavioral disturbance (Goodman & Gotlib, 1999; Goodman & Halperin, 2018; Stein et al., 2014). Since the mean level of mothers' depression was relatively stable over the duration of the study (Table A6), it seems likely that even after adjustment by matching or weighting, children in the depressed group were exposed to higher levels of maternal depression *in utero* and prior to age 2. As with genetic influences, some portion of the confounding effect of mother's history of maternal

[4] GWAS scores for internalizing behavior are being created in this study and will be available at a later date.

depression may already have been captured in the set of covariates I balanced on at child age 2.

Common informant bias. Many of the largest adjusted effects were observed on mother report of child behavior. This may reflect common informant bias, since mothers reported both their own level of depression and their child's behavior. However, this could also be explained by mothers being more accurate reporters than secondary caregivers, or mothers reporting unique effects on home behavior that would not be captured in teacher ratings of school behavior (De Los Reyes et al., 2015; De Los Reyes, Thomas, Goodman, & Kundey, 2013; Funder & West, 1993).

Conceptualization of Causation

Consistent with the potential outcomes approach, this design sought to mimic a hypothetical randomized experiment in which mothers were randomly assigned to levels of maternal depression, but were otherwise similar across all characteristics at baseline (Rubin, 2007, 2008). Accordingly, I estimated the causal effect of depression *net* of the effects of all the covariates in the balancing pool (i.e., Table A3). My analysis asked the following question:

> Do later outcomes differ in children with (a) depressed vs. (b) non-depressed mothers at child age 2, *if these families were similar on current income, child aggression, parenting laxness, neighborhood danger, and the rest of the covariates on which I balanced*?

This question may not be sensible if covariates in the balancing pool are in fact part of the construct of depression. For example, it would not be sensible to estimate the effect of maternal depression while equating the depressed and non-depressed mothers on a

measure of hopelessness, because hopelessness is a core feature of the construct of depression.

All covariates were measured concurrently in a questionnaire/assessment battery conducted at child age 2. Measurement of mother's depression was part of this same battery. This procedure ensured that I did not equate the depressed and non-depressed groups on a measurement that itself had been affected by maternal depression, which would produce biased and difficult to interpret estimates (Rosenbaum, 1984; Rubin, 2004). The groups were *not* equated on covariates that vary over time between baseline and outcome measurement. For example, even after successfully equating the groups on parenting overreactivity at child age 2 ($d = 0.03$ after propensity score matching), mothers in the depressed group reported greater ($d = 0.22$) parenting overreactivity at child age 3. This underscores the fact that equating the depressed and non-depressed mothers on total daily hassles at child age 2 does not prohibit parenting overreactivity from mediating subsequent effects on child behavior. To further illustrate this point, Figure A3 shows standardized mean differences on a sampling of other covariates measured at age 3.

In selecting the covariates on which to balance, I strove to avoid including variables that would produce interpretive quandaries. I equated the depressed and non-depressed groups only on covariates measured concurrently with depression at child age 2 and only on covariates that I felt could be clearly distinguished from the construct of depression. For example, although a measure of general life satisfaction was collected (Crnic, Greenberg, Ragozin, Robinson, & Basham, 1983), I did not include it as covariate because I conceptualized poor life satisfaction as a core feature of the construct of

depression. Even after matching, the depressed and non-depressed groups differed meaningfully on life satisfaction at child age 2 (d = 0.20), illustrating how one's definition of the "treatment" will by necessity define the causal effect that is estimated (Holland, 1986). Others may conceptualize depression differently and disagree with the variables I included in the balancing pool. This issue could be addressed by rerunning analyses while permitting the depressed and non-depressed groups to be imbalanced (i.e., not matched) on controversial covariates, then reporting both sets of results side by side (see Future Directions).

Measurement of Maternal Depression

Given this study's measurement of maternal depression, I estimated the causal effect of having a mother with CES-D total score greater than or equal to 16 *during the past week at a single measurement when child was 2 years old*. More severe symptomatology (e.g., mother has CES-D total score > 30 or has been diagnosed with Major Depressive Disorder) or more sustained exposure (e.g., mother is depressed at all of child ages 2, 3, 4, and 5) may be associated with larger negative effects on children.

However, several lines of evidence support the expectation that maternal depression as measured in this study would show effects on child behavior. First, Goodman et al.'s (2011) meta-analysis of the effect of maternal depression on child behavior suggests that effect sizes are similar when depression is measured with (a) depressive symptoms vs. (b) clinical diagnosis: d = 0.52 vs. 0.45 for child internalizing behavior, and d = 0.43 vs. 0.43 for child externalizing behavior. 70% of the studies included in their meta-analysis used a depression rating scale, suggesting that a large portion of the available literature has detected effects using rating scales. Second, the

correlation of maternal depression across study waves was moderate to high in magnitude (Table A6), indicating depression was relatively stable over time. Correlations of mothers' CES-D total scores at child age 2 with their CES-D total scores at later child ages ranged from 0.31 (age 2 with age 14) to 0.43 (age 2 with age 3). These numbers are similar to the test-retest correlations observed in the original CES-D development samples (0.32 and 0.49) over the interval of just one year (Radloff, 1977). This finding suggests that although the questionnaire assessed symptoms over just the past week, the mothers were not reporting on a transient fluctuation in mood. Third, children of depressed (vs. non-depressed) mothers at child age 2 were also exposed to greater levels of depression over the remainder of their childhood. Even after matching, mothers in the depressed group exhibited more depressive symptoms at *all* later waves, with differences ranging from $d = 0.38$ to $d = 0.73$ (Figure 17).[5] Moreover, the mean CES-D total score in the matched depressed group remained above the clinical cutoff at all timepoints (Figure 17). In other words, the effect size difference in level of depression between the depressed and non-depressed mothers was moderate in magnitude at ages 3, 4, 5, 7, 8, 9, 10, and 14. As such, children's longitudinal outcomes reflect a sustained mean difference in exposure.

Since the *unadjusted* effects of maternal depression were detected on nearly all outcomes and were quite like those indicated by meta-analysis of past literature, the measurement of maternal depression does not seem a plausible explanation for why my

[5] Regression to the mean (Campbell & Kenny, 1999) would be expected to reduce the difference between the depressed and non-depressed groups defined at child age 2 at later waves. Here, the difference in mean level of depression was reduced from $d = 1.60$ at child age 2 to between $d = 0.38$ and $d = 0.73$ at subsequent waves.

adjusted effects differ from past literature. This difference is most likely explained by confounding variables. However, one alternative possibility is that, although depressive symptoms and MDD correlate with child outcomes at a similar magnitude, the symptom-outcome relationship may be driven more by confounding factors than the MDD-outcome relationship. For example, perhaps elevated maternal depressive symptoms are more likely than MDD to arise from stressful life events that also cause child misbehavior, such as a major change in family living situation (e.g., moving to a dangerous neighborhood). Adjusting for stressful life events (a confounding variable) would have less impact on the relation of MDD to child outcomes than on the relation of depressive symptoms to child outcomes.[6] In this way, although adjustment for confounding variables eliminated much of the effect of depression as defined by symptom rating scales, it might have eliminated less of the effect of depression had it been defined by clinical diagnoses.

High-Risk Nature of Early Steps Sample

The Early Steps sample is a high-risk sample from a high-risk population: families in in the Women, Infants, and Children (WIC) Nutritional Supplement program. Mothers are eligible for the WIC program when they are pregnant or have a child up to age 5 years, report gross income at or below 185% of the U.S. Poverty Income Guidelines, and are deemed to exhibit nutritional risk. These eligibility criteria increase

[6] Another way of framing this idea is to consider that compared to Major Depressive Disorder, depressive symptoms might reflect greater state (vs. trait) variance (Steyer, Ferring, & Schmitt, 1992; Steyer, Mayer, Geiser, & Cole, 2015). Confounding factors such as living situation might explain more state variance in mood than trait variance in mood.

the prevalence of low-income, single-parent families, so participants in WIC would be expected to exhibit greater effects of maternal depression on child behavior than families in the general population (Goodman et al., 2011).

Study participants were WIC families that displayed multiple risk factors for child conduct problems (Dishion et al., 2008). One of these potential risk factors was mother having a CES-D total score greater than or equal to 9. Thus, it is possible that the screening procedure led to restriction of range on maternal depression, which could have attenuated the magnitude of the causal effects (Bland & Altman, 2011). When compared to screened-out families, screened-in families reported lower monthly income (~ $1700 vs. ~ $1960) and were more likely to be unmarried mothers (36% vs. 45%), two factors associated with increased effects of maternal depression (Goodman et al., 2011). Past literature suggests that the effect of maternal depression would be stronger in the WIC population than the general population, and stronger in the study participants than in the WIC population.

Strengths and Weaknesses

Strengths of this study include its large sample size ($N = 707$), use of multiple raters of children's behavioral outcomes (mother, secondary caregiver, teacher, and child), prospective design, and extended period of follow-up (eight measurements over the span of 12 years). Most importantly, a broad array of measures was collected at baseline, such that I could adjust for 89 variables that could potentially confound the relation of maternal depression to subsequent child behavior. As described above, the unadjusted effects of maternal depression were quite similar to those expected based on previous meta-analyses (Goodman et al., 2011), increasing confidence that findings are

not due to an aberrant pattern of maternal depression/child behavior relations in the current sample. Finally, the depressed and non-depressed groups were equated and compared in local conditions (i.e., drawn from same sample, measured at the same time, completed an identical battery of covariates), a feature of observational studies that produce estimates closer to the true causal effect (Cook, Shadish, & Wong, 2008).

The biggest weakness of this study, the measurement of maternal depression, has already been discussed. Other weaknesses arise from the measurement of child outcomes. First, secondary caregivers were a heterogeneous group and likely possessed mixed ability to accurately rate child behavior (e.g., 16 to 20% were not living with the child at the time of reporting).[7] Second, there were substantial missing data on the teacher report of child behavior (48 to 61%), limiting statistical power to detect effects. Third, because maternal depression was initially measured when the child was 2 years old, non-caregiver informants were not available until age 7 (for teacher) and 10 (for child). Finally, measures that were administered repeatedly (e.g., CBCL) did not satisfy measurement invariance (e.g., changing items, changing teachers, changing secondary caregivers), limiting my ability to compare effect sizes over time (Millsap, 2011).

Future Directions

Define maternal depression using more extreme symptom thresholds. More severe maternal depression may have larger causal effects on child behavior. In defining the groups, I used the standard cutpoint of a CES-D total score ≥ 16, the value proposed

[7] This could be addressed in future analyses by specifying a set of relations that are deemed more likely to be valid reporters (e.g., only biological fathers), or by analyzing outcomes separately by relation of secondary caregiver.

by Lewinsohn et al. (1997) as indicating a significant risk of Major Depressive Disorder. This cutpoint produced a large difference (d = 1.60) between the depressed and non-depressed groups in mothers' CES-D total score at child age 2 (Figure 17). However, results may still have been attenuated by dichotomization of the continuous measure (MacCallum, Zhang, Preacher, & Rucker, 2002). This design could be repeated while applying more stringent cutoffs that increase the separation between groups (i.e., extreme groups design; see Preacher, Rucker, MacCallum, & Nicewander, 2005 for discussion of advantages and limitations of this approach). For example, the non-depressed group could be defined as mothers having CES-D total score less than or equal to 10, and the depressed group could be defined as mothers having CES-D total score greater than or equal to 20.

Define maternal depression using multiple timepoints. More sustained maternal depression may have larger causal effects on child behavior. Sustained depressed vs. non-depressed groups could be defined using measurements of maternal depression at both child ages 2 and 3. This would increase reliability of measurement and minimize the impact of regression to the mean. 172 mothers obtained a total CES-D score greater than or equal to 16 at both age 2 and age 3, and 253 mothers did so at neither age 2 nor age 3. Thus, the sample size would be 425 (prior to matching) when investigating if results changed when depression was defined as a two-year sustained exposure. Such an analysis would require careful thought about how to appropriately control for covariates measured at both child age 2 and child age 3.

Investigate moderators of causal effects. A number of mother, child, and family characteristics have been shown to moderate the correlation of maternal

depression with children's behavioral outcomes (Goodman et al., 2011). My findings of very small overall effects of maternal depression may mask meaningful variability that includes more substantial effects on certain children. The current design could be used to evaluate whether the same characteristics that moderate the correlation of maternal depression and child behavior also moderate the causal effect. By testing the interaction of maternal depression and child sex within the matched or weighted samples, I could evaluate whether the causal effect of maternal depression is also greater for girls than boys. One moderator of particular interest might be child's level of externalizing and internalizing behavior at baseline. Perhaps maternal depression has larger effects on the behavior of the most (or least) at-risk children.[8]

Use latent variable modeling to integrate multiple informants' reports of child behavior. Children's longitudinal behavioral outcomes were assessed separately by informant. An alternative strategy would be to use latent variable modeling to integrate multiple informants' reports collected at the same timepoint. Mother and secondary caregiver reports could be integrated at child ages 2, 3, 4, and 5, and then mother, secondary caregiver, and teacher reports could be integrated at child ages 7, 8, 9, and 10. This approach might improve statistical power to detect the effect of maternal depression and help to summarize results. However, it is not straightforward to identify and interpret such effects (Eid et al., 2008).

Rerun analyses while permitting imbalance on subsets of covariates at child age 2. Analyses treated all covariates in the balancing pool at child age 2 (i.e., those

[8] This may be particularly relevant given that this was not a clinical sample, and thus there are low levels of behavior problems overall (Table A4).

listed in Table A3) equivalently. Instead, covariates could be organized into conceptually similar sets, such as demographics, characteristics of living situation, mother characteristics, and child characteristics. Analyses could then be repeated while equating the depressed and non-depressed groups on all but one of these sets of covariates (Cohen, Cohen, West, & Aiken, 2013). This approach could (a) shed light on which set of covariates was driving the attenuation of the *prima facie* effect and (b) address concerns of readers who thought that the depressed or non-depressed groups should *not* be equated on some of our covariates (e.g., a controversial batch).

Implications for Clinical Practice

If the true causal effect of maternal depression at age 2 on children's behavior is indeed very small (e.g., $d = 0.10$, or $r = 0.05$), this implies that treatment of maternal depression *alone* should not be expected to resolve (or substantially reduce) child behavior problems. For example, if CBT reduces maternal depression with $d = 0.60$ (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016), an effect of maternal depression on child behavior of $r = 0.05$ will convey a difference of only $d = 0.03$ to child externalizing or internalizing behavior.[9] This notion is consistent with the finding that treatment of maternal depression has only a small effect on child behavior (Cuijpers et al., 2015; Goodman et al., 2018). It is also consistent with the finding that concurrent

[9] Let assignment to CBT be binary, and characterize its effect on maternal depression as a standardized difference (here, $d = 0.60$, based on the meta-analysis of Cuijpers et al., 2016). Level of maternal depression following CBT will not be binary, so its effect on child behavior must be characterized with a correlation (here, $r = 0.05$, which is the equivalent of the hypothetical $d = 0.10$ on the correlation metric). Then the effect of CBT on child behavior by way of maternal depression is calculated as $d = 0.60$ multiplied by $r = 0.05$, producing $d = 0.03$.

treatment of maternal depression and child behavior problems does not improve child behavior beyond treatment of child behavior problems alone (e.g., Chronis, Gamble, Roberts, & Pelham Jr., 2006; Chronis-Tuscano et al., 2013). As such, when families present with comorbid maternal depression and child behavior problems, a treatment package that address both concerns should be used (e.g., CBT plus behavioral parent training; Goodman & Garber, 2017).

Implications for Developmental Science

Results suggest that a substantial, consistently observed association of developmental interest (i.e., maternal depression and child behavior) may be explained in large part by confounding variables. Studies seeking to explicate the developmental pathways that explain or include this relation may be compromised if the *prima facie* effect is not equal to the causal effect. Variables that appear to mediate the effect of maternal depression on child behavior may in fact be mediating the correlational component of this association (e.g., covariation driven by family's socioeconomic status) rather than the causal component. Similarly, maternal depression may appear to mediate the effect of another developmental influence on child behavior, when in fact it simply proxies for other methods of transmission. Consider the recent report of Choi et al. (2018), which documented how postpartum and later maternal depression mediate the intergenerational transmission of child maltreatment. Inspection of their model (Figure 18) makes it clear that failure to account for variables confounding the relation of maternal depression to child behavior could have profound impact on the network of resulting estimates and theoretical conclusions.

My results also show how modern causal inference methods (e.g., propensity score matching) can be invaluable tools for developmental psychopathology, a field in which experimentation is often difficult or impossible. Statistical techniques that are commonly used to answer developmental questions (e.g., mediating process analysis; Mackinnon, 2008) can be augmented with causal inference methods (e.g., Valente et al., 2017) to give a stronger causal interpretation to otherwise correlational results. If developmental scientists can increase the correspondence between the effects they are observing and the true causal effects, their theories will become more accurate.

Finally, the presence of a large non-causal component of a *prima facie* association has profound implications for the planning of studies that seek to explore causal rather than correlational relations. If the true causal effect of maternal depression were d = 0.10, more than 3,000 cases would be required to detect this effect with 80% power using a two-sample *t*-test (Cohen, 1988). If the true causal effect were $d = 0.20$, more than 750 cases would be required. Since these samples sizes are uncommon in psychological research, these results imply that investigators interested in causal pathways that include maternal depression should focus on design and analysis techniques that can improve statistical power (e.g., repeated measurement of child outcomes, inclusion of outcome-predictive covariates when modeling, the use of twin designs; see Shadish, Cook, and Campbell, 2002). If the causal effect is truly this small, meta-analysis of multiple studies may be the optimal way to reliably probe the causal effect of maternal depression on child behavior.

Conclusion

Findings are consistent with the claim that there is a very small causal effect of exposure to maternal depression at child age 2 on child externalizing and internalizing behavior in early childhood, middle childhood, and adolescence. This study (N = 707) was underpowered to detect effect sizes of this magnitude, so causal effects were generally not statistically significant even when potentially clinically meaningful (e.g., d = 0.20). Several limitations qualify these conclusions. Replications of this design in other samples with different timing, severity, and duration of exposure to maternal depression will help clarify findings. For now, my results suggest (a) that treatment of maternal depression should not be expected to resolve (or substantially reduce) child externalizing and internalizing behavior problems; (b) that very large sample sizes will be needed to investigate causal developmental processes that link maternal depression to child behavior; and (c) that causal inference methods can be an important addition to the toolbox of developmental psychopathologists.

REFERENCES

Abayomi, K., Gelman, A., & Levy, M. (2008). Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *57*, 273–291.

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology*, *87*, 49–74.

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA School-age Forms & Profiles: Child Behavior Checklist for Ages 6-18, Teacher's Report Form, & Youth Self-Report*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Alloy, L. B., Abramson, L. Y., Metalsky, G. I., & Hartlage, S. (1988). The hopelessness theory of depression: attributional aspects. *British Journal of Clinical Psychology*, *27*, 5–21.

Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: a measure of dysfunctional parenting in discipline situations. *Psychological Assessment*, *5*, 137–144.

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low and middle-income families. *Applied Psycholinguistics*, *19*, 209–223.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*, 399–424.

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, *33*, 1057–1069.

Beardslee, W. R., Bemporad, J., Keller, M. B., & Klerman, G. L. (1983). Children of parents with major affective disorder: a review. *The American Journal of Psychiatry*, *140*, 825–832.

Beardslee, W. R., Gladstone, T. R. G., Wright, E. J., & Cooper, A. B. (2003). A family-based approach to the prevention of depressive symptoms in children at risk: evidence of parental and child change. *Pediatrics*, *112*, e119–e131.

Beck, A. T. (1963). Thinking and depression: I. Idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, *9*, 324–333.

Beck, A. T. (1987). Cognitive models of depression. *Journal of Cognitive Psychotherapy: An International Quarterly*, *1*, 5–37.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Benazon, N. R., & Coyne, J. C. (2000). Living with a depressed spouse. *Journal of Family Psychology*, *14*, 71–79.

Bland, J. M., & Altman, D. G. (2011). Correlation in restricted ranges of data. *BMJ*, *342*, d556.

Boggs, S. R., Eyberg, S., & Reynolds, L. A. (1990). Concurrent validity of the Eyberg Child Behavior Inventory. *Journal of Clinical Child Psychology*, *19*, 75–78.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Hoboken, NJ: Wiley.

Boyd, J. H., Weissman, M. M., Thompson, W. D., & Myers, J. K. (1982). Screening for depression in a community sample: understanding the discrepancies between depression symptom and diagnostic scales. *Archives of General Psychiatry*, *39*, 1195–1200.

Brooker, R. J., Neiderhiser, J. M., Leve, L. D., Shaw, D. S., Scaramella, L. V., & Reiss, D. (2015). Associations between infant negative affect and parent anxiety symptoms are bidirectional: evidence from mothers and fathers. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01875

Bruce, M. L., & Kim, K. M. (1992). Differences in the effects of divorce on major depression in men and women. *The American Journal of Psychiatry*, *149*, 914–917.

Bullock, B., & Dishion, T. J. (2004). *Family Affective Attitude Rating Scale (FAARS)*. Eugene, OR: Child and Family Center.

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical Psychology Review*, *26*, 17–31.

Caldwell, B. M., & Bradley, R. H. (2003). *HOME Inventory Administration Manual*. Little Rock, AR: University of Arkansas for Medical Sciences and University of Arkansas at Little Rock.

Campbell, D. T., & Kenny, D. A. (1999). *A Primer on Regression Artifacts*. New York, NY: Guilford Pres.

