

# Does Anyone Suffer From Teenage Motherhood? Mental Health Effects of Teen Motherhood in Great Britain Are Small and Homogeneous

Martin O’Flaherty, Sara Kalucza, and Joshua Bon

**ABSTRACT** Teen mothers experience disadvantage across a wide range of outcomes. However, previous research is equivocal with respect to possible long-term mental health consequences of teen motherhood and has not adequately considered the possibility that effects on mental health may be heterogeneous. Drawing on data from the 1970 British Birth Cohort Study, this article applies a novel statistical machine-learning approach—Bayesian Additive Regression Trees—to estimate the effects of teen motherhood on mental health outcomes at ages 30, 34, and 42. We extend previous work by estimating not only sample-average effects but also individual-specific estimates. Our results show that sample-average mental health effects of teen motherhood are substantively small at all time points, apart from age 30 comparisons to women who first became mothers at age 25–30. Moreover, we find that these effects are largely homogeneous for all women in the sample—indicating that there are no subgroups in the data who experience important detrimental mental health consequences. We conclude that there are likely no mental health benefits to policy and interventions that aim to prevent teen motherhood.

**KEYWORDS** Teenage parenthood • Mental health • Causal inference • Bayesian methods • Statistical machine learning

## Introduction

Both popular and academic narratives paint teenage motherhood as a tragedy for mother and child (Duncan 2007; Tyler 2008), and it is true that teen mothers face myriad challenges, including economic hardship, health difficulties, and disrupted relationships (Angelini and Mierau 2018; Diaz and Fiel 2016; Ermisch and Pevalin 2005; Gorry 2019; Sironi et al. 2020). Considering these stressors and often abusive public discourse, high rates of mental health problems among teen mothers are perhaps unsurprising, with evidence indicating that teen mothers suffer from pre- and postnatal depression at rates several times higher than adult mothers (Hodgkinson et al. 2014). Interpreting these observations is, however, challenging. Notably, the extent to which mental health problems are attributable to teen motherhood *per se*

(vs. preparenthood disadvantage) or persist over the long term represent points of continuing disagreement. Quantitative evidence for long-term mental health *effects* of teen motherhood remains mixed, with studies supporting detrimental, neutral, and in rare cases beneficial outcomes after adjustment for confounding (Aitken et al. 2016; Angelini and Mierau 2018; Grundy et al. 2020; Güneş 2016; Hillis et al. 2004; Kalil and Kunz 2002; Kravdal et al. 2017; Mollborn and Morningstar 2009; Patel and Sen 2012; Whitworth 2017; Xavier et al. 2017, 2018). Teenage mothers themselves, moreover, appear decidedly ambivalent about the meanings and consequences of young parenthood. While they may acknowledge disrupted life plans, negative stereotypes, and a lack of personal and financial preparedness for parenthood, teen mothers often also see the birth of their child as a source of meaning, purpose, and connection that may act as a “turning point” in their lives (Brubaker and Wright 2006; Edin and Kefalas 2005; Jones et al. 2019; Yardley 2008).

In this article we analyze the relationship between early motherhood and long-term mental health outcomes using data from a cohort of British women born in 1970. In doing so, we contribute to the literature in several ways. First, important concerns remain regarding causal interpretation of differences in mental health by motherhood timing. Women who become mothers at a young age are disadvantaged before becoming parents (Kalucza 2018; Mollborn and Morningstar 2009), and poor outcomes reported by previous studies may therefore reflect uncontrolled confounding, particularly because most previous studies controlled for only a limited set of prior confounders. Incorporating data on a rich set of controls (collected prospectively from birth to adolescence) offers stronger evidence for or against causality. Second, existing evidence is generally consistent with small detrimental effects of teen motherhood on mental health *on average*. However, it is possible that this masks subgroups in the population for whom the effects may be more strongly negative or, alternatively, beneficial among adolescents with the most favorable attitudes toward pregnancy (Mollborn 2017; Whitworth 2017). Mental health effects may also vary over the life course as teen mothers move beyond direct parenting roles and responsibilities. Indeed, several studies found that detrimental mental health effects of teen motherhood were concentrated among younger women (Aitken et al. 2016; Grundy et al. 2020; Güneş 2016), but these studies could not distinguish between life stage and cohort differences. We estimate the effects of teen motherhood on mental health at three time points (ages 30, 34, and 42) and investigate potential moderation by a wide range of preparenthood characteristics.

Analytically, we employ a novel methodology—Bayesian Additive Regression Trees (BART) (Chipman et al. 2010). BART represents a highly flexible estimation approach that allows for complex relationships between confounders, motherhood timing, and mental health outcomes, and it has important advantages as a tool for causal inference (Hill 2011). Notably, even if the set of observed confounders is sufficient for nonparametric identification of causal effects, biases may still arise if the functional form is misspecified (e.g., assumed linearity when the true relationship is nonlinear (Ho et al. 2007). Theory, while indispensable for study design and interpretation, typically provides no guidance on functional forms. The flexibility of BART offers improved estimates of both average and heterogeneous causal estimands in circumstances without requiring prior knowledge of the “true” relationships (Dorie et al. 2019; Hahn et al. 2020; Hill 2011; Wendling et al. 2018).