Caughy, M. O., Randolph, S. M., & O'Campo, P. J. (2002). The Africentric Home Environment Inventory: an observational measure of the racial socialization features of the home environment for African American preschool children. *Journal of Black Psychology*, *28*, 37–52.

Cham, H., & West, S. G. (2016). Propensity score analysis with missing data. *Psychological Methods*, *21*, 427–445.

Child and Family Center. (2001). *CFC Parent Questionnaire: Service Utilization*. Eugene, OR: Child and Family Center.

Choe, D. E., Shaw, D. S., Brennan, L. M., Dishion, T. J., & Wilson, M. N. (2014). Inhibitory control as a mediator of bidirectional effects between early oppositional behavior and maternal depression. *Development and Psychopathology*, *26*, 1129–1147.

Choi, K. W., Houts, R., Arseneault, L., Pariante, C., Sikkema, K. J., & Moffitt, T. E. (2018). Maternal depression in the intergenerational transmission of childhood maltreatment and its sequelae: testing postpartum effects in a longitudinal birth cohort. *Development and Psychopathology*, 1–14.

Chronis, A. M., Gamble, S. A., Roberts, J. E., & Pelham Jr., W. E. (2006). Cognitive-behavioral depression treatment for mothers of children with Attention-Deficit/Hyperactivity Disorder. *Behavior Therapy*, *37*, 143–158.

Chronis-Tuscano, A., Clarke, T. L., O'Brien, K. A., Raggi, V. L., Diaz, Y., Mintz, A. D., … Lewinsohn, P. (2013). Development and preliminary evaluation of an integrated treatment targeting parenting and depressive symptoms in mothers of children with Attention-Deficit/Hyperactivity Disorder. *Journal of Consulting and Clinical Psychology*, *81*, 918–925.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associaties.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. New York, NY: Routledge.

Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology*, *67*, 631–647.

Connell, A. M., & Goodman, S. H. (2002). The association between psychopathology in fathers versus mothers and children's internalizing and externalizing behavior problems: a meta-analysis. *Psychological Bulletin*, *128*, 746–773.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*, 724–750.

Coyne, J. C. (1976a). Depression and the response of others. *Journal of Abnormal Psychology*, *85*, 186–193.

Coyne, J. C. (1976b). Toward an interactional description of depression. *Psychiatry*, *39*, 28–40.

Coyne, J. C., Kessler, R. C., Tal, M., Turnbull, J., Wortman, C. B., & Greden, J. F. (1987). Living with a depressed person. *Journal of Consulting and Clinical Psychology*, *55*, 347–352.

Crnic, K. A., & Greenberg, M. T. (2008). Minor parenting stresses with young children. *Child Development*, *61*, 1628–1637.

Crnic, K. A., Greenberg, M. T., Ragozin, A. S., Robinson, N. M., & Basham, R. B. (1983). Effects of stress and social support on mothers and premature and full-term infants. *Child Development*, *54*, 209–217.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry*, *58*, 376–385.

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, *15*, 245–258.

Cuijpers, P., Weitz, E., Karyotaki, E., Garber, J., & Andersson, G. (2015). The effects of psychological treatment of maternal depression on children and parental functioning: a meta-analysis. *European Child & Adolescent Psychiatry*, *24*, 237–245.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*, 858–900.

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, *9*, 123–149.

Dishion, T. J., Brennan, L. M., Shaw, D. S., McEachern, A. D., Wilson, M. N., & Jo, B. (2014). Prevention of problem behavior through annual Family Check-Ups in early childhood: intervention effects from home to early elementary school. *Journal of Abnormal Child Psychology*, *42*, 343–354.

Dishion, T. J., & Kavanagh, K. (1997). *FAST: Family Assessment Task*. University of Oregon, Child and Family Center.

Dishion, T. J., & Kavanagh, K. (2003). *Intervening in Adolescent Problem Behavior: A Family-centered Approach*. Guilford Press.

Dishion, T. J., Shaw, D., Connell, A., Gardner, F., Weaver, C., & Wilson, M. (2008). The Family Check-Up with high-risk indigent families: preventing problem behavior by increasing parents' positive behavior support in early childhood. *Child Development*, *79*, 1395–1414.

Downey, G., & Coyne, J. C. (1990). Children of depressed parents: an integrative review. *Psychological Bulletin*, *108*, 50–76.

du Rocher Schudlich, T. D., Papp, L. M., & Cummings, E. M. (2011). Relations between spouses' depressive symptoms and marital conflict: a longitudinal investigation of the role of conflict resolution styles. *Journal of Family Psychology*, *25*, 531–540.

Duman, R. S., Heninger, G. R., & Nestler, E. J. (1997). A molecular and cellular theory of depression. *Archives of General Psychiatry*, *54*, 597–608.

Dumas, J. E., Gibson, J. A., & Albin, J. B. (1989). Behavioral correlates of maternal depressive symptomatology in conduct-disorder children. *Journal of Consulting and Clinical Psychology*, *57*, 516–521.

Dziak, J. J., & Henry, K. L. (2017). Two-part predictors in regression models. *Multivariate Behavioral Research*, *52*, 551–561.

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: different models for different types of methods. *Psychological Methods*, *13*, 230–253.

Field, T. (1992). Infants of depressed mothers. *Development and Psychopathology*, *4*, 49–66.

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*, *303*, 47–53.

Funder, D. C., & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: an introduction. *Journal of Personality, 61*(4), 457–476.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Goodman, S. H. (2007). Depression in mothers. *Annual Review of Clinical Psychology*, *3*, 107–135.

Goodman, S. H., Cullum, K. A., Dimidjian, S., River, L. M., & Kim, C. Y. (2018). Opening windows of opportunities: evidence for interventions to prevent or treat depression in pregnant women being associated with changes in offspring's developmental trajectories of psychopathology risk. *Development and Psychopathology*, *30*, 1179–1196.

Goodman, S. H., & Garber, J. (2017). Evidence-based interventions for depressed mothers and their young children. *Child Development*, *88*, 368–377.

Goodman, S. H., & Gotlib, I. H. (1999). Risk for psychopathology in the children of depressed mothers: a developmental model for understanding mechanisms of transmission. *Psychological Review*, *106*, 458–490.

Goodman, S. H., & Halperin, M. S. (2018). Perinatal depression as an early stress: risk for the development of psychopathology in children. In K. Harkness & E. P. Hayden (Eds.), *The Oxford Handbook of Stress and Mental Health*. London, UK: Oxford University Press.

Goodman, S. H., Rouse, M. H., Connell, A. M., Broth, M. R., Hall, C. M., & Heyward, D. (2011). Maternal depression and child psychopathology: a meta-analytic review. *Clinical Child and Family Psychology Review*, *14*, 1–27.

Goodman, S. H., & Tully, E. C. (2009). Recurrence of depression during pregnancy: psychosocial and personal functioning correlates. *Depression and Anxiety*, *26*, 557–567.

Gotlib, I. H., Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology*, *63*, 90–100.

Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The economic burden of adults with Major Depressive Disorder in the United States (2005 and 2010). *The Journal of Clinical Psychiatry*, *76*, 155–162.

Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology*, *125*, 761–768.

Gross, H. E., Shaw, D. S., Moilanen, K. L., Dishion, T. J., & Wilson, M. N. (2008). Reciprocal models of child behavior and depressive symptoms in mothers and fathers in a sample of children at risk for early conduct problems. *Journal of Family Psychology*, *22*, *22*, 742, 742–751.

Grund, S., Robitzsch, A., & Luedtke, O. (2018). mitml: Tools for Multiple Imputation in Multilevel Modeling (Version 0.3-6) [R]. Retrieved from https://CRAN.R-project.org/package=mitml

Grunebaum, H., Cohler, B. J., Kauffman, C., & Gallant, D. (1978). Children of depressed and schizophrenic mothers. *Child Psychiatry and Human Development*, *8*, 219–228.

Gunlicks, M. L., & Weissman, M. M. (2008). Change in child psychopathology with improvement in parental depression: a systematic review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 379–389.

Haaga, D. A., Dyck, M. J., & Ernst, D. (1991). Empirical status of cognitive theory of depression. *Psychological Bulletin*, *110*, 215–236.

Hails, K. A., Reuben, J. D., Shaw, D. S., Dishion, T. J., & Wilson, M. N. (2017). Transactional associations among maternal depression, parent–child coercion, and child conduct problems during early childhood. *Journal of Clinical Child & Adolescent Psychology*, *0*, 1–15.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62.

Hammen, C. (1992). Life events and depression: the plot thickens. *American Journal of Community Psychology*, *20*, 179–193.

Hammen, C., Gordon, D., Burge, D., Adrian, C., Jaenicke, C., & Hiroto, D. (1987). Maternal affective disorders, illness, and stress: risk for children's psychopathology. *American Journal of Psychiatry*, *144*, 736–741.

Heninger, G. R., Delgado, P. L., & Charney, D. S. (1996). The revised monoamine theory of depression: a modulatory role for monoamines, based on new findings from monoamine depletion experiments in humans. *Pharmacopsychiatry*, *29*, 2–11.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*, 199–236.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*, 1–28.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960.

Hong, G. (2004). *Causal-inference for multi-level observational data with application to kindergarten retention*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*, 309–334.

Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York, NY: Cambridge University Press.

Ingoldsby, E. M., & Shaw, D. S. (2002). Neighborhood contextual factors and early-starting antisocial pathways. *Clinical Child and Family Psychology Review*, *5*, 21–55.

Johnston, C., & Mash, E. J. (1989). A measure of parenting satisfaction and efficacy. *Journal of Clinical Child Psychology*, *18*, 167–175.

Judd, L. L., Akiskal, H. S., Zeller, P. J., Paulus, M., Leon, A. C., Maser, J. D., … Keller, M. B. (2000). Psychosocial disability during the long-term course of unipolar Major Depressive Disorder. *Archives of General Psychiatry*, *57*, 375–380.

Kang, J. D. Y., & Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, *22*, 523–539.

Kendler, K. S., & Gardner, C. O. (2011). A longitudinal etiologic model for symptoms of anxiety and depression in women. *Psychological Medicine*, *41*, 2035–2045.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The epidemiology of Major Depressive Disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*, *289*, 3095–3105.

Kochanska, G., Kuczynski, L., Radke-Yarrow, M., & Welsh, J. D. (1987). Resolutions of control episodes between well and affectively ill mothers and their young children. *Journal of Abnormal Child Psychology*, *15*, 441–456.

Lewinsohn, P. M. (1974). A behavioral approach to depression. In R. Friedman & M. Katz (Eds.), *The Psychology of Depression: Contemporary Theory and Research* (pp. 157–185). London, UK: John Wiley & Sons.

Lewinsohn, P. M., Seeley, J. R., Roberts, R. E., & Allen, N. B. (1997). Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and Aging*, *12*, 277–287.

Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: their reliability and validity. *Marriage and Family Living*, *21*, 251–255.

Lockwood, J. R., & McCaffrey, D. F. (2016). Matching and weighting with functions of error-prone covariates for causal inference. *Journal of the American Statistical Association*, *111*, 1831–1839.

Lumley, T. (2003). Analysis of complex survey samples. *Journal of Statistical Software*, *9*, 1–19.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*, 747.

March, J. S., & Sullivan, K. (1999). Test-retest reliability of the Multidimensional Anxiety Scale for Children. *Journal of Anxiety Disorders*, *13*, 349–358.

Matheny, A. P., Wachs, T. D., Ludwig, J. L., & Phillips, K. (1995). Bringing order out of chaos: psychometric characteristics of the Confusion, Hubbub, And Order Scale. *Journal of Applied Developmental Psychology*, *16*, 429–444.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, *9*, 403–425.

Meng, X.-L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, *79*, 103–111.

Milberger, S., Biederman, J., Faraone, S. V., Chen, L., & Jones, J. (1996). Is maternal smoking during pregnancy a risk factor for Attention Deficit Hyperactivity Disorder in children? *The American Journal of Psychiatry*, *153*, 1138–1142.

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York, NY: Routledge.

Neyman, J. (1923). On the application of probability theory to agriculture experiments: principles. *Roczniki Nauk Rolniczch*, *10*, 1–51.

Neyman, J. (1990). On the application of probability theory to agriculture experiments: essays on principles (Section 9). *Statistical Science*, *5*, 465–471.

Nicholson, J. S., Deboeck, P. R., Farris, J. R., Boker, S. M., & Borkowski, J. G. (2011). Maternal depressive symptomatology and child behavior: transactional relationship with simultaneous bidirectional coupling. *Developmental Psychology*, *47*, 1312–1323.

Nylen, K. J., Moran, T. E., Franklin, C. L., & O'Hara, M. W. (2006). Maternal depression: a review of relevant treatment approaches for mothers and infants. *Infant Mental Health Journal*, *27*, 327–343.

Oatley, K., & Bolton, W. (1985). A social-cognitive theory of depression in reaction to life events. *Psychological Review*, *92*, 372–388.

O'Leary, K. D., Christian, J. L., & Mendell, N. R. (1994). A closer look at the link between marital discord and depressive symptomatology. *Journal of Social and Clinical Psychology*, *13*, 33–41.

Pappa, I., Pourcain, B. S., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M. G., … Tiemeier, H. (2016). A genome-wide approach to children's aggressive behavior: the EAGLE consortium. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *171*, 562–572.

Pelham, W. E., Gnagy, E. M., & Milich, R. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *31*, 210–218.

Pelham, W. E., Lang, A. R., Atkeson, B., Murphy, D. A., Gnagy, E. M., Greiner, A. R., … Greenslade, K. E. (1997). Effects of deviant child behavior on parental distress and alcohol consumption in laboratory interactions. *Journal of Abnormal Child Psychology*, *25*, 413–424.

Phinney, J. S. (1992). The Multigroup Ethnic Identity Measure: a new scale for use with diverse groups. *Journal of Adolescent Research*, *7*, 156–176.

Pianta, R. C. (1995). *Child-Parent Relationship Scale*. Charlottesville, VA: University of Virginia.

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*, 702–709.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, *10*, 178–192.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundating for Statistical Computing. Retrieved from http://www.R-project.org/

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Raghunathan, T. E., Lepowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*, 85–95.

Reuben, J. D., Shaw, D. S., Brennan, L. M., Dishion, T. J., & Wilson, M. N. (2015). A family-based intervention for improving children's emotional problems through

effects on maternal depressive symptoms. *Journal of Consulting and Clinical Psychology*. doi:10.1037/ccp0000049

Rolf, J. E. (1972). The social and academic competence of children vulnerable to schizophrenia and other behavior pathologies. *Journal of Abnormal Psychology*, *80*, 225–243.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, *147*, 656–666.

Rosenbaum, P. R. (1986). Dropping out of high school in the United States: an observational study. *Journal of Educational Statistics*, *11*, 207–224.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 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*, 516–524.

Rothbart, R. M., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2003). Investigations of temperament at three to seven years: the Children's Behavior Questionnaire. *Child Development*, *72*, 1394–1408.

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

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

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

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, *6*, 34–58.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169–188.

Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, *31*, 161–170.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, *100*, 322–331.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine*, *26*, 20–36.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, *2*, 808–840.

Rubin, D. B. (2010). Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010). *Psychological Methods*, *15*, 38–46.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, *95*, 573–585.

Rutter, M. (1966). *Children of Sick Parents: An Environmental and Psychiatric Study*. London: Oxford University Press.

Saylor, C. F., Finch, A. J., Spirito, A., & Bennett, B. (1984). The Children's Depression Inventory: a systematic evaluation of psychometric properties. *Journal of Consulting and Clinical Psychology*, *52*, 955–967.

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

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*, 28–38.

Shaw, D. S., Connell, A., Dishion, T. J., Wilson, M. N., & Gardner, F. (2009). Improvements in maternal depression as a mediator of intervention effects on early childhood problem behavior. *Development and Psychopathology*, *21*, 417–439.

Smith, R. S. (1991). The macrophage theory of depression. *Medical Hypotheses*, *35*, 298–306.

Stein, A., Pearson, R. M., Goodman, S. H., Rapa, E., Rahman, A., McCallum, M., … Pariante, C. M. (2014). Effects of perinatal mental disorders on the fetus and child. *The Lancet*, *384*, 1800–1819.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *8(2)*. Retrieved from http://orbilu.uni.lu/handle/10993/12805

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits—revised. *Annual Review of Clinical Psychology*, *11*, 71–98.

Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science*, *25*, 1–21.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry*, *157*, 1552–1562.

Taylor, A. B., West, S. G., & Aiken, L. S. (2006). Loss of power in logistic, ordinal logistic, and probit regression when an outcome variable is coarsely categorized. *Educational and Psychological Measurement*, *66*, 228–239.

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*, 90–118.

Valente, M. J., Pelham III, W. E., Smyth, H., & MacKinnon, D. P. (2017). Confounding in statistical mediation analysis: what it is and how to address it. *Journal of Counseling Psychology*, *64*, 659–671.

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*, 137–150.

van Buuren, S. van. (2012). *Flexible Imputation of Missing Data*. Boca Raton, FL: CRC Press.

van Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67.

Weaver, C. M., Shaw, D. S., Dishion, T. J., & Wilson, M. N. (2008). Parenting self-efficacy and problem behavior in children at high risk for early conduct problems: The mediating role of maternal depression. *Infant Behavior and Development*, *31*, 594–605.

Weissman, M. M., Pilowsky, D. J., Wickramaratne, P. J., Talati, A., Wisniewski, S. R., Fava, M., … Team, for the S.-C. (2006). Remissions in maternal depression and child psychopathology: a STAR*D-child report. *JAMA*, *295*, 1389–1398.

West, S. G., Cham, H., Thoemmes, F., Renneberg, B., Schulze, J., & Weiler, M. (2014). Propensity scores as a basis for equating groups: basic principles and application in clinical treatment outcome research. *Journal of Consulting and Clinical Psychology*, *82*, 906–919.

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*, 18–37.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*, 377–399.

Zuckerman, B., Bauchner, H., Parker, S., & Cabral, H. (1990). Maternal depressive symptoms during pregnancy, and newborn irritability. *Journal of Developmental & Behavioral Pediatrics*, *11*, 190.

APPENDIX A

TABLES AND FIGURES

Table 1
Balance on Age 2 Variables After Matching

Variable	N		Means			Variances		
	Ctrl.	Treat.	Ctrl.	Treat.	SMD	Ctrl.	Treat.	Ratio
Binary variables								
Family assigned to intervention condition	179	179	55 %	53 %	-	-	-	-
Site is Charlottesville, VA	179	179	74 %	73 %	-	-	-	-
Site is Pittsburgh, PA	179	179	36 %	40 %	-	-	-	-
Child is male	179	179	46 %	44 %	-	-	-	-
Child is hispanic	179	179	13 %	12 %	-	-	-	-
Child is black	179	179	73 %	69 %	-	-	-	-
Child is biracial	179	179	13 %	12 %	-	-	-	-
Mother was teen parent	179	179	74 %	77 %	-	-	-	-
Mother is hispanic	179	179	11 %	9 %	-	-	-	-
Mother is black	179	179	73 %	69 %	-	-	-	-
Mother is married or living with significant other	179	179	56 %	57 %	-	-	-	-
Mother reports having a live-in partner	179	179	58 %	58 %	-	-	-	-
Mother endorses having religious / spiritual beliefs	179	179	30 %	34 %	-	-	-	-
Child has been cared for by person other than mother more than 5 hrs/wk	179	179	21 %	19 %	-	-	-	-
Family is below poverty line	179	179	22 %	26 %	-	-	-	-
Family received food stamps	179	179	61 %	63 %	-	-	-	-
Family received medical assistance	179	179	30 %	32 %	-	-	-	-
Family received social security income	179	179	8 %	9 %	-	-	-	-
Family receives child support	179	179	20 %	19 %	-	-	-	-
Family owns a home	179	179	14 %	16 %	-	-	-	-
Person in home had trouble with law since child's birth	179	179	32 %	33 %	-	-	-	-
Person in home reported for child abuse since child's birth	179	179	7 %	7 %	-	-	-	-
Person in home treated by a mental health professional since child's birth	179	179	36 %	37 %	-	-	-	-
Mother endorses support from extended family as family strength	179	179	63 %	65 %	-	-	-	-
Mother endorses employment situation as family strength	179	179	22 %	23 %	-	-	-	-
Mother endorses church, religion, spirituality as family strength	179	179	32 %	28 %	-	-	-	-
Mother endorses conflict or violence as impacting family	179	179	15 %	15 %	-	-	-	-
Mother endorses drug use by parent as impacting family	179	179	11 %	9 %	-	-	-	-
Mother endorses high crime neighborhood as impacting family	179	179	13 %	15 %	-	-	-	-
Mother endorses parent being absent as impacting family	179	179	24 %	27 %	-	-	-	-
Mother endorses stress between home and school as impacting family	179	179	12 %	13 %	-	-	-	-
Mother endorses unstable home situation as impacting family	179	179	7 %	8 %	-	-	-	-
Mother endorses death in family as impacting family	179	179	17 %	17 %	-	-	-	-
Mother endorses past traumatic experience as impacting family	179	179	11 %	12 %	-	-	-	-
Mother: no organized groups are a source of support	179	179	46 %	49 %	-	-	-	-
Mother: has not visited friends within past week	179	179	74 %	69 %	-	-	-	-
Mother: how many times per week speak with friends and family on phone	179	179	42 %	42 %	-	-	-	-
Mother: ever drinks alcohol	175	175	69 %	71 %	-	-	-	-
Mother: drinks alcohol at least monthly	175	175	28 %	30 %	-	-	-	-
Mother: drinks alcohol at least weekly	175	175	9 %	9 %	-	-	-	-
Mother: ever stopped drinking due to problems with use	174	173	7 %	7 %	-	-	-	-
Mother: ever uses marijuana	175	175	16 %	15 %	-	-	-	-
Mother: uses marijuana at least monthly	175	175	7 %	6 %	-	-	-	-
Mother: ever stopped using marijuana due to problems with use	174	173	5 %	6 %	-	-	-	-
Mother: currently smoke cigarettes	148	148	55 %	58 %	-	-	-	-
Mother: any hard drug use	175	175	7 %	9 %	-	-	-	-

	Variable	N		Means			Variances		
		Ctrl.	Treat.	Ctrl.	Treat.	SMD	Ctrl.	Treat.	Ratio
	Live-in partner: ever drinks alcohol	66	65	27 %	26 %	-	-	-	-
	Live-in partner: drinks alcohol at least monthly	66	65	61 %	63 %	-	-	-	-
	Live-in partner: drinks alcohol at least weekly	66	65	12 %	14 %	-	-	-	-
	Live-in partner: ever stopped drinking due to problems with use	63	62	10 %	11 %	-	-	-	-
	Live-in partner: ever uses marijuana	66	65	15 %	14 %	-	-	-	-
	Live-in partner: uses marijuana at least monthly	66	65	8 %	6 %	-	-	-	-
	Live-in partner: ever stopped using marijuana due to problems with use	65	63	6 %	5 %	-	-	-	-
	Live-in partner: currently smoke cigarettes	54	54	44 %	50 %	-	-	-	-
	Metric variables								
	Child age in months at baseline	179	179	28.26	28.00	-0.08	11.01	10.09	0.92
	Mother's age in years	179	179	26.58	26.32	-0.04	36.39	36.14	0.99
	Mother's level of education	179	179	5.15	5.20	0.04	1.27	1.43	1.12
	Mother's gross monthly income	179	179	3.74	3.83	0.05	3.44	4.25	1.23
	Number of people living in home	179	179	4.50	4.50	0.00	2.35	2.93	1.24
	Number of children living in home	179	179	2.44	2.42	-0.01	1.35	1.59	1.18
	Number of adults living in home	179	179	2.07	2.08	0.01	0.77	1.17	1.52
	Gross monthly income per number of persons in home	179	179	0.90	0.94	0.08	0.23	0.33	1.42
	Mother: adult-child relationship positive score (ACRS)	179	179	8.02	8.19	0.06	8.53	9.72	1.14
	Mother: adult-child relationship conflict score (ACRS)	179	179	28.13	28.61	0.07	49.25	54.92	1.12
	Mother: live-in partner relationship score (LOCKE)	76	77	57.75	56.87	-0.09	88.11	100.46	1.14
	Mother: neighborhood cohesion (MMNQ)	179	179	15.41	15.19	-0.03	55.09	71.91	1.31
	Mother: neighborhood danger (MMNQ)	179	179	8.62	8.98	0.05	56.91	67.79	1.19
	Mother: rating of child attention problems (CBC)	179	179	4.17	4.16	-0.01	3.39	3.37	0.99
	Mother: rating of child 'other' behavioral problems (CBC)	179	179	14.68	14.30	-0.07	30.94	24.61	0.80
	Mother: rating of child sleep problems (CBC)	179	179	4.77	4.59	-0.07	6.95	7.39	1.06
	Mother: rating of child somatic complaints (CBC)	179	179	2.37	2.22	-0.08	3.60	3.14	0.87
	Mother: rating of child externalizing behavior (CBC)	179	179	20.89	21.03	0.02	53.53	48.50	0.91
	Mother: rating of child internalizing behavior (CBC)	179	179	12.56	12.41	-0.03	34.54	32.93	0.95
	Mother: rating of child inhibitory control (TEMP)	176	177	3.99	3.95	-0.04	0.66	0.68	1.04
*	Mother: frequency of total daily hassles (HASSL)	179	179	46.22	47.19	0.12	60.98	68.25	1.12
	Mother: parent competency score (BEPAR)	179	179	62.59	62.84	0.02	114.56	123.22	1.08
	Mother: parenting laxness score(PARTS)	176	177	3.16	3.14	-0.02	0.85	1.08	1.27
	Mother: parenting overreactivity score (PARTS)	176	177	2.67	2.69	0.03	0.58	0.58	0.99
	Mother: perception of total daily hassles (HASSL)	179	179	47.16	48.27	0.10	134.09	177.76	1.33
	Mother: rating of child total behavior problems (EYBC)	179	179	14.32	14.66	0.05	40.68	43.01	1.06
	Mother: rating of child total intensity of behavior problems (EYBC)	179	179	126.97	129.66	0.09	826.98	761.70	0.92
	Mother: total number of words understood by child (MACDI)	179	179	60.92	61.60	0.03	709.62	681.49	0.96
	Live-in partner: total depressive symptoms (CESD)	65	64	10.35	10.19	-0.02	57.38	47.33	0.82
	Live-in partner: rating of child externalizing behavior (CBC)	64	64	15.48	15.47	-0.00	46.03	64.25	1.40
	Live-in partner: rating of child internalizing behavior (CBC)	64	64	9.61	9.64	0.01	28.91	50.33	1.74
	Home visitor: parent involvement score	179	179	2.20	2.23	0.04	0.79	0.84	1.07
	Home visitor: total chaos score	179	179	5.18	5.41	0.06	12.43	13.25	1.07
	Eagle Consortium GWAS score for early childhood aggression	124	127	2.28	2.21	-0.02	18.47	11.00	0.60
*	Eagle Consortium GWAS score for middle childhood aggression	124	127	-27.98	-27.22	0.12	43.64	35.59	0.82

Note. Ctrl. = non-depressed mothers, Treat. = depressed mothers, SMD = standardized mean difference. Binary covariates are listed first, followed by metric covariates. For binary covariates, means reflect proportions, and SMD and variances are omitted. SMD calculated as difference between means of non-depressed and depressed groups, divided by the standard deviation in the non-depressed group. Asterisk in leftmost column indicates that variable was included as covariate in subsequent outcome analyses.milberger
eeemill

Table 2
Balance on Age 2 Variables After Weighting

Variable	N		Means			Variances		
	Ctrl.	Treat.	Ctrl.	Treat.	SMD	Ctrl.	Treat.	Ratio
Binary variables								
Family assigned to intervention condition	385	322	55 %	53 %	-	-	-	-
Site is Charlottesville, VA	385	322	70 %	75 %	-	-	-	-
Site is Pittsburgh, PA	385	322	36 %	41 %	-	-	-	-
Child is male	385	322	47 %	52 %	-	-	-	-
Child is hispanic	385	322	16 %	15 %	-	-	-	-
Child is black	385	322	67 %	72 %	-	-	-	-
Child is biracial	385	322	10 %	11 %	-	-	-	-
Mother was teen parent	385	322	78 %	73 %	-	-	-	-
Mother is hispanic	385	322	12 %	13 %	-	-	-	-
Mother is black	385	322	67 %	72 %	-	-	-	-
Mother is married or living with significant other	385	322	54 %	55 %	-	-	-	-
Mother reports having a live-in partner	385	322	55 %	59 %	-	-	-	-
Mother endorses having religious / spiritual beliefs	385	322	29 %	28 %	-	-	-	-
Child has been cared for by person other than mother more than 5 hrs/wk	385	322	18 %	18 %	-	-	-	-
Family is below poverty line	385	322	23 %	20 %	-	-	-	-
Family received food stamps	385	322	62 %	58 %	-	-	-	-
Family received medical assistance	385	322	29 %	32 %	-	-	-	-
Family received social security income	385	322	9 %	9 %	-	-	-	-
Family receives child support	385	322	21 %	17 %	-	-	-	-
Family owns a home	385	322	15 %	18 %	-	-	-	-
Person in home had trouble with law since child's birth	385	322	35 %	30 %	-	-	-	-
Person in home reported for child abuse since child's birth	385	322	5 %	7 %	-	-	-	-
Person in home treated by a mental health professional since child's birth	385	322	36 %	34 %	-	-	-	-
Mother endorses support from extended family as family strength	385	322	63 %	65 %	-	-	-	-
Mother endorses employment situation as family strength	385	322	23 %	23 %	-	-	-	-
Mother endorses church, religion, spirituality as family strength	385	322	33 %	32 %	-	-	-	-
Mother endorses conflict or violence as impacting family	385	322	20 %	15 %	-	-	-	-
Mother endorses drug use by parent as impacting family	385	322	11 %	9 %	-	-	-	-
Mother endorses high crime neighborhood as impacting family	385	322	13 %	15 %	-	-	-	-
Mother endorses parent being absent as impacting family	385	322	23 %	23 %	-	-	-	-
Mother endorses stress between home and school as impacting family	385	322	13 %	12 %	-	-	-	-
Mother endorses unstable home situation as impacting family	385	322	8 %	7 %	-	-	-	-
Mother endorses death in family as impacting family	385	322	18 %	17 %	-	-	-	-
Mother endorses past traumatic experience as impacting family	385	322	12 %	13 %	-	-	-	-
Mother: no organized groups are a source of support	385	322	49 %	48 %	-	-	-	-
Mother: has not visited friends within past week	385	322	71 %	73 %	-	-	-	-
Mother: how many times per week speak with friends and family on phone	385	322	44 %	44 %	-	-	-	-
Mother: ever drinks alcohol	375	317	69 %	73 %	-	-	-	-
Mother: drinks alcohol at least monthly	375	317	29 %	32 %	-	-	-	-
Mother: drinks alcohol at least weekly	375	317	7 %	8 %	-	-	-	-
Mother: ever stopped drinking due to problems with use	368	314	5 %	6 %	-	-	-	-
Mother: ever uses marijuana	374	315	12 %	13 %	-	-	-	-
Mother: uses marijuana at least monthly	374	315	5 %	6 %	-	-	-	-
Mother: ever stopped using marijuana due to problems with use	371	310	4 %	5 %	-	-	-	-
Mother: currently smoke cigarettes	328	274	62 %	59 %	-	-	-	-
Mother: any hard drug use	375	317	6 %	7 %	-	-	-	-

	Variable	N		Means			Variances		
		Ctrl.	Treat.	Ctrl.	Treat.	SMD	Ctrl.	Treat.	Ratio
	Live-in partner: ever drinks alcohol	159	109	26 %	25 %	-	-	-	-
	Live-in partner: drinks alcohol at least monthly	159	109	61 %	58 %	-	-	-	-
	Live-in partner: drinks alcohol at least weekly	159	109	85 %	84 %	-	-	-	-
	Live-in partner: ever stopped drinking due to problems with use	156	106	12 %	13 %	-	-	-	-
	Live-in partner: ever uses marijuana	161	109	14 %	13 %	-	-	-	-
	Live-in partner: uses marijuana at least monthly	161	109	6 %	5 %	-	-	-	-
	Live-in partner: ever stopped using marijuana due to problems with use	157	106	7 %	10 %	-	-	-	-
	Live-in partner: currently smoke cigarettes	138	94	54 %	52 %	-	-	-	-
	Metric variables								
	Child age in months at baseline	385	322	27.85	28.36	0.15	11.17	11.38	1.02
	Mother's age in years	385	322	26.45	26.71	0.04	36.73	33.38	0.91
	Mother's level of education	385	322	5.15	5.09	-0.06	1.12	1.69	1.51
	Mother's gross monthly income	385	322	3.60	3.70	0.05	3.54	3.82	1.08
	Number of people living in home	385	322	4.38	4.53	0.10	2.08	3.07	1.48
	Number of children living in home	385	322	2.38	2.41	0.03	1.17	1.48	1.26
	Number of adults living in home	385	322	2.01	2.12	0.12	0.81	1.20	1.49
	Gross monthly income per number of persons in home	385	322	0.89	0.88	-0.03	0.27	0.24	0.89
	Mother: adult-child relationship positive score (ACRS)	385	322	7.91	8.02	0.04	7.36	9.69	1.32
	Mother: adult-child relationship conflict score (ACRS)	385	322	28.37	27.90	-0.07	51.27	57.90	1.13
	Mother: live-in partner relationship score (LOCKE)	195	130	58.88	58.03	-0.10	76.10	80.99	1.06
	Mother: neighborhood cohesion (MMNQ)	385	322	14.99	15.27	0.04	51.32	61.63	1.20
	Mother: neighborhood danger (MMNQ)	385	322	8.34	8.75	0.06	56.48	67.43	1.19
	Mother: rating of child attention problems (CBC)	385	322	4.07	4.01	-0.03	4.04	3.71	0.92
	Mother: rating of child 'other' behavioral problems (CBC)	385	322	13.90	14.20	0.05	32.61	32.03	0.98
	Mother: rating of child sleep problems (CBC)	385	322	4.68	4.57	-0.04	6.98	8.06	1.16
	Mother: rating of child somatic complaints (CBC)	385	322	2.23	2.32	0.05	3.47	3.70	1.07
	Mother: rating of child externalizing behavior (CBC)	385	322	20.38	20.32	-0.01	54.99	58.25	1.06
	Mother: rating of child internalizing behavior (CBC)	385	322	12.12	12.65	0.09	35.17	41.57	1.18
	Mother: rating of child inhibitory control (TEMP)	379	318	3.96	4.03	0.09	0.65	0.78	1.20
*	Mother: frequency of total daily hassles (HASSL)	385	322	47.37	46.63	-0.08	87.03	74.22	0.85
	Mother: parent competency score (BEPAR)	385	322	62.39	62.71	0.03	145.67	146.76	1.01
	Mother: parenting laxness score(PARTS)	378	318	3.18	3.18	-0.00	0.84	1.06	1.27
	Mother: parenting overreactivity score (PARTS)	378	318	2.70	2.67	-0.03	0.64	0.59	0.92
	Mother: perception of total daily hassles (HASSL)	385	322	48.86	47.42	-0.10	222.10	196.06	0.88
	Mother: rating of child total behavior problems (EYBC)	385	322	14.77	14.50	-0.04	44.39	38.69	0.87
	Mother: rating of child total intensity of behavior problems (EYBC)	385	322	126.66	128.20	0.05	810.50	731.40	0.90
	Mother: total number of words understood by child (MACDI)	385	322	57.35	59.79	0.09	724.74	732.49	1.01
	Live-in partner: total depressive symptoms (CESD)	160	108	9.33	9.53	0.03	58.40	53.60	0.92
	Live-in partner: rating of child externalizing behavior (CBC)	160	108	15.21	14.97	-0.03	50.84	59.85	1.18
	Live-in partner: rating of child internalizing behavior (CBC)	160	108	9.66	9.67	0.00	41.52	45.42	1.09
	Home visitor: parent involvement score	385	322	2.10	2.24	0.14	0.93	0.84	0.91
	Home visitor: total chaos score	385	322	5.22	5.33	0.03	13.44	13.53	1.01
	Eagle Consortium GWAS score for early childhood aggression	262	236	2.08	2.11	0.01	17.15	11.48	0.67
*	Eagle Consortium GWAS score for middle childhood aggression	262	236	-26.75	-27.19	-0.06	53.73	35.60	0.66

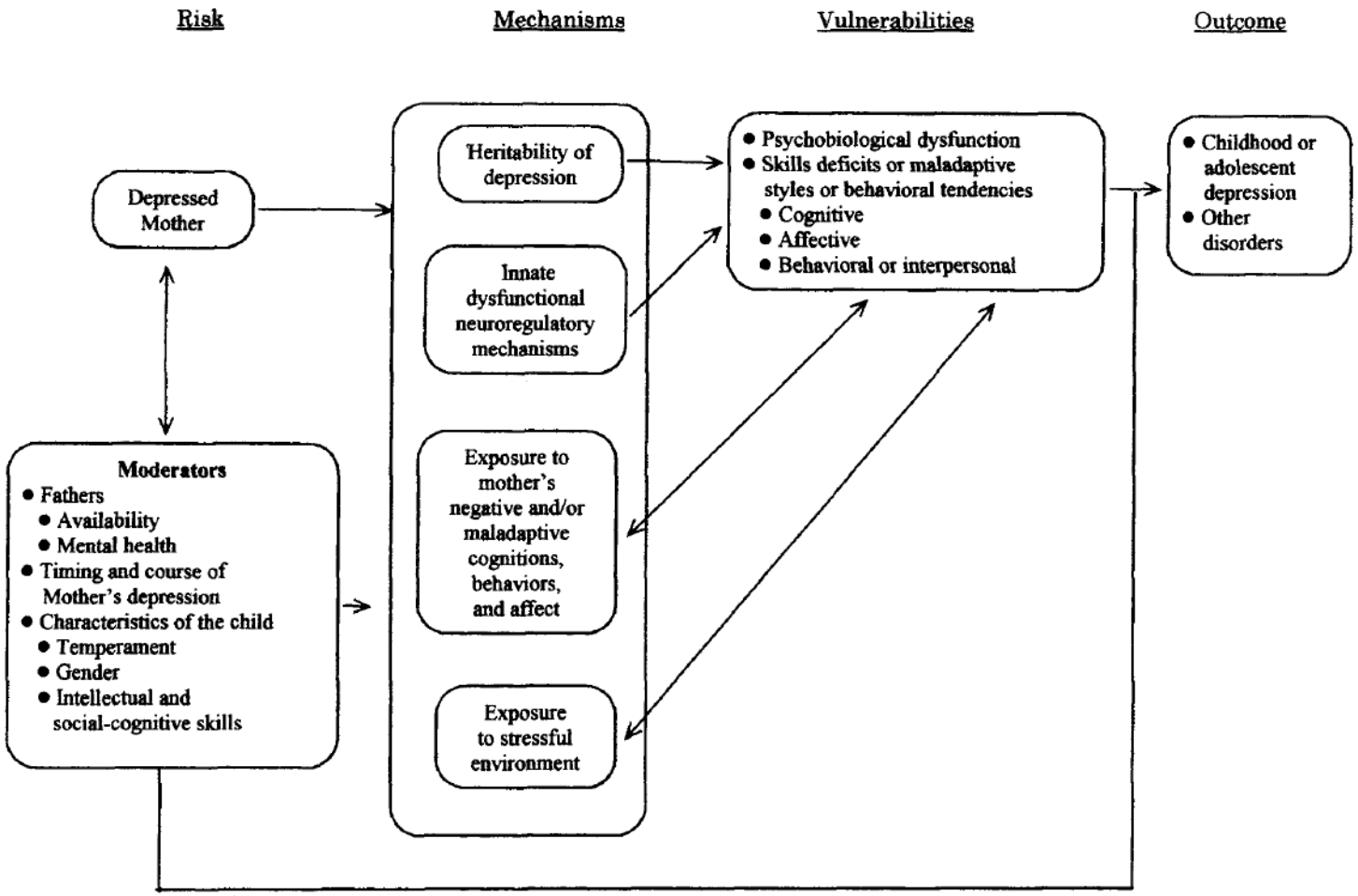
Note. Ctrl. = non-depressed mothers, Treat. = depressed mothers, SMD = standardized mean difference. Binary covariates are listed first, followed by metric covariates. For binary covariates, means reflect proportions, and SMD and variances are omitted. SMD calculated as difference between means of non-depressed and depressed groups, divided by the standard deviation in the non-depressed group. Asterisk in leftmost column indicates that variable was included as covariate in subsequent outcome analyses.

Table 3
Sequential Comparisons of Mixed Models

Informant	Measure	Scale	Method	Model 2 vs. Model 1: Adding fixed effects for age of assessment ("time effect").	Model 3 vs. Model 2: Adding fixed effect of maternal depression group ("group effect").	Model 4 vs. Model 3: Adding interaction of age of assessment and maternal depression group ("time by group interaction").	Pseudo R^2 for each model			
							M1	M2	M3	M4
Mother	Externalizing	CBCL	Unadjusted	$F(7, 42205.4) = 271.32$, $p < .05$	$F(1, 175826.5) = 50.25$, $p < .05$	$F(7, 50262.4) = 3.41$, $p < .05$	0.41	0.65	0.65	0.65
			Matched	$F(7, 41609.9) = 132.32$, $p < .05$	$F(1, 223247.0) = 1.79$, *ns*	$F(7, 50703.5) = 1.24$, *ns*	0.40	0.64	0.64	0.64
Mother	Internalizing	CBCL	Unadjusted	$F(7, 43028.5) = 29.59$, $p < .05$	$F(1, 115296.2) = 44.87$, $p < .05$	$F(7, 45394.7) = 0.97$, *ns*	0.46	0.49	0.49	0.49
			Matched	$F(7, 49617.1) = 14.33$, $p < .05$	$F(1, 127810.6) = 0.63$, *ns*	$F(7, 48465.4) = 0.69$, *ns*	0.43	0.46	0.46	0.46
2nd caregiver	Externalizing	CBCL	Unadjusted	$F(7, 11866.7) = 90.75$, $p < .05$	$F(1, 7987.2) = 34.69$, $p < .05$	$F(7, 13265.2) = 0.75$, *ns*	0.26	0.43	0.43	0.43
			Matched	$F(7, 14875.9) = 51.42$, $p < .05$	$F(1, 11514.3) = 0.73$, *ns*	$F(7, 14522.0) = 0.23$, *ns*	0.27	0.44	0.44	0.44
2nd caregiver	Internalizing	CBCL	Unadjusted	$F(7, 11087.2) = 12.52$, $p < .05$	$F(1, 5068.9) = 25.63$, $p < .05$	$F(7, 12501.4) = 0.58$, *ns*	0.27	0.29	0.28	0.29
			Matched	$F(7, 14587.7) = 6.31$, $p < .05$	$F(1, 6121.6) = 0.33$, *ns*	$F(7, 15810.4) = 0.74$, *ns*	0.25	0.27	0.27	0.27
Teacher	Externalizing	TRF	Unadjusted	$F(3, 4203.2) = 0.48$, *ns*	$F(1, 4594.4) = 17.14$, $p < .05$	$F(3, 4857.6) = 0.13$, *ns*	0.45	0.45	0.45	0.45
			Matched	$F(3, 5477.6) = 0.35$, *ns*	$F(1, 4939.4) = 5.61$, $p < .05$	$F(3, 5548.3) = 0.15$, *ns*	0.45	0.45	0.44	0.44
Teacher	Internalizing	TRF	Unadjusted	$F(3, 3468.1) = 0.40$, *ns*	$F(1, 2069.4) = 7.29$, $p < .05$	$F(3, 4469.9) = 0.86$, *ns*	0.33	0.32	0.30	0.30
			Matched	$F(3, 4879.8) = 0.19$, *ns*	$F(1, 2511.6) = 2.25$, *ns*	$F(3, 5376.1) = 0.59$, *ns*	0.33	0.31	0.31	0.30

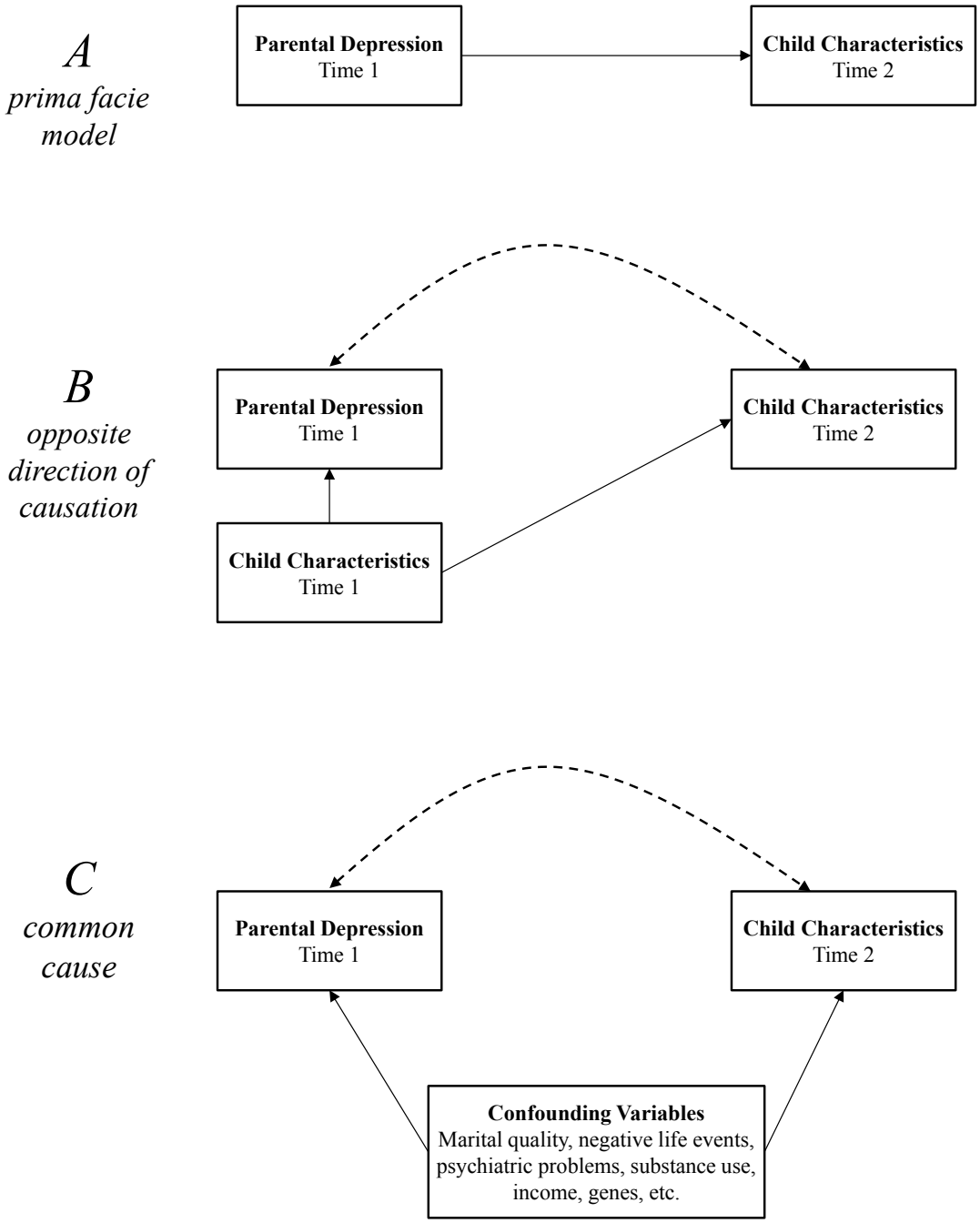
Note. Each row reports a sequence of model comparisons for a specific series of repeated measures (e.g., in first row, mother report of externalizing problems at ages 3, 4, 5, 7, 8, 9, 10, and 14). In this way, procedure mimics Type I rather than Type III tests in the analysis of variance terminology. See text of Methods for Step 2 for specification of Models 1, 2, 3, and 4. Pseudo R^2 calculated as squared correlation between actual and model-predicted response values (i.e., outcome). The change in pseudo R^2 from Model 3 to Model 4 can be conceptualized as an effect size for the interaction between maternal depression group and age of assessment. Differences were minimal, suggesting the interaction effect size is negligible. Results are not reported for weighting because survey weights cannot be incorporated into the mixed modeling framework in a straightforward way.

Figure 1
Goodman and Gotlib's (1999) Model of Transmission of Risk to Children of Depressed Mothers



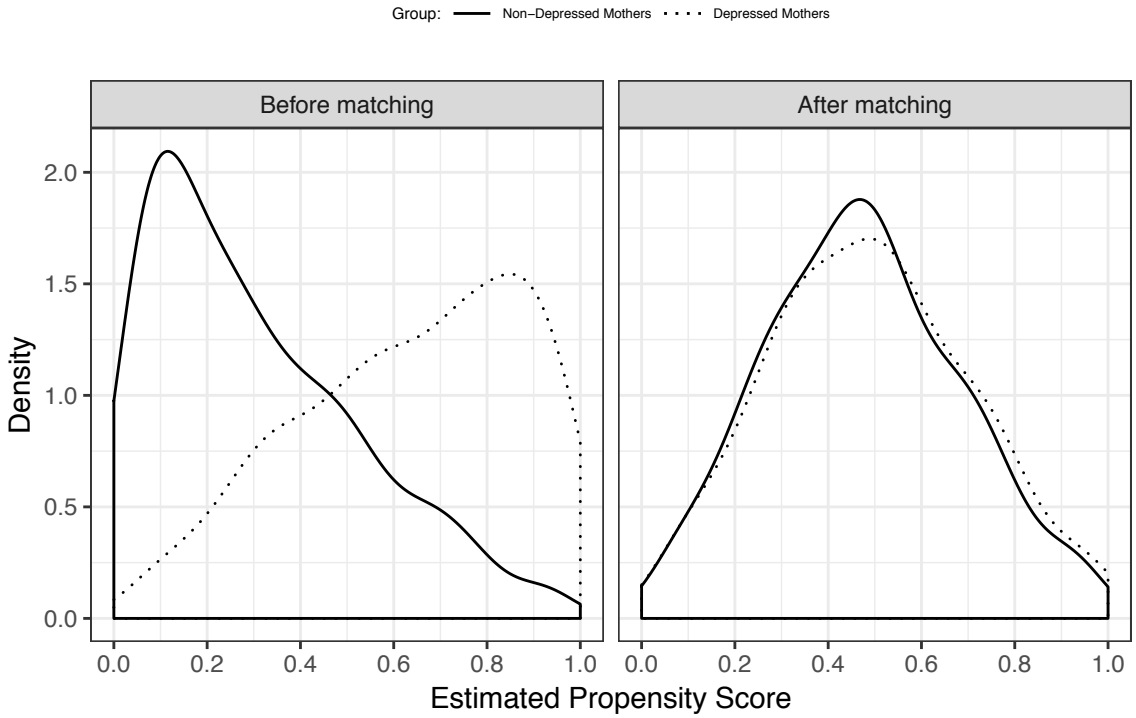
Note. Reproduced from Goodman and Gotlib (1999). Bidirectional arrows indicate bidirectional or transactional influences.

Figure 2
Causal and Alternative Models of the Relationship Between Parental Depression and Child Outcomes



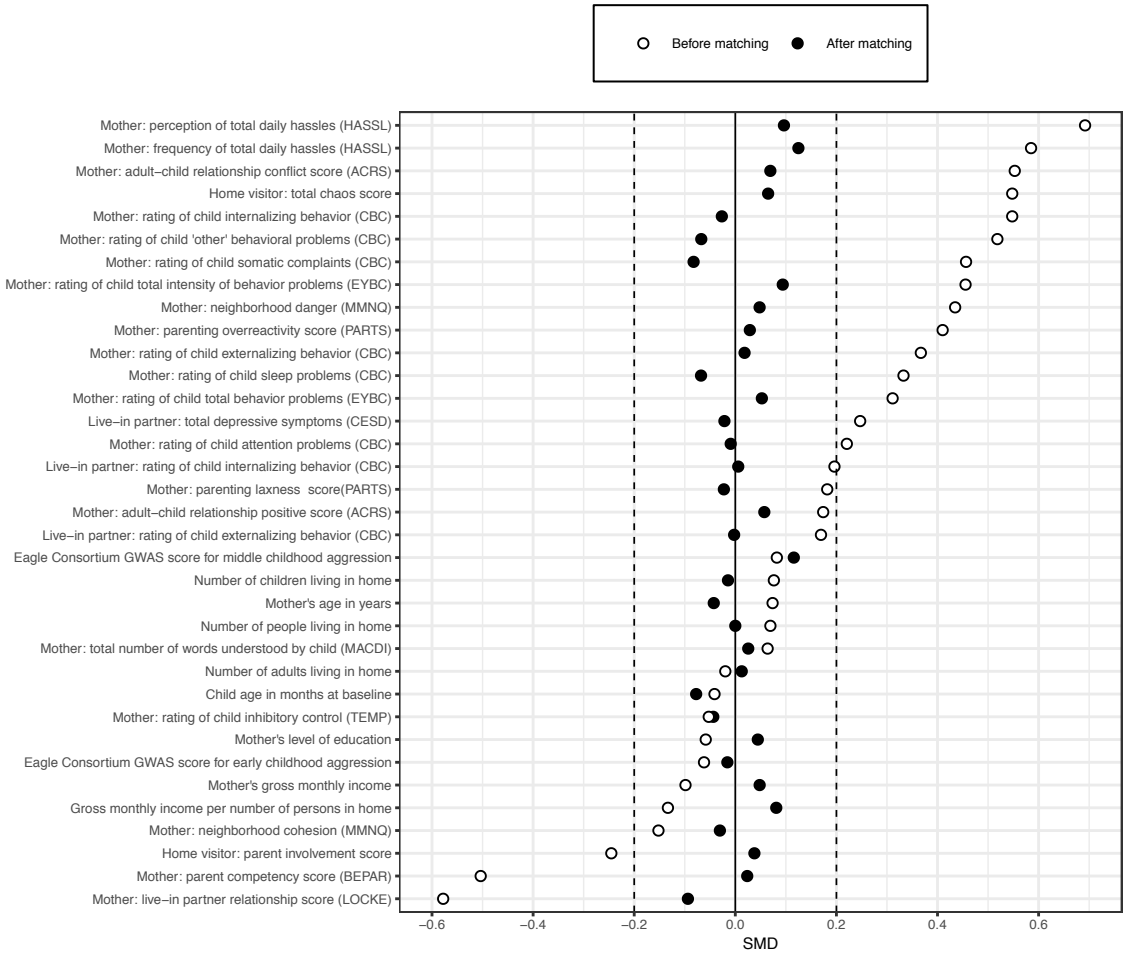
Note. Dashed lines indicate spurious (i.e., non-causal) relations. Bidirectional (or "transactional") models are not considered.

Figure 3
Distribution of Estimated Propensity Scores Before and After Matching



Note. Curve is kernel density estimate. Kernel density plot depicts estimate of the underlying distribution in each group in the population.

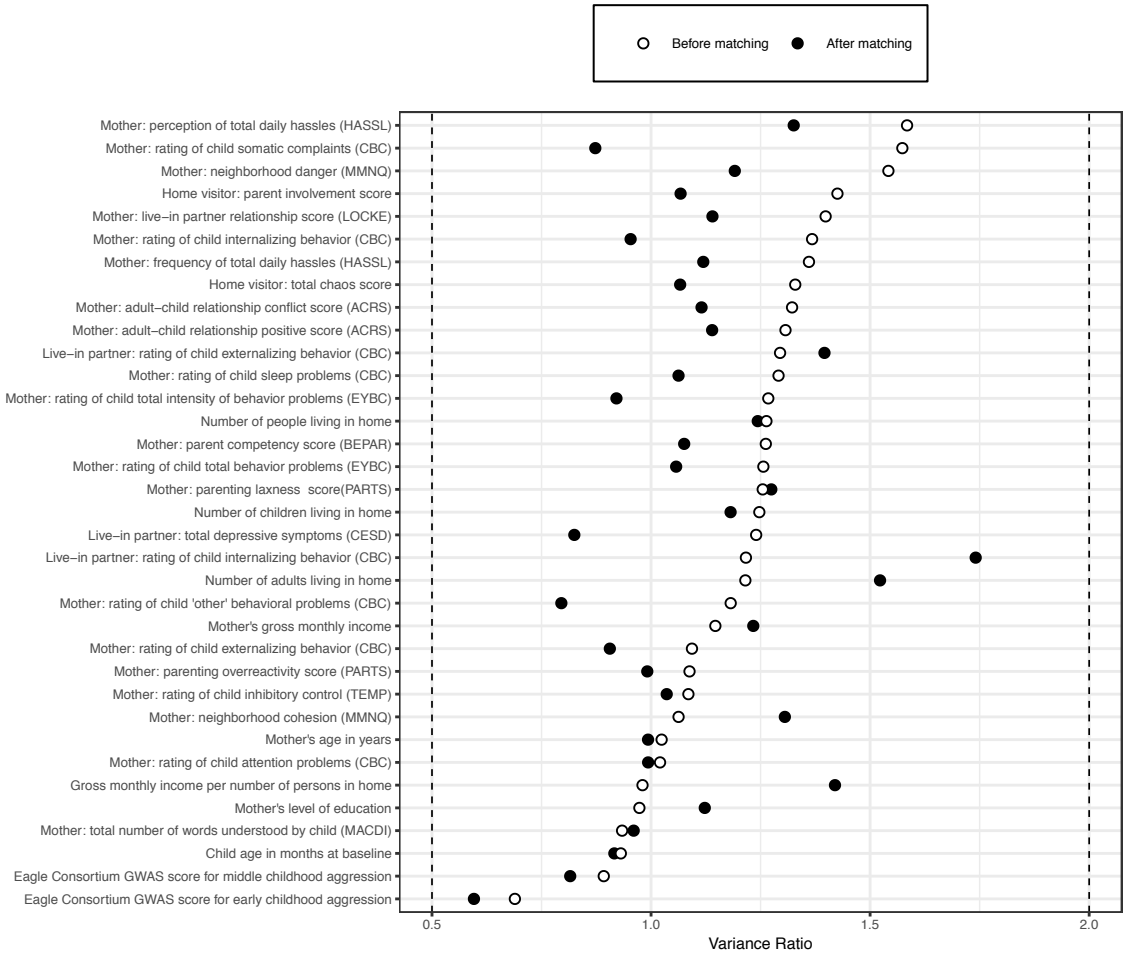
Figure 4
Balance on SMDs of Metric Covariates After Matching



Note. Standardized mean difference calculated as [mean(depressed) – mean(non-depressed)] / sd(non-depressed). Variables are sorted from highest to lowest SMD prior to matching. Open circles to the left or right of the band indicated by vertical dashed lines indicate variables with substantial (|smd| > 0.20) imbalance between depressed and non-depressed groups *prior to matching*. Filled circles are all within the band indicated by vertical dashed lines, confirming that depressed and non-depressed groups have similar means on these covariates *after matching*.

Figure 5

Balance on Variance Ratios of Metric Covariates After Matching

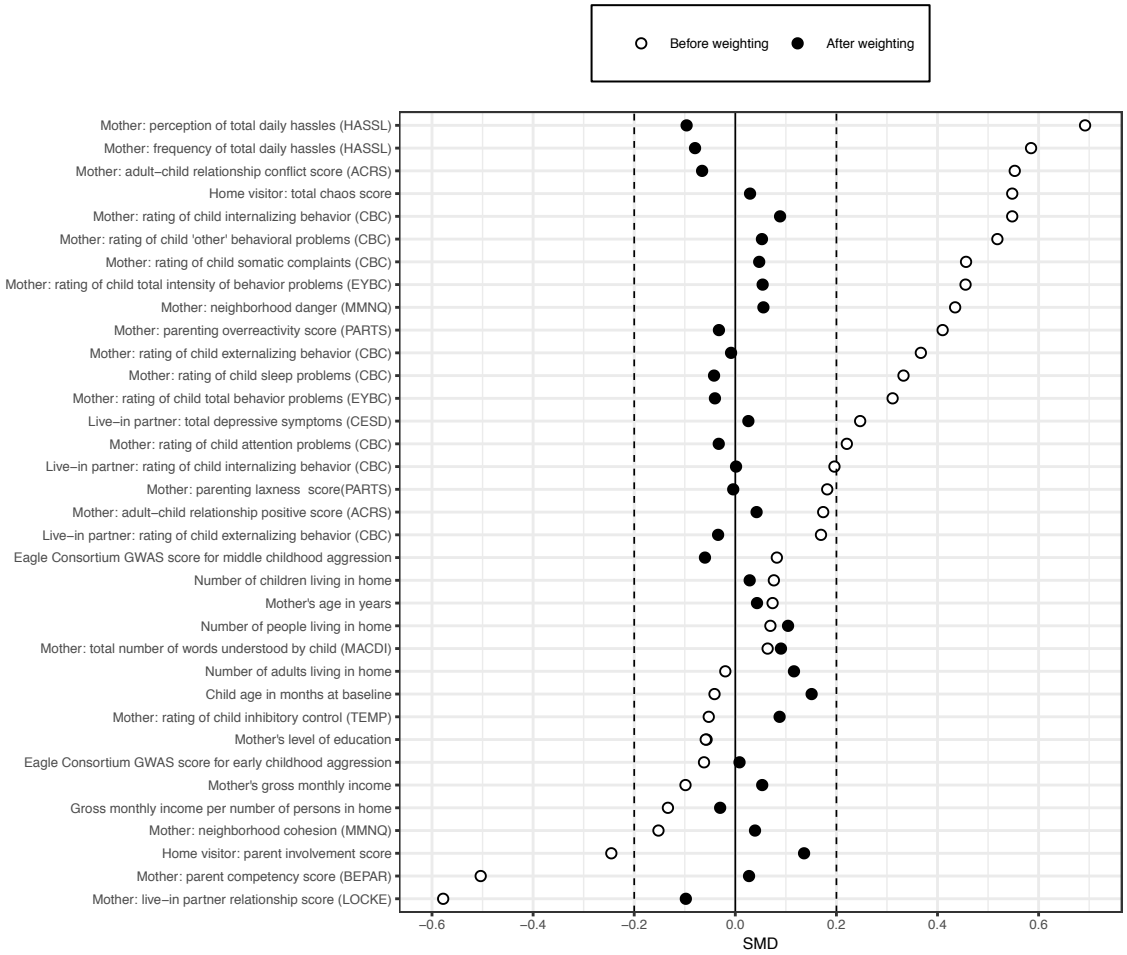


Note. Variance ratio calculated as var(depressed) / var(non-depressed). Variables are sorted from highest to lowest variance ratio prior to matching. All dots are within the band indicated by vertical dashed lines, indicating that the depressed and non-depressed groups have similar variances on these covariates both before and after matching.

Figure 6
Balance on Prevalence of Binary Covariates After Matching

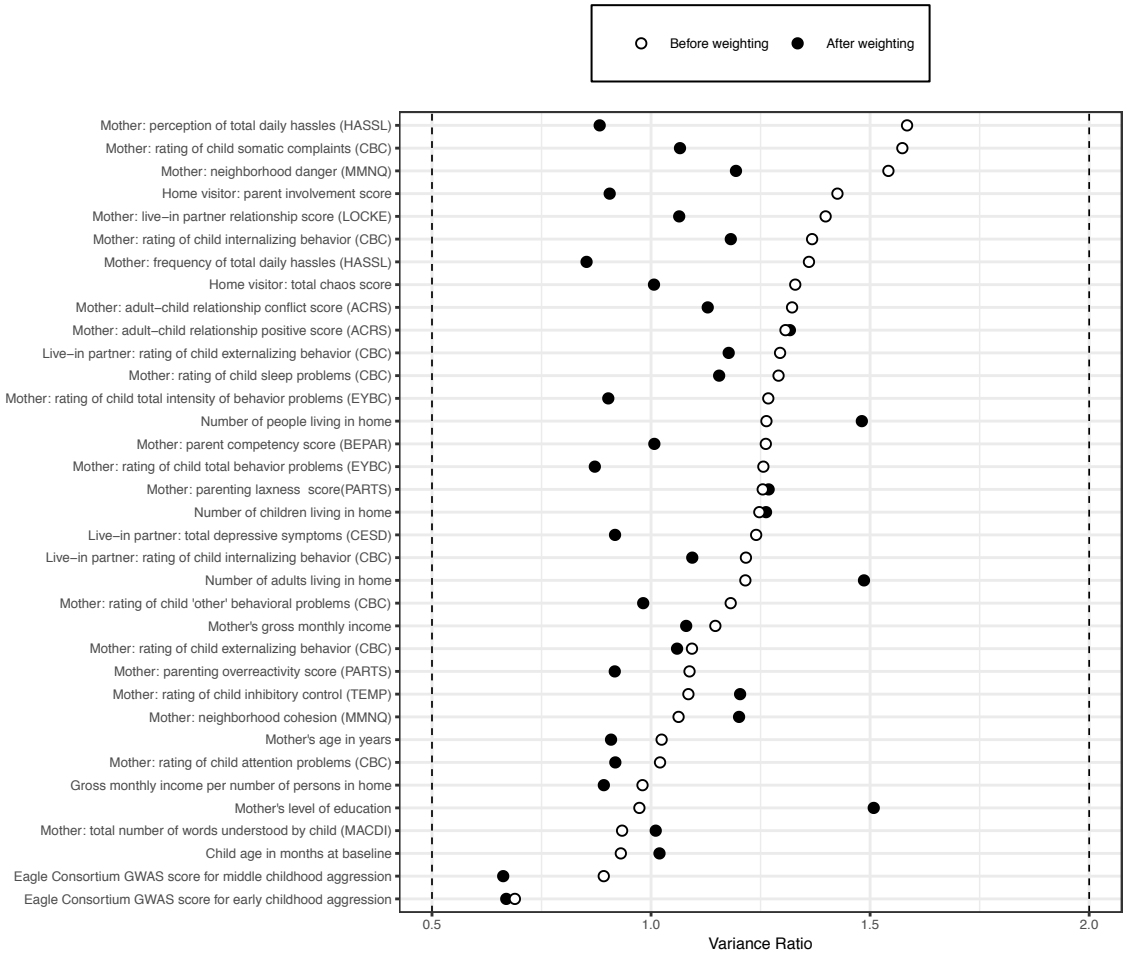
Note. Difference in prevalence calculated as prev(depressed) – prev(non-depressed). Variables are sorted from highest to lowest difference in prevalence prior to matching. Open circles to the left or right of the band indicated by vertical dashed lines indicate variables with more than a 5 percentage point difference in prevalence in the depressed and non-depressed groups prior to matching. Filled circles are almost all within the band indicated by vertical dashed lines, confirming that depressed and non-depressed groups have similar prevalences of these covariates after matching.

Figure 7
Balance on SMDs of Metric Covariates After Weighting



Note. Standardized mean difference calculated as [mean(depressed) – mean(non-depressed)] / sd(non-depressed). Variables are sorted from highest to lowest SMD prior to weighting. Open circles to the left or right of the band indicated by vertical dashed lines indicate variables with substantial (|smd| > 0.20) imbalance between depressed and non-depressed groups prior to weighting. Filled circles are all within the band indicated by vertical dashed lines, confirming that depressed and non-depressed groups have similar means on these covariates after weighting.

Figure 8
Balance on Variance Ratios of Metric Covariates After Weighting



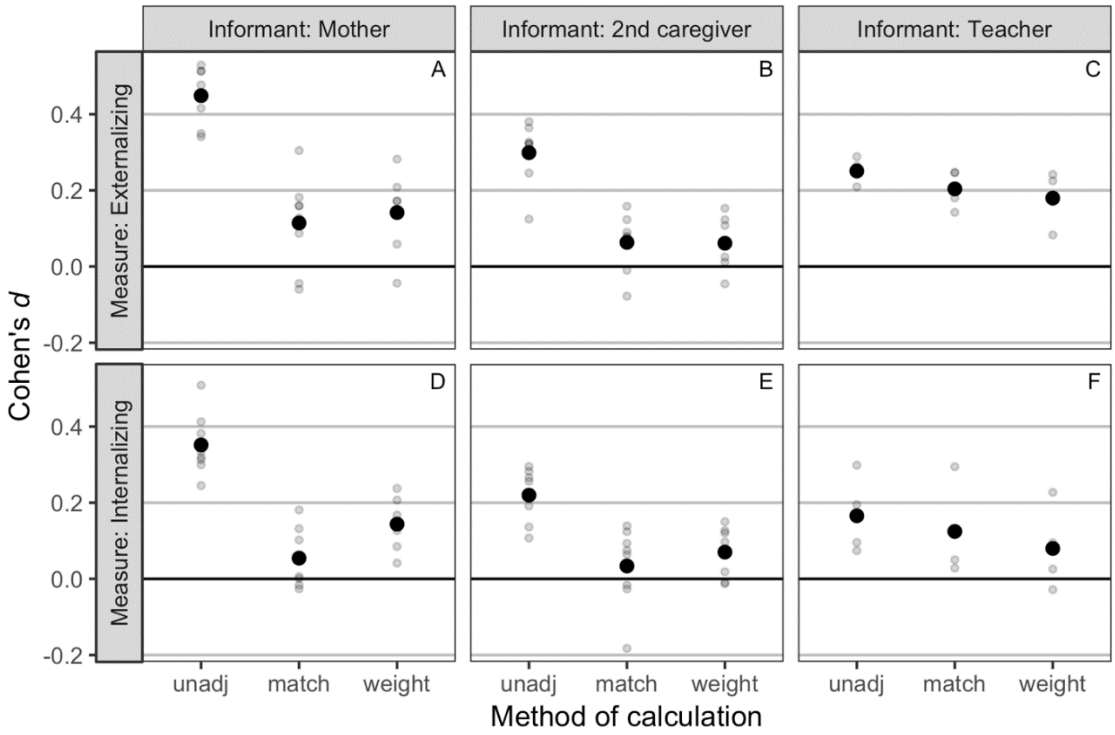
Note. Variance ratio calculated as var(depressed) / var(non-depressed). Variable are sorted from highest to lowest variance ratio prior to weighting. All dots are within the band indicated by vertical dashed lines, indicating that the depressed and non-depressed groups have similar variances on these covariates both before and after weighting.

Figure 9
Balance on Prevalence of Binary Covariates After Weighting

Note. Difference in prevalence calculated as prev(depressed) – prev(non-depressed). Variables are sorted from highest to lowest difference in prevalence prior to weighting. Open circles to the left or right of the band indicated by vertical dashed lines indicate variables with more than a 5 percentage point difference in prevalence in the depressed and non-depressed groups prior to weighting. Filled circles are almost all within the band indicated by vertical dashed lines, confirming that depressed and non-depressed groups have similar prevalences of these covariates after weighting.

Figure 10

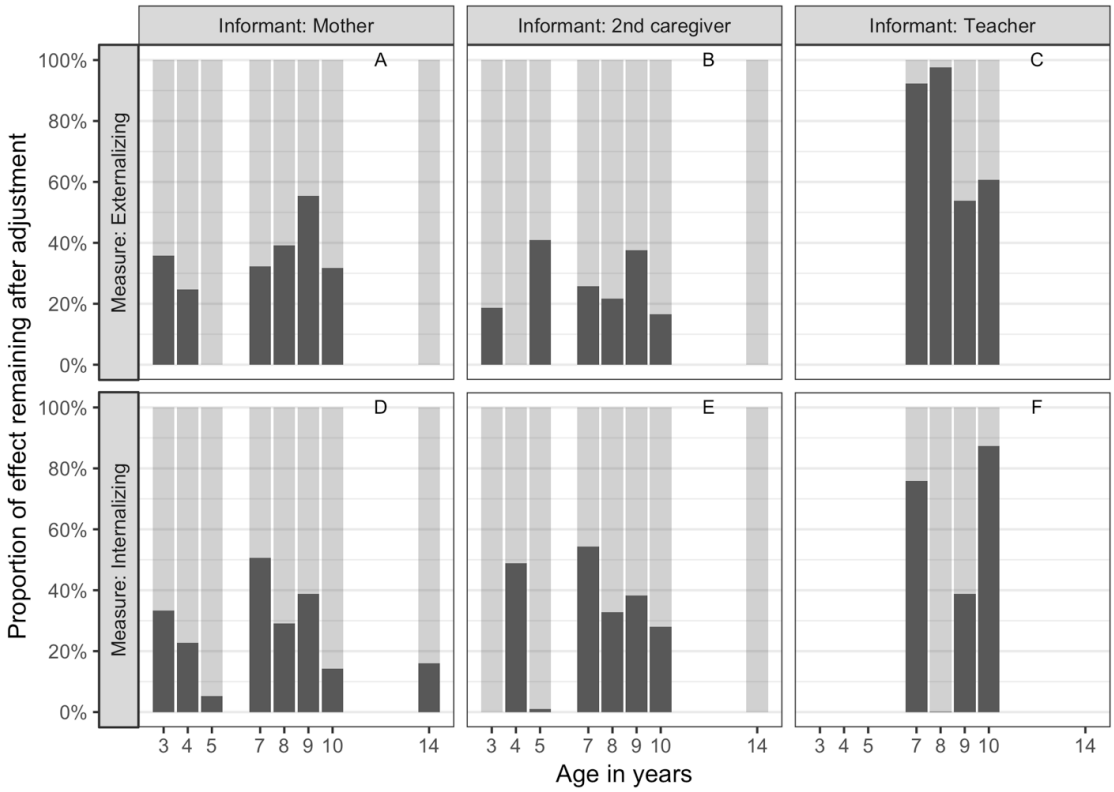
Mean Effects of Maternal Depression at Age 2 on Externalizing and Internalizing Behavior



Note. unadj = unadjusted *prima facie* estimate, match = matching estimate, weight = weighting estimate. Large black dots indicate mean effect size across age for each combination of outcome, informant, and method of calculation. Lighter, smaller black dots indicate individual effect sizes contributing to each mean (i.e., effects at one specific age).

Figure 11

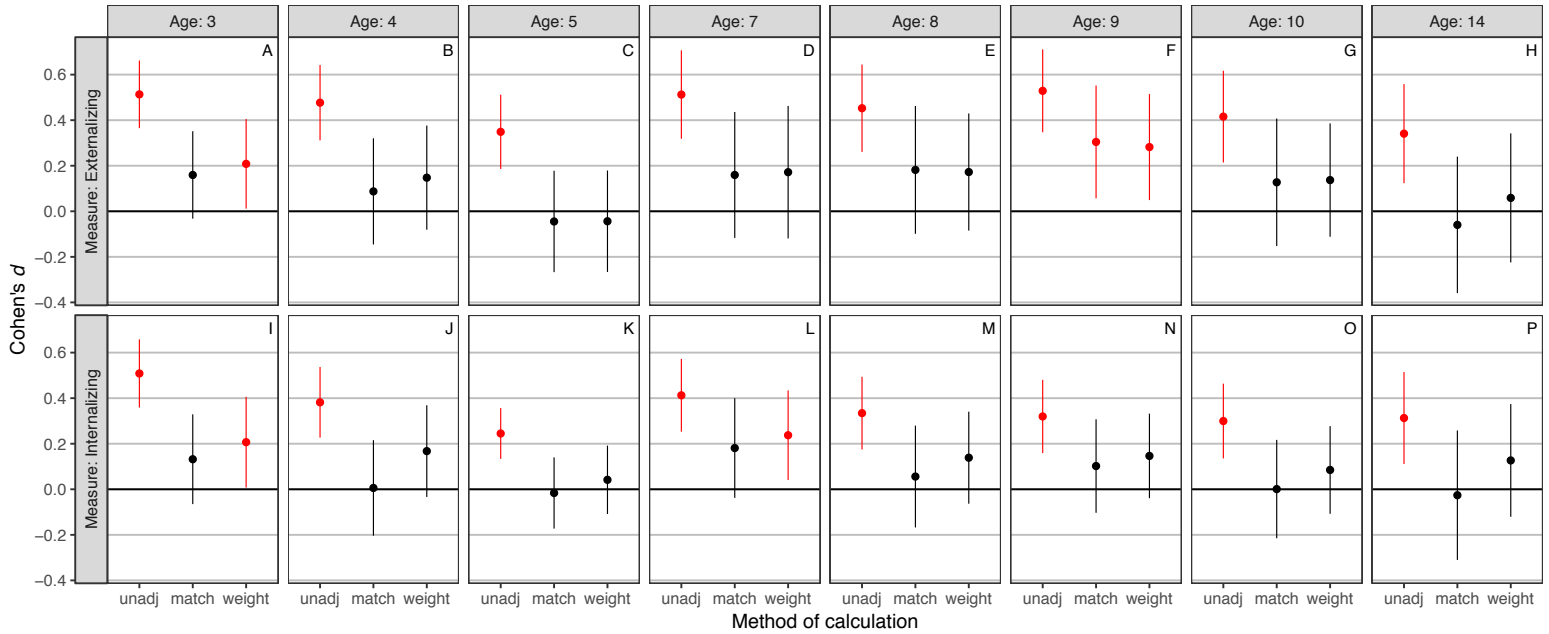
Proportion of Effect of Maternal Depression at Age 2 Remaining After Adjustment for Confounding



Note. For calculation of proportions, adjusted effect size was the mean of the matched and weighted estimates. Dark grey bars indicate the proportion of an unadjusted effect that remained after adjustment for confounding. Light grey bars indicate what theoretically would have occurred if confounding had no impact on the effect size (i.e., 100% of unadjusted effect would have remained). For example, a dark bar reaching 40% on the y-axis would indicate that 40% of the unadjusted effect remained after adjustment (e.g., Cohen's *d* was reduced from 0.50 to 0.20).

Figure 12

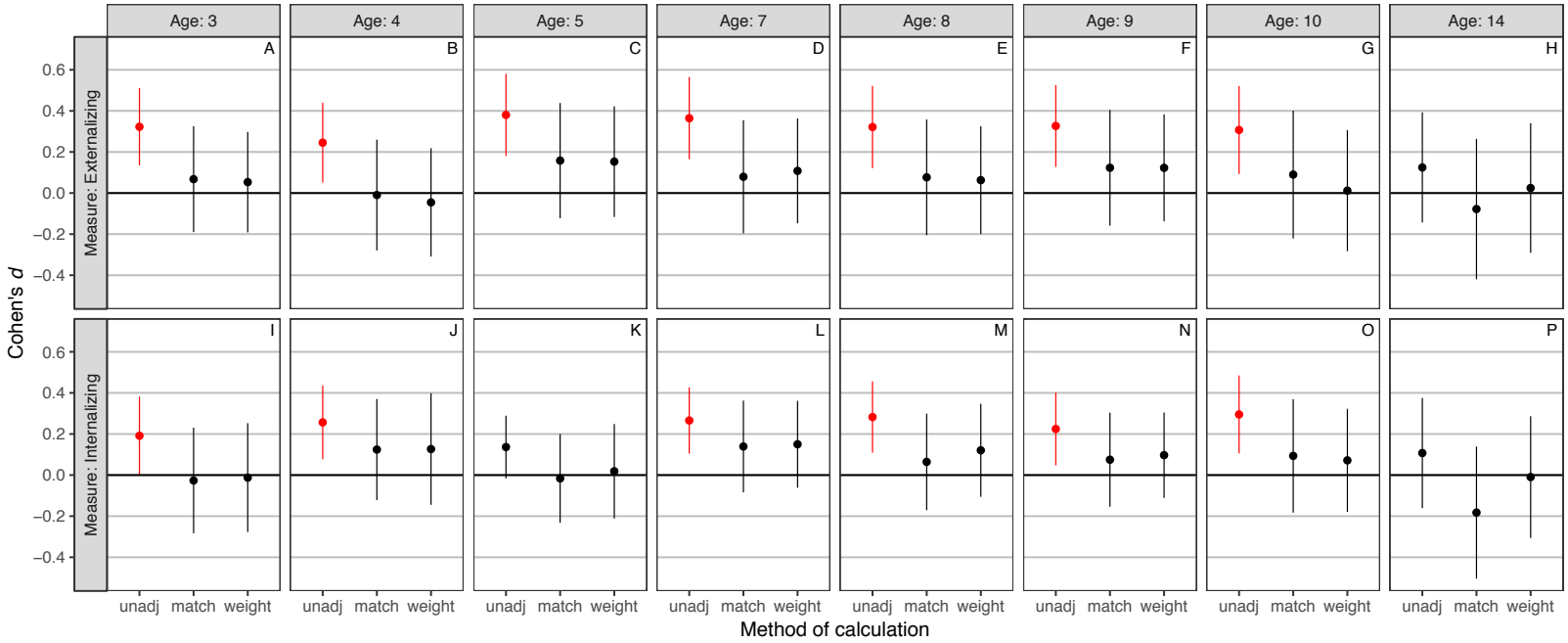
Mother-Reported Outcomes: Effects of Maternal Depression at Child Age 2



Note. unadj = unadjusted *prima facie* estimate, match = matching estimate, weight = weighting estimate.

Figure 13

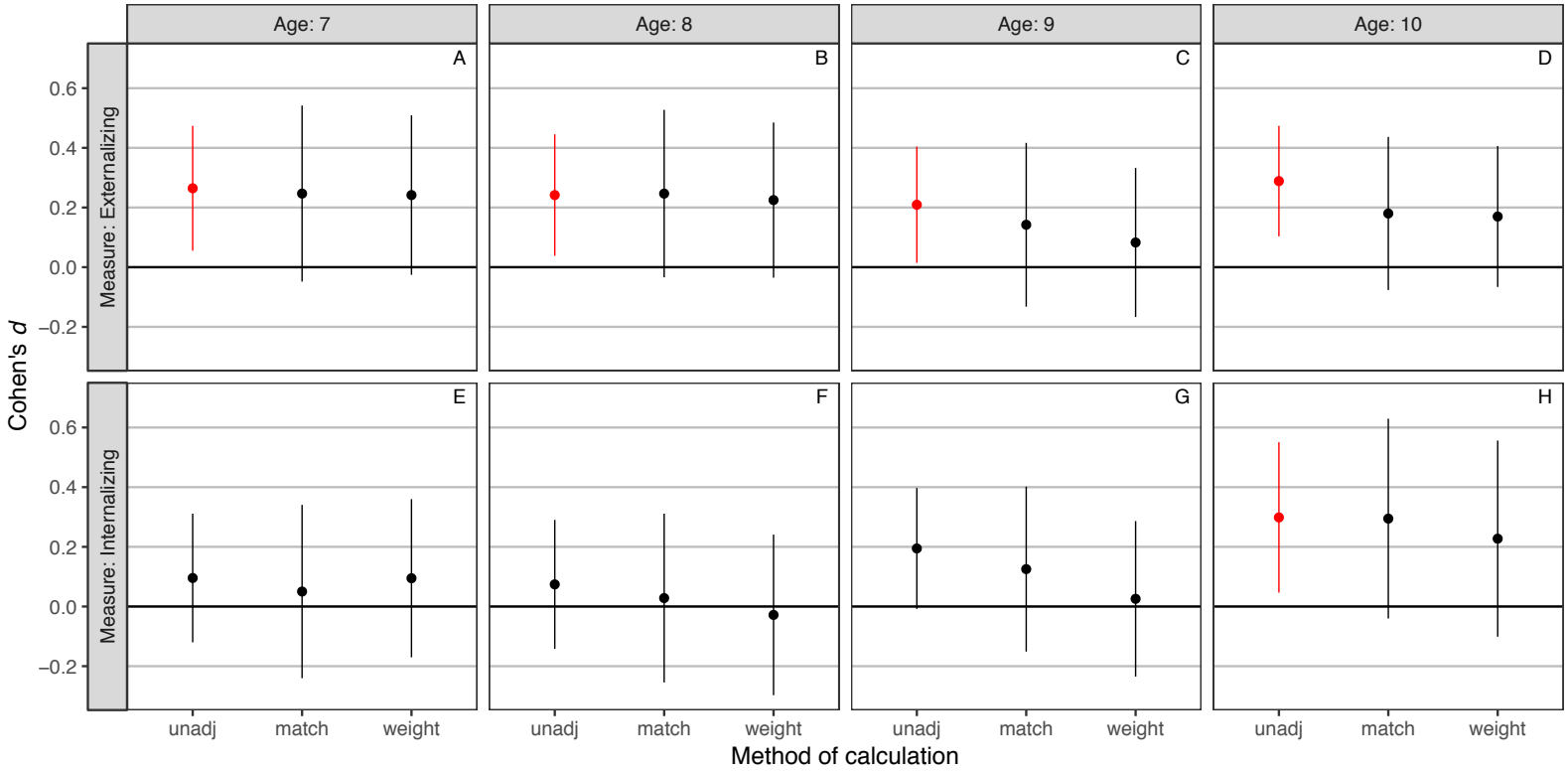
Secondary-Caregiver-Reported Outcomes: Effects of Maternal Depression at Child Age 2



Note. unadj = unadjusted *prima facie* estimate, match = matching estimate, weight = weighting estimate.

Figure 14

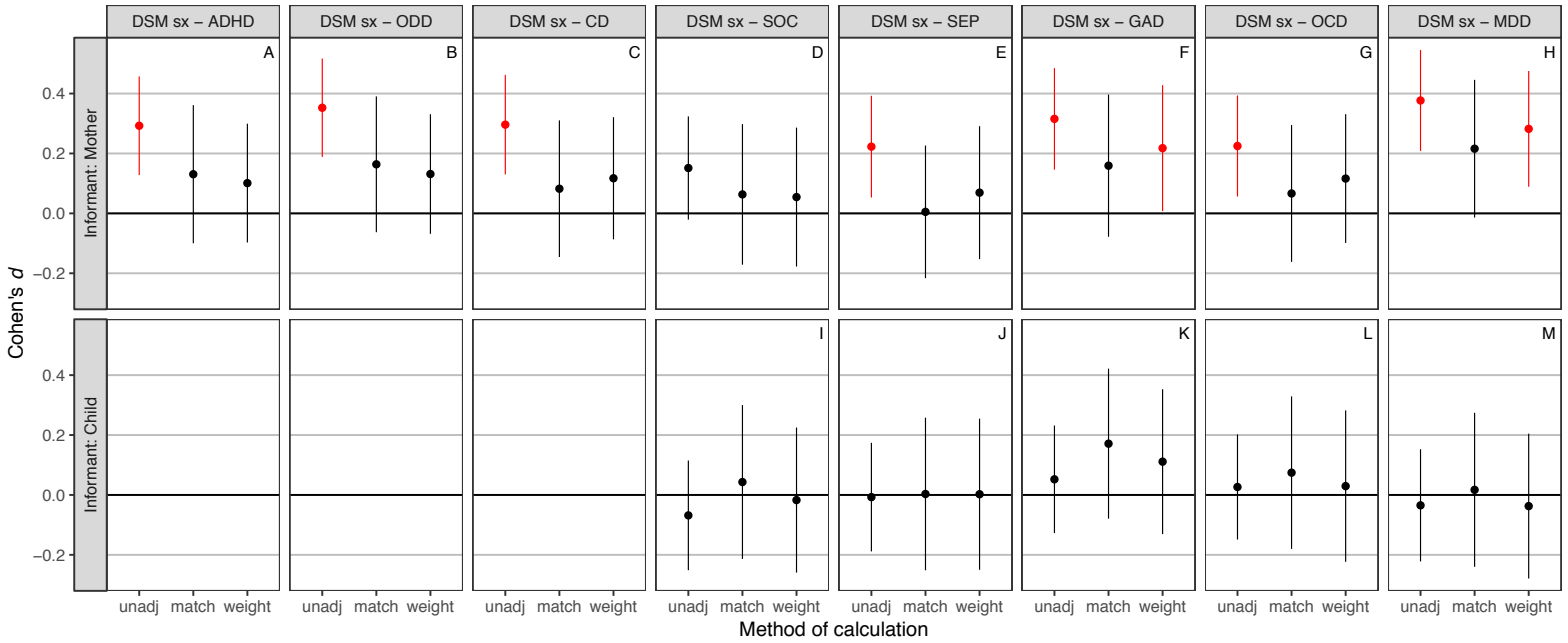
Teacher-Reported Outcomes: Effects of Maternal Depression at Child Age 2



Note. unadj = unadjusted *prima facie* estimate, match = matching estimate, weight = weighting estimate.

Figure 15

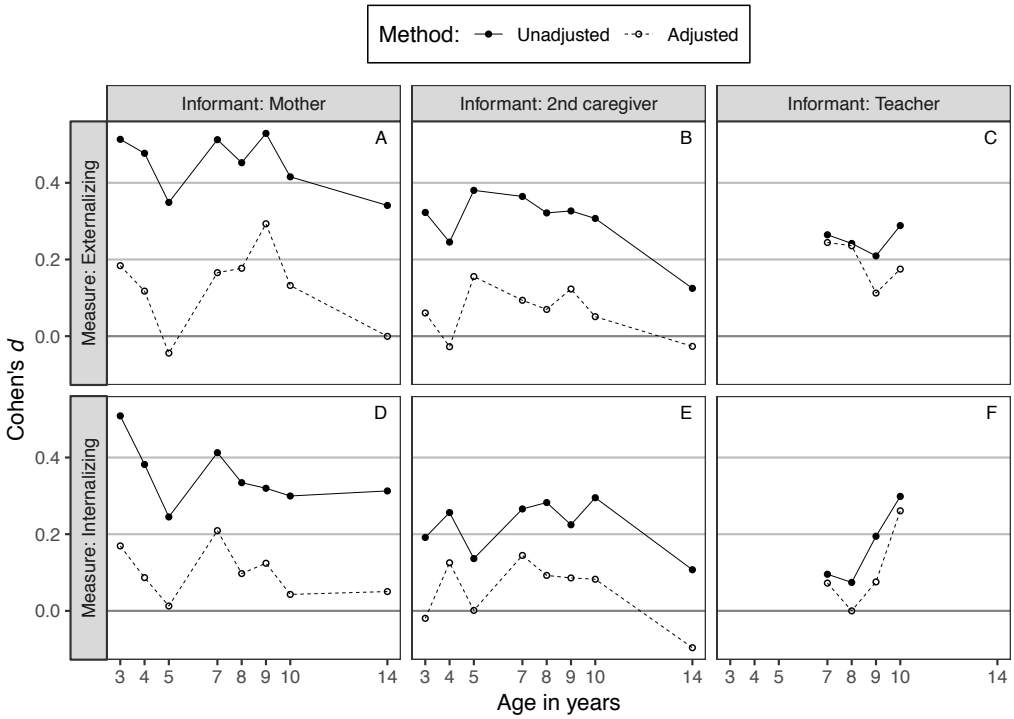
DSM Symptom Outcomes at Age 10: Effects of Maternal Depression at Child Age 2



Note. ADHD = attention deficit hyperactivity disorder, ODD = oppositional defiant disorder, CD = conduct disorder, SOC = social phobia disorder, SEP = separation anxiety disorder, GAD = generalizing anxiety disorder, OCD = obsessive-compulsive disorder, MDD = major depressive disorder, unadj = unadjusted estimate, match = matching estimate, weight = weighting estimate.

Figure 16

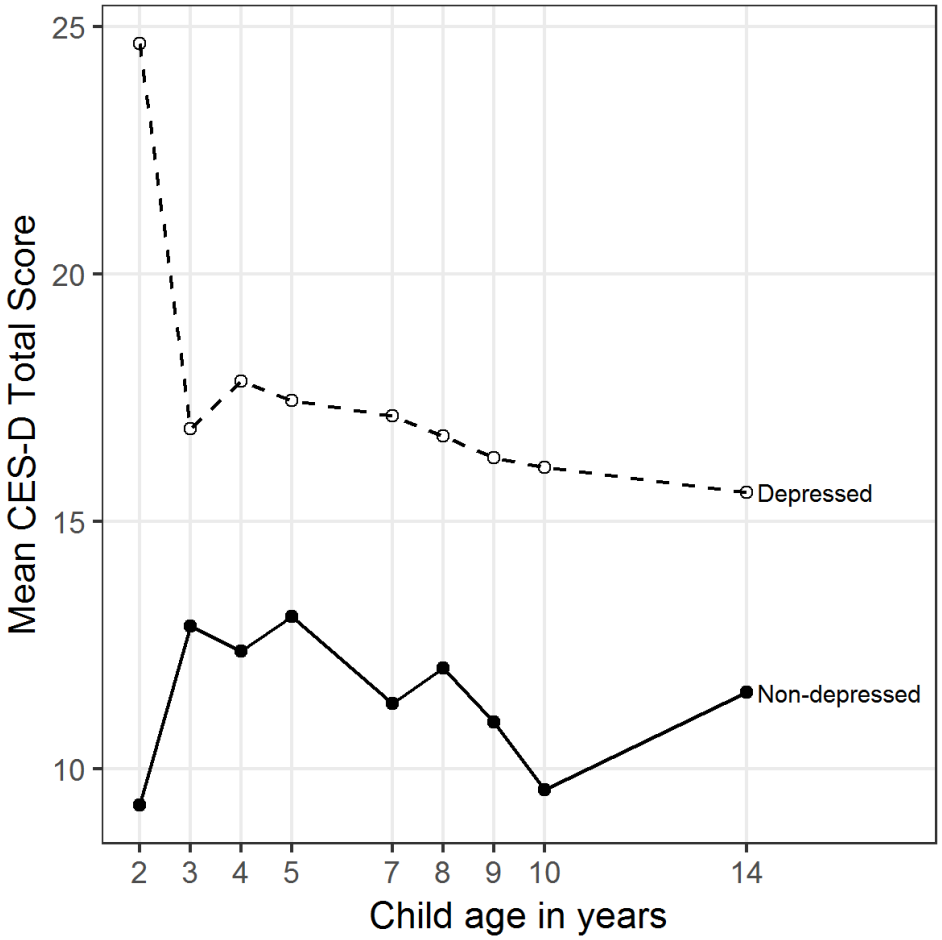
Unadjusted vs. Adjusted Effects of Maternal Depression at Age 2 Over Time



Note. Adjusted effect sizes were calculated as the mean of the matched and weighted estimates.

Figure 17

Longitudinal Mean Maternal Depression among the Matched Depressed and Non-Depressed Groups Defined at Child Age 2



Note. Compares mean CES-D total scores of the matched depressed (N = 179) and non-depressed (N = 179) groups defined at child age 2. At follow-up waves, standardized differences between depressed and non-depressed groups defined at child age 2 ranged from d = 0.38 (at child age 3) to d = 0.73 (at child age 10). Calculated using available data at each measurement. Change in the groups' mean CES-D from child age 2 to child age 3 is partially attributable to regression to the mean.

Figure 18

Example of Model in Which Adjustment for Variables Confounding the Relation of Maternal Depression with Child Behavior May Profoundly Affect Results

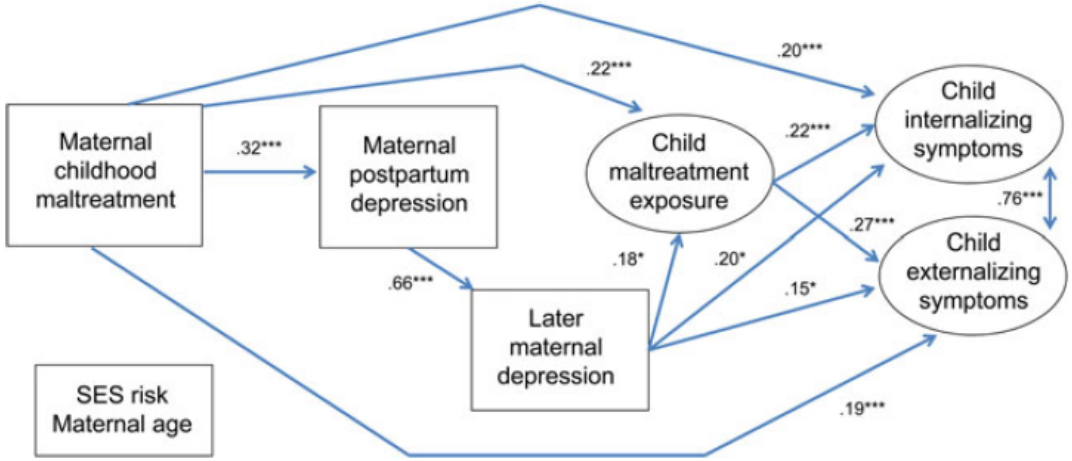


Figure 3. (Color online) Adjusted structural model including later maternal depression. Rectangles represent observed variables; ovals represent latent variables on which younger and elder twin scores have been regressed (paths not shown to improve readability). Solid lines represent paths with coefficients significant at $p < .05$* or $p < .001$***, with only significant paths shown in the model.

Note. Reproduced from Choi et al. (2018), in which it is Figure 3.

APPENDIX B

SUPPLEMENTARY MATERIAL

Study-Specific Scale Assessing Adolescents' Deviant Behavior (ASRD)

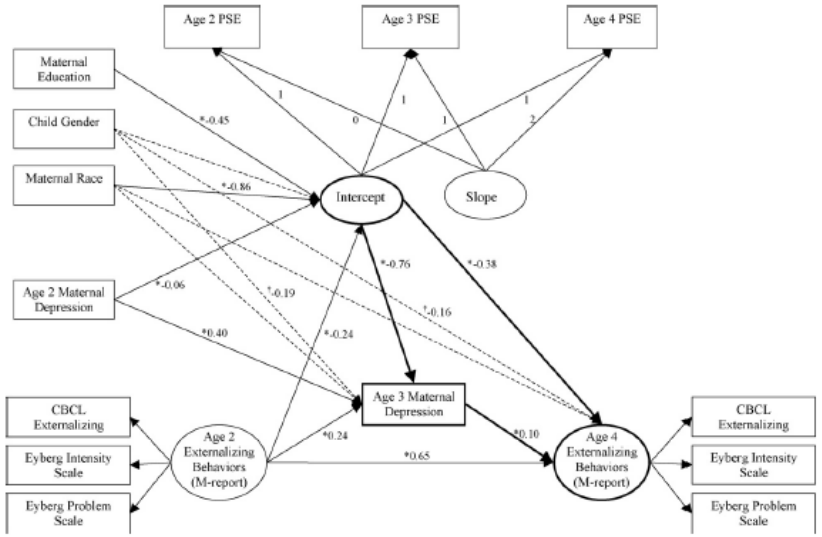
At age 14, adolescents completed a self-report questionnaire including 27 items measuring deviant behaviors. They were instructed to indicate how often they engaged in each behavior in the past year. Sample items include:

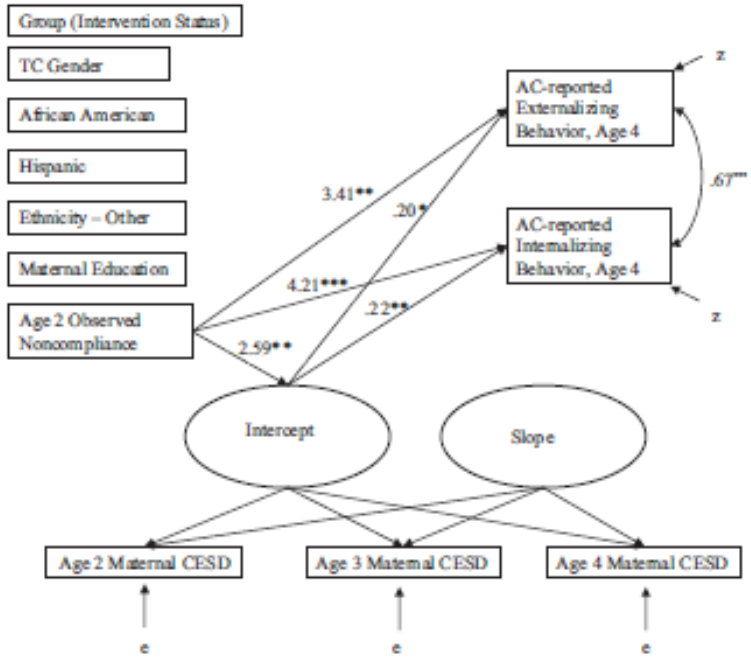
- Have you received an in-school detention?
- Have you skipped school without an excuse?
- Have you on purpose broken or damaged or destroyed something belonging to a school?
- Have you taken something from a store without paying for it?
- Have you bullied, threatened, or intimidated someone else?

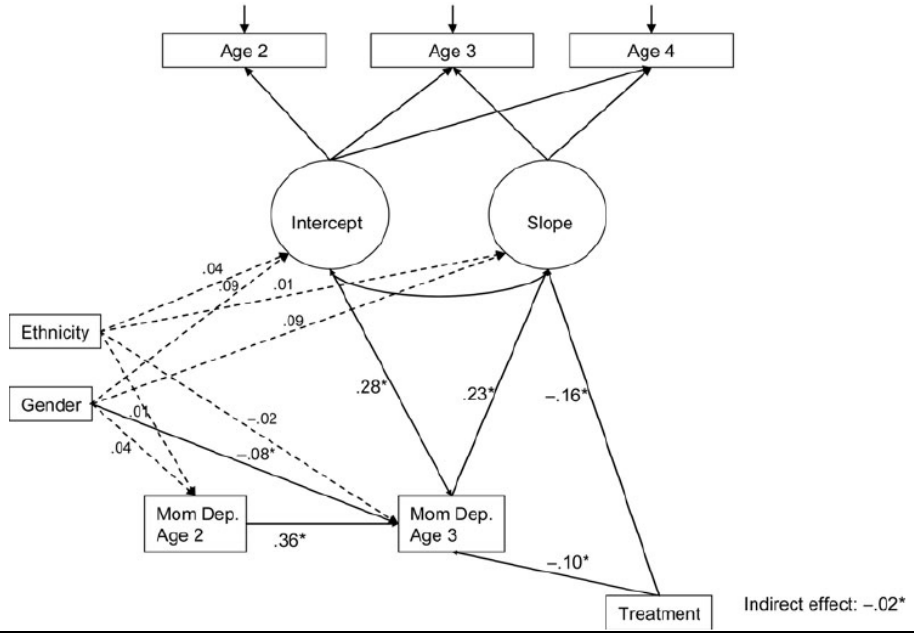
Response options included *never* (1)*, once or twice* (2)*,* or *more often* (3). Responses to these 27 items were summed to produce a score measuring adolescents' self-reported deviant behavior.

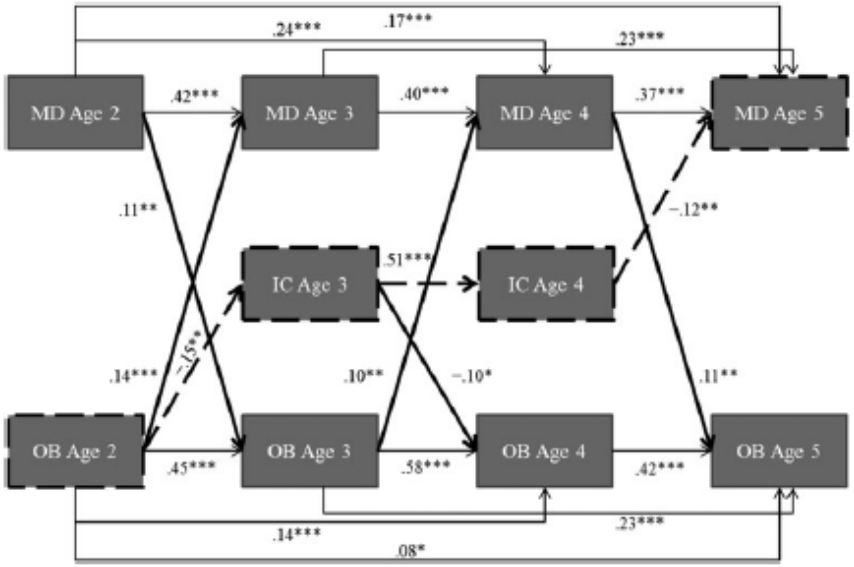
Table A1

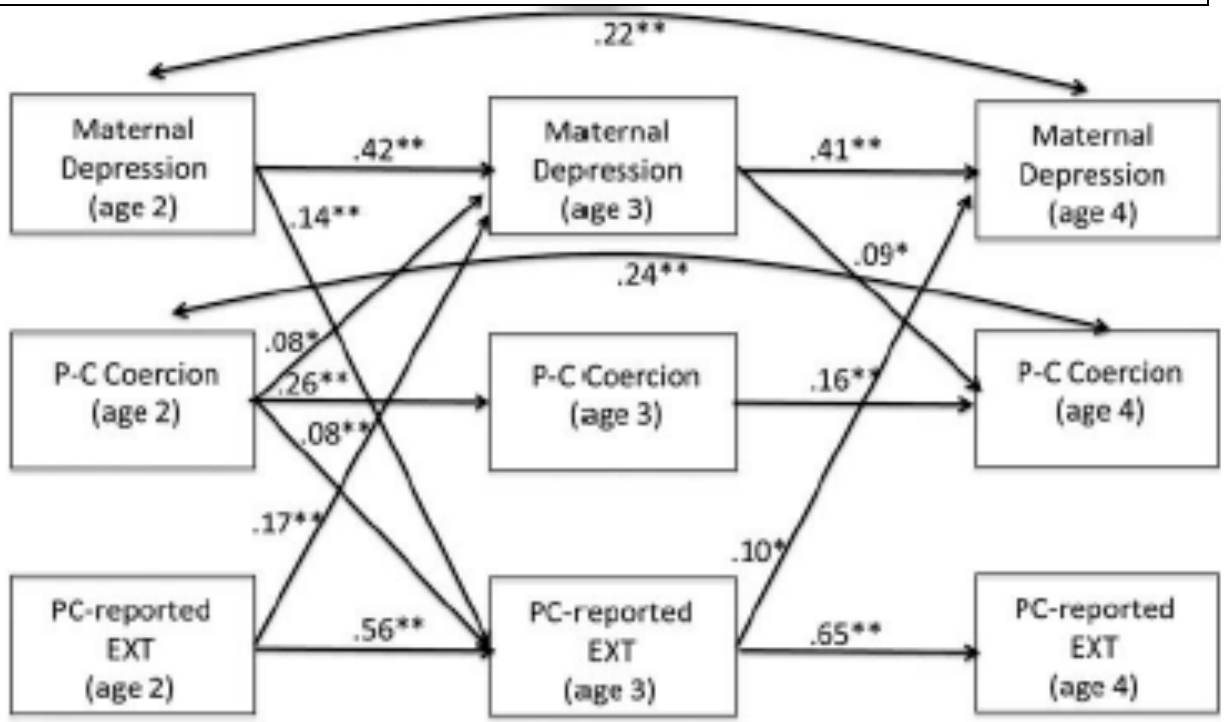
Models Fit in Previous Early Steps Studies

Study	Description	Path Model
Weaver et al. 2008	Weaver et al. (2008) investigated growth in parenting self-efficacy across ages 2 to 4, finding that maternal depression mediated the relationship between parent self-efficacy and subsequent child externalizing behavior. Increased self-efficacy was associated with reduced maternal depression, which was in turn associated with reduced externalizing behavior.	

Study	Description	Path Model
Gross et al. 2008	Gross et al. (2008) modeled the interactive influences of child noncompliance and parental depression on subsequent child internalizing and externalizing outcomes. Results indicated that directly-observed child noncompliance at child age 2 years old predicted higher levels of maternal and paternal depression, which in turn were associated with increased internalizing problems at age 4.	

Study	Description	Path Model
Shaw et al. 2009	Shaw et al. (2009) investigated maternal depression as a mediator of *intervention* effects, rather than of naturally occurring development processes.	

Study	Description	Path Model
Choe et al. 2014	Choe et al. (2014) examined what constructs might mediate the relationship between maternal depression and child's oppositional behavior, rather than considering maternal depression itself a mediator. Modeling revealed that maternal depression predicted subsequent levels of oppositional behavior even after controlling for prior levels of oppositional behavior, and that this effect was mediated by reductions in the child's inhibitory control.	

Study	Description	Path Model
Hails et al. 2017	Hails et al. (2017) modeled the transactional influences of maternal depression, parent-child coercion, and child behavior problems using repeated assessment of each construct at child ages 2, 3, and 4. A series of cross-lagged effects emerged, showing that depression, coercion, and child behavior predicted each other over time, with depression and coercion both associated with greater externalizing and internalizing problems.	

Note. Each path models is an exemplar from the associated manuscript, although in several, multiple models were fit. See following section of manuscript for discussion: Specific Findings on Effects of Parental Depression in the Early Steps Sample.

Table A2
Measures Collected at Baseline

Measure	Informant	Citation
Being a Parent Questionnaire (BEPAR)	Prim. Caregiver	(Johnston & Mash, 1989)
Confusion, Hubbub, and Order Scale	Prim. Caregiver	(Matheny, Wachs, Ludwig, & Phillips, 1995)
Parenting Daily Hassles Scale (HASSL)	Prim. Caregiver	(Crnic & Greenberg, 2008)
Demographics Questionnaire	Prim. Caregiver	*N/A* - developed for Early Steps study.
Financial Stress Questionnaire	Prim. Caregiver	(Conduct Problems Prevention Research Group, 1999)
General Life Satisfaction	Prim. Caregiver	(Crnic et al., 1983)
MacArthur Developmental Inventory (MACDI)	Prim. Caregiver	(Arriaga, Fenson, Cronan, & Pethick, 1998)
Me and My Neighborhood Questionnaire	Prim. Caregiver	(Ingoldsby & Shaw, 2002)
Multigroup Ethnic Identity Measure	Prim. Caregiver	(Phinney, 1992)
Service Provider Questionnaire	Prim. Caregiver	(Child and Family Center, 2001)
Adult-Child Relationship Scale (ACRS)	Prim. Caregiver, Alt. Caregiver	(Pianta, 1995)
Adult Substance Use - Short Version	Prim. Caregiver, Alt. Caregiver	(Dishion & Kavanagh, 2003)
Center for Epidemiologic Studies Depression Scale (CESD)	Prim. Caregiver, Alt. Caregiver	(Radloff, 1977)
Child Behavior Checklist 1.5-5 (CBC)	Prim. Caregiver, Alt. Caregiver	(Achenbach & Rescorla, 2001)
Eyberg Child Behavior Inventory (EYBC)	Prim. Caregiver, Alt. Caregiver	(Boggs, Eyberg, & Reynolds, 1990)
Family Assessment Task	Prim. Caregiver, Alt. Caregiver	(Dishion & Kavanagh, 1997)
Children's Behavior Questionnaire (TEMP)	Prim. Caregiver, Alt. Caregiver	(Rothbart, Ahadi, Hershey, & Fisher, 2003)
Parenting Scale (PARTS)	Prim. Caregiver, Alt. Caregiver	(Arnold, O'Leary, Wolff, & Acker, 1993)
Short Marital Adjustment Scale (LOCKE)	Prim. Caregiver, Alt. Caregiver	(Locke & Wallace, 1959)
Speech Sample	Prim. Caregiver, Alt. Caregiver	(Bullock & Dishion, 2004)
Ethnic Home Environment Inventory	Staff	(Caughy, Randolph, & O'Campo, 2002)
Infant/Toddler Home Inventory (IT-HOME)	Staff	(Caldwell & Bradley, 2003)

Note. Prim. = primary, Alt. = alternate. Acronyms in parentheses under "Measure" column show scale abbreviation that may be referenced in covariate labels in Tables 1, 2, and A3.

Table A3

Descriptive Statistics for Covariates Included in Propensity Score Model

Variable	N	Distinct	Mean	SD	Min	p25	Median	p75	Max
Binary variables									
Family assigned to intervention condition	707	2	50 %	-	-	-	-	-	-
Site is Charlottesville, VA	707	2	26 %	-	-	-	-	-	-
Site is Pittsburgh, PA	707	2	37 %	-	-	-	-	-	-
Child is male	707	2	50 %	-	-	-	-	-	-
Child is hispanic	705	2	14 %	-	-	-	-	-	-
Child is black	706	2	28 %	-	-	-	-	-	-
Child is biracial	706	2	13 %	-	-	-	-	-	-
Mother was teen parent	707	2	23 %	-	-	-	-	-	-
Mother is hispanic	705	2	11 %	-	-	-	-	-	-
Mother is black	706	2	28 %	-	-	-	-	-	-
Mother is married or living with significant other	707	2	57 %	-	-	-	-	-	-
Mother reports having a live-in partner	707	2	60 %	-	-	-	-	-	-
Mother endorses having religious / spiritual beliefs	707	2	69 %	-	-	-	-	-	-
Child has been cared for by person other than mother more than 5 hrs/wk	707	2	80 %	-	-	-	-	-	-
Family is below poverty line	699	2	75 %	-	-	-	-	-	-
Family received food stamps	707	2	62 %	-	-	-	-	-	-
Family received medical assistance	707	2	69 %	-	-	-	-	-	-
Family received social security income	707	2	10 %	-	-	-	-	-	-
Family receives child support	707	2	18 %	-	-	-	-	-	-
Family owns a home	705	2	16 %	-	-	-	-	-	-
Person in home had trouble with law since child's birth	701	2	35 %	-	-	-	-	-	-
Person in home reported for child abuse since child's birth	703	2	7 %	-	-	-	-	-	-
Person in home treated by a mental health professional since child's birth	707	2	38 %	-	-	-	-	-	-
Mother endorses support from extended family as family strength	707	2	66 %	-	-	-	-	-	-
Mother endorses employment situation as family strength	707	2	21 %	-	-	-	-	-	-
Mother endorses church, religion, spirituality as family strength	707	2	31 %	-	-	-	-	-	-
Mother endorses conflict or violence as impacting family	707	2	15 %	-	-	-	-	-	-
Mother endorses drug use by parent as impacting family	707	2	9 %	-	-	-	-	-	-
Mother endorses high crime neighborhood as impacting family	707	2	13 %	-	-	-	-	-	-
Mother endorses parent being absent as impacting family	707	2	25 %	-	-	-	-	-	-
Mother endorses stress between home and school as impacting family	707	2	12 %	-	-	-	-	-	-
Mother endorses unstable home situation as impacting family	707	2	7 %	-	-	-	-	-	-
Mother endorses death in family as impacting family	707	2	15 %	-	-	-	-	-	-
Mother endorses past traumatic experience as impacting family	707	2	12 %	-	-	-	-	-	-
Mother: no organized groups are a source of support	706	2	53 %	-	-	-	-	-	-
Mother: has not visited friends within past week	705	2	28 %	-	-	-	-	-	-
Mother: how many times per week speak with friends and family on phone	706	2	45 %	-	-	-	-	-	-
Mother: ever drinks alcohol	692	2	72 %	-	-	-	-	-	-
Mother: drinks alcohol at least monthly	692	2	32 %	-	-	-	-	-	-
Mother: drinks alcohol at least weekly	692	2	8 %	-	-	-	-	-	-
Mother: ever stopped drinking due to problems with use	682	2	6 %	-	-	-	-	-	-
Mother: ever uses marijuana	689	2	14 %	-	-	-	-	-	-
Mother: uses marijuana at least monthly	689	2	6 %	-	-	-	-	-	-
Mother: ever stopped using marijuana due to problems with use	681	2	6 %	-	-	-	-	-	-
Mother: currently smoke cigarettes	602	2	43 %	-	-	-	-	-	-
Mother: any hard drug use	692	2	7 %	-	-	-	-	-	-
Live-in partner: ever drinks alcohol	404	2	75 %	-	-	-	-	-	-

Variable	N	Distinct	Mean	SD	Min	p25	Median	p75	Max
Live-in partner: drinks alcohol at least monthly	404	2	41 %	-	-	-	-	-	-
Live-in partner: drinks alcohol at least weekly	404	2	14 %	-	-	-	-	-	-
Live-in partner: ever stopped drinking due to problems with use	393	2	12 %	-	-	-	-	-	-
Live-in partner: ever uses marijuana	407	2	16 %	-	-	-	-	-	-
Live-in partner: uses marijuana at least monthly	407	2	7 %	-	-	-	-	-	-
Live-in partner: ever stopped using marijuana due to problems with use	392	2	7 %	-	-	-	-	-	-
Live-in partner: currently smoke cigarettes	349	2	50 %	-	-	-	-	-	-
Metric variables									
Child age in months at baseline	707	17	28.17	3.29	18	25	28	31	35
Mother's age in years	701	30	26.52	5.87	16	22	25	30	46
Mother's level of education	707	7	5.17	1.13	2	5	5	6	8
Mother's gross monthly income	699	11	3.78	1.93	1	2	4	5	11
Number of people living in home	707	11	4.48	1.61	2	3	4	5	12
Number of children living in home	707	8	2.42	1.20	1	2	2	3	8
Number of adults living in home	707	8	2.06	0.94	1	1	2	2	8
Gross monthly income per number of persons in home	699	48	0.92	0.52	0.11	0.50	0.83	1.25	4.50
Mother: adult-child relationship positive score (ACRS)	706	16	8.09	2.98	5	6	7	10	20
Mother: adult-child relationship conflict score (ACRS)	706	43	28.37	7.45	11	23	28	33	49
Mother: live-in partner relationship score (LOCKE)	398	49	57.14	10.13	17	52	59	64	74
Mother: neighborhood cohesion (MMNQ)	704	33	14.86	7.61	5	8	14	20	35
Mother: neighborhood danger (MMNQ)	703	46	8.72	7.81	0	2	7	13	39
Mother: rating of child attention problems (CBC)	706	11	4.12	1.90	0	3	4	5	10
Mother: rating of child 'other' behavioral problems (CBC)	706	34	14.26	5.86	1	10	14	18	41
Mother: rating of child sleep problems (CBC)	706	14	4.64	2.78	0	3	4	6	13
Mother: rating of child somatic complaints (CBC)	706	12	2.26	1.96	0	1	2	3	11
Mother: rating of child externalizing behavior (CBC)	706	42	20.72	7.26	1	16	20	25	46
Mother: rating of child internalizing behavior (CBC)	706	35	12.47	6.42	0	8	12	16	38
Mother: rating of child inhibitory control (TEMP)	697	172	3.96	0.81	1.33	3.46	4.00	4.46	7.00
Mother: frequency of total daily hassles (HASSL)	707	61	46.73	8.60	25	41	46	52	80
Mother: parent competency score (BEPAR)	705	84	62.79	11.82	21.00	55.00	63.00	71.00	92.00
Mother: parenting laxness score(PARTS)	696	66	3.13	0.98	1.00	2.45	3.09	3.73	6.45
Mother: parenting overreactivity score (PARTS)	696	42	2.68	0.77	1	2	3	3	5
Mother: perception of total daily hassles (HASSL)	707	93	47.32	13.10	20	38	46	56	93
Mother: rating of child total behavior problems (EYBC)	705	34	14.21	6.47	0	10	14	19	33
Mother: rating of child total intensity of behavior problems (EYBC)	707	130	128.64	27.98	53	108	127	146	237
Mother: total number of words understood by child (MACDI)	704	98	60.14	25.45	1	41	63	80	100
Live-in partner: total depressive symptoms (CESD)	407	53	11.12	8.64	0	5	9	15	49
Live-in partner: rating of child externalizing behavior (CBC)	409	39	15.57	7.85	0	10	15	21	40
Live-in partner: rating of child internalizing behavior (CBC)	409	38	10.52	7.78	0	5	9	14	45
Home visitor: parent involvement score	706	4	2.21	0.89	0	2	2	3	3
Home visitor: total chaos score	703	16	5.33	3.60	0	2	5	8	15
Eagle Consortium GWAS score for early childhood aggression	498	498	2.19	3.85	-11.13	-0.25	2.46	4.70	13.00
Eagle Consortium GWAS score for middle childhood aggression	498	496	-27.34	6.35	-44.74	-31.97	-27.99	-22.51	-12.08

Note. "distinct" = number of distinct non-missing values, p25 = 25% percentile, p75 = 75% percentile. Binary variables are always coded as 0 = no, 1 = yes, such that means reflect proportions of sample meeting criterion.

Table A4
Descriptive Statistics for Outcome Variables

Informant	Scale	Measure	Age	N	Proportion Missing	Distinct	Mean	SD	Min	p25	Median	p75	Max
Mother	CBCL	Externalizing	3	638	10 %	42	0.74	0.33	0.00	0.50	0.71	0.96	1.75
Mother	CBCL	Externalizing	4	609	14 %	43	0.66	0.36	0.00	0.38	0.62	0.88	1.83
Mother	CBCL	Externalizing	5	594	16 %	41	0.33	0.24	0.00	0.14	0.29	0.46	1.23
Mother	CBCL	Externalizing	7	550	22 %	46	0.37	0.28	0.00	0.17	0.31	0.51	1.37
Mother	CBCL	Externalizing	8	535	24 %	44	0.31	0.27	0.00	0.11	0.23	0.46	1.43
Mother	CBCL	Externalizing	9	566	20 %	41	0.31	0.26	0.00	0.11	0.26	0.43	1.46
Mother	CBCL	Externalizing	10	544	23 %	44	0.30	0.28	0.00	0.09	0.23	0.40	1.43
Mother	CBCL	Externalizing	14	529	25 %	47	0.31	0.30	0.00	0.09	0.23	0.46	1.49
Mother	CBCL	Internalizing	3	638	10 %	36	0.32	0.20	0.00	0.17	0.28	0.42	1.39
Mother	CBCL	Internalizing	4	609	14 %	37	0.30	0.20	0.00	0.14	0.28	0.39	1.22
Mother	CBCL	Internalizing	5	594	16 %	29	0.19	0.16	0.00	0.06	0.16	0.25	0.97
Mother	CBCL	Internalizing	7	550	22 %	37	0.26	0.22	0.00	0.09	0.19	0.34	1.31
Mother	CBCL	Internalizing	8	535	24 %	34	0.24	0.22	0.00	0.09	0.19	0.34	1.22
Mother	CBCL	Internalizing	9	566	20 %	35	0.24	0.22	0.00	0.09	0.19	0.34	1.56
Mother	CBCL	Internalizing	10	543	23 %	33	0.24	0.23	0.00	0.06	0.19	0.38	1.34
Mother	CBCL	Internalizing	14	532	25 %	40	0.30	0.27	0.00	0.09	0.22	0.44	1.50
2nd caregiver	CBCL	Externalizing	3	403	43 %	37	0.58	0.33	0.00	0.33	0.58	0.79	1.58
2nd caregiver	CBCL	Externalizing	4	381	46 %	39	0.53	0.33	0.00	0.29	0.50	0.75	1.71
2nd caregiver	CBCL	Externalizing	5	331	53 %	35	0.26	0.22	0.00	0.10	0.23	0.37	1.37
2nd caregiver	CBCL	Externalizing	7	414	41 %	42	0.29	0.25	0.00	0.11	0.23	0.40	1.46
2nd caregiver	CBCL	Externalizing	8	399	44 %	38	0.28	0.24	0.00	0.10	0.23	0.40	1.31
2nd caregiver	CBCL	Externalizing	9	420	41 %	40	0.27	0.24	0.00	0.09	0.20	0.37	1.31
2nd caregiver	CBCL	Externalizing	10	385	46 %	38	0.24	0.24	0.00	0.06	0.17	0.34	1.43
2nd caregiver	CBCL	Externalizing	14	290	59 %	40	0.27	0.26	0.00	0.09	0.20	0.40	1.43
2nd caregiver	CBCL	Internalizing	3	403	43 %	33	0.28	0.20	0.00	0.14	0.22	0.38	1.03
2nd caregiver	CBCL	Internalizing	4	381	46 %	33	0.25	0.18	0.00	0.11	0.22	0.33	1.00
2nd caregiver	CBCL	Internalizing	5	331	53 %	22	0.16	0.16	0.00	0.06	0.12	0.22	1.22
2nd caregiver	CBCL	Internalizing	7	414	41 %	30	0.21	0.19	0.00	0.06	0.16	0.28	1.34
2nd caregiver	CBCL	Internalizing	8	399	44 %	30	0.21	0.20	0.00	0.06	0.16	0.28	1.44
2nd caregiver	CBCL	Internalizing	9	419	41 %	30	0.20	0.20	0.00	0.06	0.16	0.28	1.19
2nd caregiver	CBCL	Internalizing	10	385	46 %	33	0.23	0.22	0.00	0.06	0.16	0.31	1.25
2nd caregiver	CBCL	Internalizing	14	290	59 %	31	0.24	0.24	0.00	0.09	0.16	0.31	1.31
Teacher	TRF	Externalizing	7	304	57 %	37	0.25	0.32	0.00	0.00	0.09	0.35	1.56
Teacher	TRF	Externalizing	8	368	48 %	42	0.25	0.33	0.00	0.00	0.12	0.34	1.53
Teacher	TRF	Externalizing	9	370	48 %	42	0.24	0.31	0.00	0.00	0.09	0.34	1.59
Teacher	TRF	Externalizing	10	347	51 %	41	0.23	0.31	0.00	0.00	0.09	0.34	1.41
Teacher	TRF	Internalizing	7	285	60 %	32	0.21	0.22	0.00	0.06	0.15	0.30	1.21
Teacher	TRF	Internalizing	8	312	56 %	30	0.22	0.21	0.00	0.06	0.15	0.31	1.15
Teacher	TRF	Internalizing	9	319	55 %	32	0.23	0.22	0.00	0.06	0.18	0.33	1.42
Teacher	TRF	Internalizing	10	277	61 %	32	0.22	0.23	0.00	0.06	0.15	0.33	1.24
Child	ASRD	Deviancy	14	534	24 %	24	3.64	4.61	0	1	2	5	40

Informant	Scale	Measure	Age	N	Proportion Missing	Distinct	Mean	SD	Min	p25	Median	p75	Max
Child	MASC	Anxiety	14	530	25 %	42	1.15	0.57	0.00	0.70	1.10	1.60	2.80
Child	CDI	Depression	14	530	25 %	39	0.27	0.36	0.00	0.00	0.10	0.40	2.00
Mother	DBD-RS	DSM sx - ADHD	10	544	23 %	95	0.88	0.60	0.00	0.39	0.72	1.23	2.89
Mother	DBD-RS	DSM sx - ODD	10	543	23 %	28	0.84	0.64	0.00	0.38	0.75	1.25	3.00
Mother	DBD-RS	DSM sx - CD	10	544	23 %	26	0.14	0.18	0.00	0.00	0.07	0.20	1.13
Mother	DISC	DSM sx - SOC	10	520	26 %	13	2.53	3.22	0	0	1	4	12
Mother	DISC	DSM sx - SEP	10	519	27 %	13	2.24	2.27	0	0	1	4	12
Mother	DISC	DSM sx - GAD	10	520	26 %	11	2.54	2.25	0	1	2	4	10
Mother	DISC	DSM sx - OCD	10	519	27 %	6	0.39	0.77	0	0	0	1	5
Mother	DISC	DSM sx - MDD	10	520	26 %	19	3.61	3.45	0	1	3	5	18
Child	DISC	DSM sx - SOC	10	471	33 %	13	2.62	3.08	0	0	1	5	12
Child	DISC	DSM sx - SEP	10	471	33 %	12	2.80	2.42	0	1	2	4	11
Child	DISC	DSM sx - GAD	10	467	34 %	11	2.50	2.27	0	1	2	4	10
Child	DISC	DSM sx - OCD	10	459	35 %	7	1.20	1.31	0	0	1	2	6
Child	DISC	DSM sx - MDD	10	466	34 %	21	5.25	4.35	0	2	4	8	21

Note. CBCL = Child Behavior Checklist, TRF = Teacher Report Form, MASC = Multidimensional Anxiety Scale for Children, CDI = Child Depression Inventory, ASRD = study-specific scale described in Appendix, DSM sx = Diagnostic and Statistical Manual symptoms, ADHD = attention deficit hyperactivity disorder, ODD = oppositional defiant disorder, CD = conduct disorder, SOC = social phobia disorder, SEP = separation anxiety disorder, GAD = generalizing anxiety disorder, OCD = obsessive-compulsive disorder, MDD = major depressive disorder, "distinct" = number of distinct non-missing values, p25 = 25% percentile, p75 = 75% percentile.

Table A5

Estimates of Effect of Maternal Depression at Age 2 on Later Child Outcomes

Informant	Scale	Outcome	Age	Cohen's *d* [95% CI] for effect of depressed vs. non-depressed mother at age 2		
				Unadjusted	Matching	Weighting
Mother	CBCL	Externalizing	3	0.51*** [0.36, 0.66]	0.16 [-0.03, 0.35]	0.21* [0.01, 0.40]
Mother	CBCL	Externalizing	4	0.48*** [0.31, 0.64]	0.09 [-0.15, 0.32]	0.15 [-0.08, 0.38]
Mother	CBCL	Externalizing	5	0.35*** [0.19, 0.51]	-0.04 [-0.27, 0.18]	-0.04 [-0.27, 0.18]
Mother	CBCL	Externalizing	7	0.51*** [0.32, 0.71]	0.16 [-0.12, 0.44]	0.17 [-0.12, 0.46]
Mother	CBCL	Externalizing	8	0.45*** [0.26, 0.64]	0.18 [-0.10, 0.46]	0.17 [-0.08, 0.43]
Mother	CBCL	Externalizing	9	0.53*** [0.35, 0.71]	0.30* [0.06, 0.55]	0.28* [0.05, 0.51]
Mother	CBCL	Externalizing	10	0.42*** [0.21, 0.62]	0.13 [-0.15, 0.41]	0.14 [-0.11, 0.39]
Mother	CBCL	Externalizing	14	0.34** [0.12, 0.56]	-0.06 [-0.36, 0.24]	0.06 [-0.22, 0.34]
Mother	CBCL	Internalizing	3	0.51*** [0.36, 0.66]	0.13 [-0.07, 0.33]	0.21* [0.01, 0.41]
Mother	CBCL	Internalizing	4	0.38*** [0.23, 0.54]	0.01 [-0.20, 0.22]	0.17 [-0.03, 0.37]
Mother	CBCL	Internalizing	5	0.25*** [0.13, 0.36]	-0.02 [-0.17, 0.14]	0.04 [-0.11, 0.19]
Mother	CBCL	Internalizing	7	0.41*** [0.25, 0.57]	0.18 [-0.04, 0.40]	0.24* [0.04, 0.43]
Mother	CBCL	Internalizing	8	0.33*** [0.17, 0.49]	0.06 [-0.17, 0.28]	0.14 [-0.06, 0.34]
Mother	CBCL	Internalizing	9	0.32*** [0.16, 0.48]	0.10 [-0.10, 0.31]	0.15 [-0.04, 0.33]
Mother	CBCL	Internalizing	10	0.30*** [0.14, 0.46]	0.00 [-0.21, 0.22]	0.08 [-0.11, 0.28]
Mother	CBCL	Internalizing	14	0.31** [0.11, 0.51]	-0.03 [-0.31, 0.26]	0.13 [-0.12, 0.37]
Mother	DBD-RS	DSM sx - ADHD	10	0.29*** [0.13, 0.46]	0.13 [-0.10, 0.36]	0.10 [-0.10, 0.30]
Mother	DBD-RS	DSM sx - ODD	10	0.35*** [0.19, 0.52]	0.16 [-0.06, 0.39]	0.13 [-0.07, 0.33]
Mother	DBD-RS	DSM sx - CD	10	0.30*** [0.13, 0.46]	0.08 [-0.15, 0.31]	0.12 [-0.09, 0.32]
Mother	DISC	DSM sx - SOC	10	0.15† [-0.02, 0.32]	0.06 [-0.17, 0.30]	0.05 [-0.18, 0.29]
Mother	DISC	DSM sx - SEP	10	0.22* [0.05, 0.39]	0.01 [-0.22, 0.23]	0.07 [-0.15, 0.29]
Mother	DISC	DSM sx - GAD	10	0.32*** [0.15, 0.48]	0.16 [-0.08, 0.40]	0.22* [0.01, 0.43]
Mother	DISC	DSM sx - OCD	10	0.23** [0.06, 0.39]	0.07 [-0.16, 0.29]	0.12 [-0.10, 0.33]
Mother	DISC	DSM sx - MDD	10	0.38*** [0.21, 0.55]	0.22† [-0.01, 0.45]	0.28** [0.09, 0.48]
2nd caregiver	CBCL	Externalizing	3	0.32*** [0.13, 0.51]	0.07 [-0.19, 0.33]	0.05 [-0.19, 0.30]
2nd caregiver	CBCL	Externalizing	4	0.25* [0.05, 0.44]	-0.01 [-0.28, 0.26]	-0.05 [-0.31, 0.22]
2nd caregiver	CBCL	Externalizing	5	0.38*** [0.18, 0.58]	0.16 [-0.12, 0.44]	0.15 [-0.12, 0.42]
2nd caregiver	CBCL	Externalizing	7	0.36*** [0.16, 0.56]	0.08 [-0.20, 0.35]	0.11 [-0.15, 0.36]
2nd caregiver	CBCL	Externalizing	8	0.32** [0.12, 0.52]	0.08 [-0.21, 0.36]	0.06 [-0.20, 0.33]
2nd caregiver	CBCL	Externalizing	9	0.33** [0.13, 0.53]	0.12 [-0.16, 0.40]	0.12 [-0.14, 0.38]
2nd caregiver	CBCL	Externalizing	10	0.31** [0.09, 0.52]	0.09 [-0.22, 0.40]	0.01 [-0.28, 0.31]
2nd caregiver	CBCL	Externalizing	14	0.12 [-0.14, 0.39]	-0.08 [-0.42, 0.26]	0.02 [-0.29, 0.34]
2nd caregiver	CBCL	Internalizing	3	0.19* [0.00, 0.38]	-0.03 [-0.28, 0.23]	-0.01 [-0.28, 0.25]
2nd caregiver	CBCL	Internalizing	4	0.26** [0.08, 0.44]	0.12 [-0.12, 0.37]	0.13 [-0.14, 0.40]
2nd caregiver	CBCL	Internalizing	5	0.14† [-0.02, 0.29]	-0.02 [-0.23, 0.20]	0.02 [-0.21, 0.25]
2nd caregiver	CBCL	Internalizing	7	0.27** [0.10, 0.43]	0.14 [-0.08, 0.36]	0.15 [-0.06, 0.36]
2nd caregiver	CBCL	Internalizing	8	0.28** [0.11, 0.46]	0.06 [-0.17, 0.30]	0.12 [-0.11, 0.35]
2nd caregiver	CBCL	Internalizing	9	0.22* [0.05, 0.40]	0.07 [-0.15, 0.30]	0.10 [-0.11, 0.30]
2nd caregiver	CBCL	Internalizing	10	0.30** [0.11, 0.48]	0.09 [-0.18, 0.37]	0.07 [-0.18, 0.32]
2nd caregiver	CBCL	Internalizing	14	0.11 [-0.16, 0.38]	-0.18 [-0.50, 0.14]	-0.01 [-0.31, 0.29]
Teacher	TRF	Externalizing	7	0.26* [0.06, 0.47]	0.25 [-0.05, 0.54]	0.24† [-0.03, 0.51]
Teacher	TRF	Externalizing	8	0.24* [0.04, 0.45]	0.25† [-0.03, 0.53]	0.22† [-0.04, 0.49]
Teacher	TRF	Externalizing	9	0.21* [0.01, 0.40]	0.14 [-0.13, 0.42]	0.08 [-0.17, 0.33]
Teacher	TRF	Externalizing	10	0.29** [0.10, 0.47]	0.18 [-0.08, 0.44]	0.17 [-0.07, 0.41]
Teacher	TRF	Internalizing	7	0.10 [-0.12, 0.31]	0.05 [-0.24, 0.34]	0.09 [-0.17, 0.36]
Teacher	TRF	Internalizing	8	0.07 [-0.14, 0.29]	0.03 [-0.25, 0.31]	-0.03 [-0.30, 0.24]
Teacher	TRF	Internalizing	9	0.19† [-0.01, 0.40]	0.13 [-0.15, 0.40]	0.03 [-0.23, 0.29]
Teacher	TRF	Internalizing	10	0.30* [0.05, 0.55]	0.29† [-0.04, 0.63]	0.23 [-0.10, 0.56]
Child	ASRD	Deviancy	14	0.06 [-0.11, 0.22]	-0.07 [-0.31, 0.16]	-0.02 [-0.23, 0.19]
Child	MASC	Anxiety	14	0.04 [-0.13, 0.21]	0.07 [-0.16, 0.31]	0.10 [-0.13, 0.32]
Child	CDI	Depression	14	0.08 [-0.09, 0.24]	0.00 [-0.22, 0.23]	0.06 [-0.16, 0.28]
Child	DISC	DSM sx - SOC	10	-0.07 [-0.25, 0.11]	0.04 [-0.21, 0.30]	-0.02 [-0.26, 0.23]
Child	DISC	DSM sx - SEP	10	-0.01 [-0.19, 0.17]	0.00 [-0.25, 0.26]	0.00 [-0.25, 0.25]
Child	DISC	DSM sx - GAD	10	0.05 [-0.13, 0.23]	0.17 [-0.08, 0.42]	0.11 [-0.13, 0.35]
Child	DISC	DSM sx - OCD	10	0.03 [-0.15, 0.20]	0.07 [-0.18, 0.33]	0.03 [-0.22, 0.28]
Child	DISC	DSM sx - MDD	10	-0.03 [-0.22, 0.15]	0.02 [-0.24, 0.27]	-0.04 [-0.28, 0.20]

Note. CBCL = Child Behavior Checklist, TRF = Teacher Report Form, MASC = Multidimensional Anxiety Scale for Children, CDI = Child Depression Inventory, ASRD = study-specific scale described in Appendix. Effect sizes calculated as described in Methods. 95% confidence intervals are in brackets.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table A6

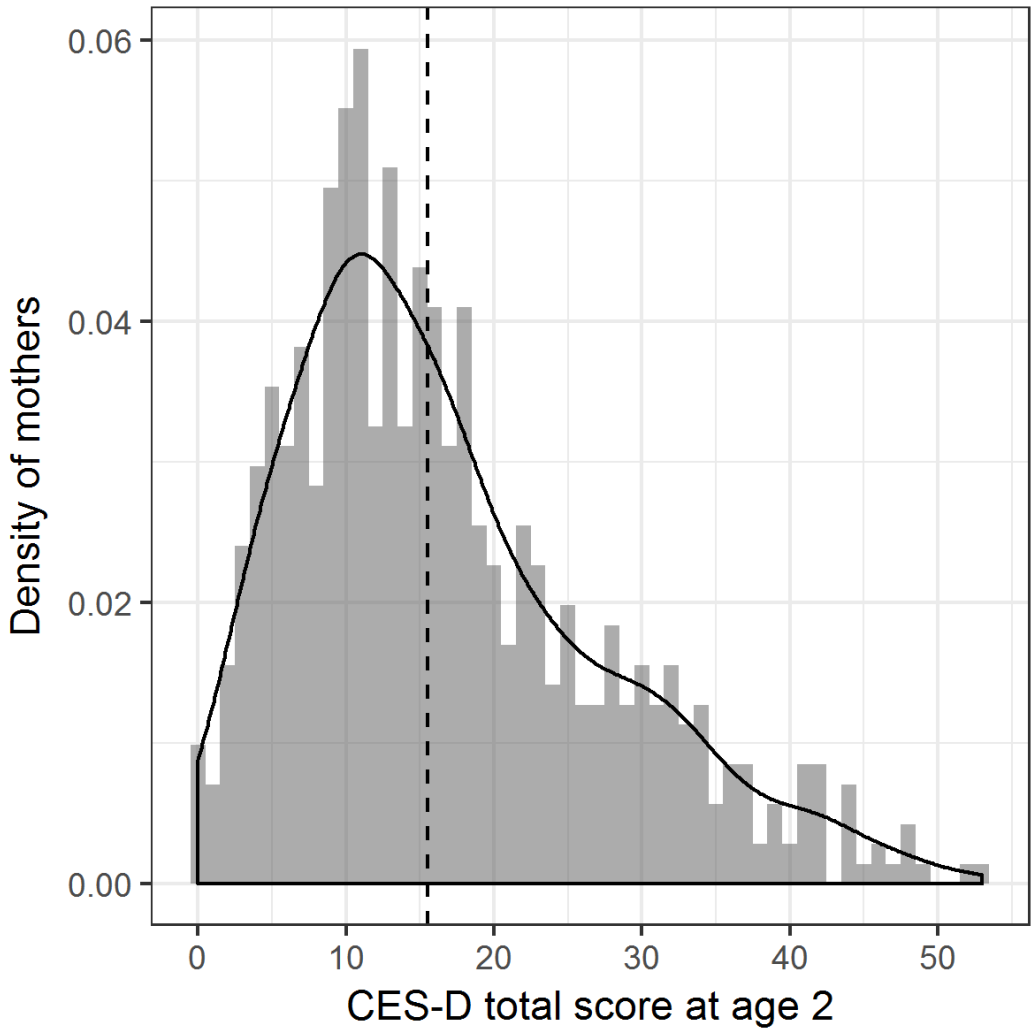
Correlation Matrix for Mothers' CES-D Total Scores Across Waves

	Child Age in Years								
Child Age in Years	**Age 02**	**Age 03**	**Age 04**	**Age 05**	**Age 07**	**Age 08**	**Age 09**	**Age 10**	**Age 14**
Age 02	1.00	0.43	0.40	0.40	0.34	0.34	0.32	0.40	0.31
Age 03	-	1.00	0.54	0.52	0.48	0.38	0.37	0.45	0.32
Age 04	-	-	1.00	0.57	0.51	0.44	0.47	0.45	0.39
Age 05	-	-	-	1.00	0.53	0.47	0.49	0.53	0.47
Age 07	-	-	-	-	1.00	0.56	0.56	0.52	0.45
Age 08	-	-	-	-	-	-	0.55	0.53	0.42
Age 09	-	-	-	-	-	-	1.00	0.64	0.49
Age 10	-	-	-	-	-	-	-	1.00	0.58
Age 14	-	-	-	-	-	-	-	-	1.00

Note. Values are Pearson correlations relating mothers' CES-D total scores at one age (e.g., child age 2) with mothers' CES-D total scores at another age (e.g., child age 3). Correlations calculated based on available data within $N = 707$.

Figure A1

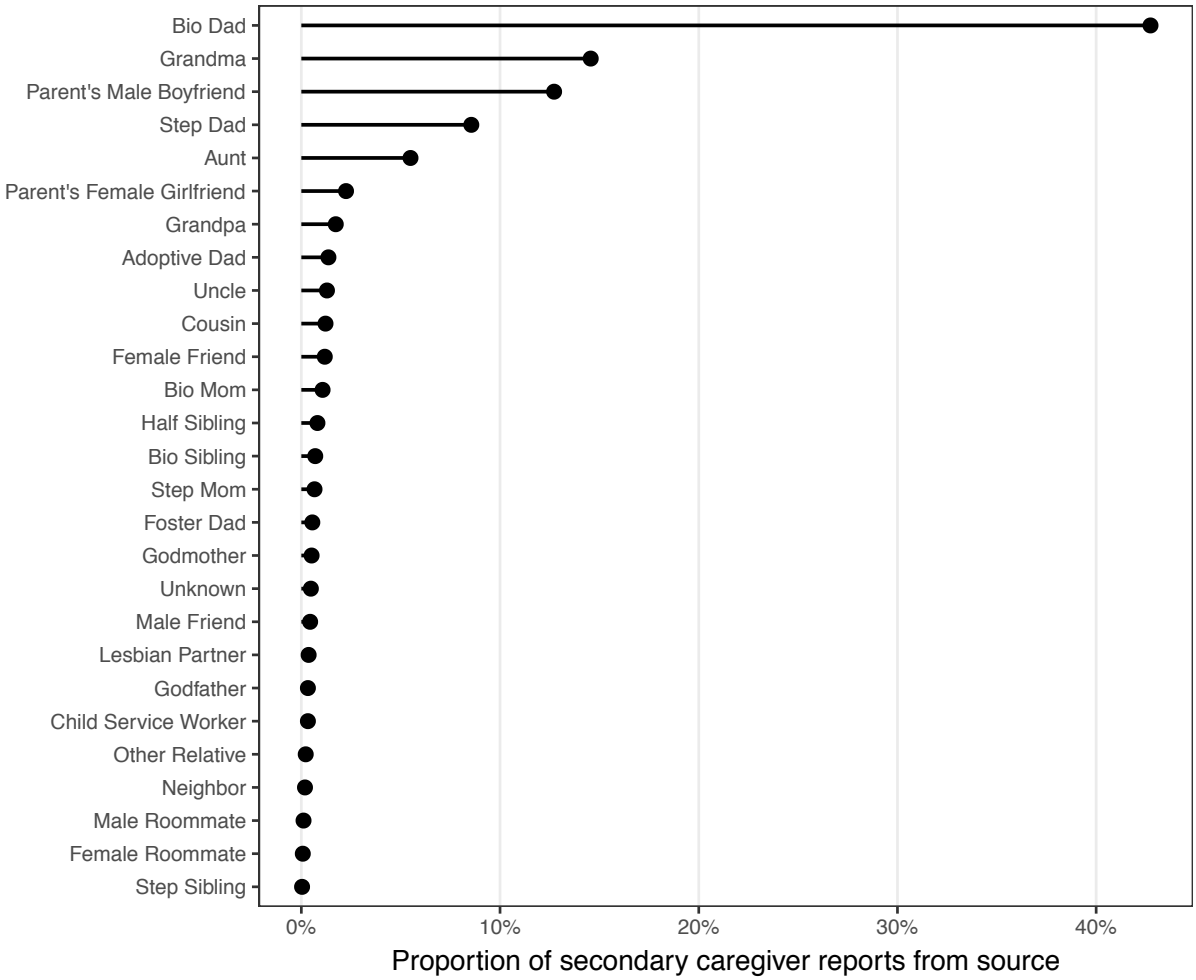
Histogram and Kernel Density Plot of Mothers' CES-D Total Scores at Age 2



Note. Vertical line separates those mothers defined as depressed (>= 16) from those mothers defined as non-depressed. The kernel density plot depicts an estimate of the underlying distribution in the population.

Figure A2

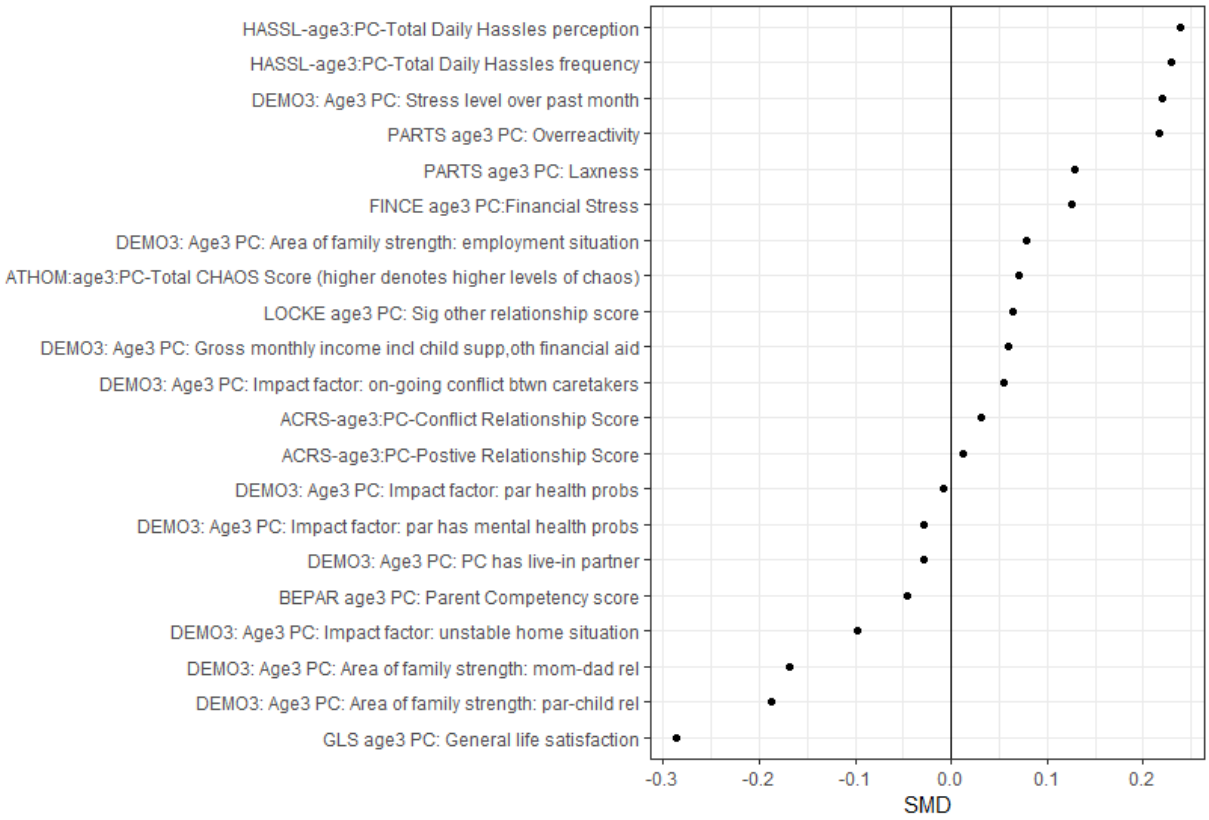
Proportion of Secondary Caregiver Reports of Child Behavior Provided by Persons of Different Relations to Child



Note. Proportion of evaluations of child behavior (i.e., CBCL) provided by secondary caregivers with various relations to the child. Based on available data within the analyzed sample ($N = 707$) at child ages 3, 4, 5, 7, 8, 9, and 10 (relations at age 14 have not yet been tabulated). Note that a very small percentage (~ 1%) of secondary caregiver reports came from biological mother; this could occur at later waves if the child was not under direct care of mother.

Figure A3

Matched Depressed and Non-Depressed Groups Differ on Covariate Values at Child Age 3



Note. Shows select standardized mean differences on age 3 variables between the matched families with depressed (N = 179) and non-depressed (N = 179) mothers at child age 2. Although the two groups were similar on total daily hassles frequency/perception, stress level, and parenting overreactivity at child age 2, the families with depressed mothers are substantially higher on these variables at child age 3.