## Understanding the Relationship Between Teen Motherhood and Mental Health

### The Case for Causal Effects: Evidence and Mechanisms

Studies that have identified detrimental long-term effects of young motherhood on mental health span a range of countries and assessed mental health at different points in the life course, from around age 30 to midlife (see Xavier et al. 2018 for a recent review). In Great Britain, Maughan and Lindelow (1997) analyzed the 1946 and 1958 British birth cohorts and reported elevated psychiatric morbidity in the mid-30s among teenage mothers born in 1958 (but not 1946). Using the English Longitudinal Study of Ageing, Grundy et al. (2020) similarly found that teen motherhood increased the log-odds of experiencing high depressive symptoms in 2010 (at ages 55–64) by 0.95 (equivalent to an odds ratio of 2.59) among women born in 1946–1955. They also reported smaller (nonsignificant) effects of teen motherhood among older women (65+). Outside Great Britain, recent studies include Angelini and Mierau's (2018) analysis of 13 European countries. After adjusting for family background, adolescent health, and academic performance, these authors found that the marginal probability of experiencing symptoms of depression in midlife (ages 49–87 in 2008–2009) is five to six percentage points higher among teenage mothers. Australian evidence also supports detrimental mental health effects of teen motherhood among women in midlife or older (40+), with effects ranging from one quarter to one half of the sample standard deviation of the outcome (mental health component of the SF-36 Health Survey) dependent on birth cohort (Aitken et al. 2016).

Most evidence supporting detrimental effects of teen motherhood comes from studies that rely on selection-on-observables assumptions (i.e., regression adjustment or propensity score approaches). These designs are generally considered to provide weaker evidence of causality; however, two recent studies using sibling or twin fixed-effects designs also indicated poorer mental health among teen mothers (Güneş 2016; Kravdal et al. 2017). Güneş (2016) presented both sibling and twin fixed-effects analyses of a sample of U.S. women, finding that the mental health consequences of teen motherhood varied by age, with no effect among older women (46 or older in 1996) but substantively large negative effects on the likelihood of reporting good mental health among younger women (25–45 in 1996). Kravdal et al. (2017) analyzed Norwegian register data from 2004–2008 for sisters aged 45–73, finding increased odds of purchasing antidepressants among women whose first birth occurred at age 21 or earlier (compared with 26 or older). The magnitude of these effects appears to increase with completed parity, from 17% higher odds among those with only one child to 58% for those with four or more children.

Teen motherhood may be causally related to later-life mental health through multiple interdependent mechanisms, including (1) stress proliferation, (2) stigmatization, and (3) sensitive period effects. Occurring at a stage of the life course when key financial and social supports have not yet been established, young motherhood may result in a stressful parenting environment in which inadequate resources are available to meet the demands of raising children. This may produce role strain and contribute to a process of *stress proliferation*, whereby the occurrence of one stressor (e.g., teen motherhood) increases the likelihood of experiencing a range

of other stressors (e.g., low income, single parenting) that accumulate over time (Pearlin 1989). This process is likely multidimensional, as teen motherhood has been linked to a range of intermediate outcomes that may be consequential for mental health. The mother's own economic outcomes are perhaps the most widely studied. Estimates vary widely, but the best available evidence is broadly consistent with small negative effects on educational attainment, earnings, and employment (e.g., Ashcraft et al. 2013; Diaz and Fiel 2016; Gorry 2019; Kane et al. 2013). We note, however, that Ermisch and Pevalin's (2005) analysis of the 1970 British birth cohort (the same women we study in the present article) found no negative effects of teen motherhood on post-16 education, employment, earnings, or occupation at age 30.

The literature examining relationships between teenage motherhood and other potential stressors is comparatively less developed, but includes evidence of adverse effects on physical health (Güneş 2016; Patel and Sen 2012; Sironi et al. 2020; Webbink et al. 2008), health behaviors (Güneş 2016; Webbink et al. 2008; Wolfe 2009; see Fletcher 2012 for a contrasting study that found no effect, or protective effects, of teen motherhood on health behavior), home ownership and housing wealth (Ermisch and Pevalin 2004; Maughan and Lindelow 1997), and various indicators of partner "quality," including education, employment, smoking, and alcohol consumption (Ermisch and Pevalin 2005; Webbink et al. 2008). Several studies also analyzed potential mediators of the relationship between teen motherhood and adult mental health, reporting findings that are generally consistent with stress proliferation as a mechanism. Studies that have assessed mediation include Angelini and Mierau (2018), Grundy et al. (2020), Maughan and Lindelow (1997), and Falci et al. (2010), all of which report that effects of teen motherhood on mental health are attenuated or rendered nonsignificant after adjusting for mid-life circumstances. The mediators considered vary by study, but include educational attainment, income, wealth, family size, relationship history, financial strain, and perceived personal control.

A second major mechanism for effects of teen motherhood is *stigmatization*, defined by Pescosolido and Martin (2015:92) as "a social process embedded in social relationships that devalues through conferring labels and stereotyping." Teen mothers are often stereotyped as being welfare dependent, promiscuous, and irresponsible, encapsulated in the UK by the derogatory "pramface" label (Nayak and Kehily 2014; SmithBattle 2013; Yardley 2008). Stigma affects mental health through discriminatory treatment (e.g., in schools or health services), social isolation, and internalization of negative stereotypes and judgments (Hatzenbuehler et al. 2013). A range of evidence indicates that teen mothers are aware of stigmatizing behavior and attitudes from family, peers, teachers, media, and health professionals, and that perceived stigma is associated with social isolation (McMichael 2012; Wiemann et al. 2005; Yardley 2008). Given the high rates of antenatal and postnatal mental health problems among adolescent mothers (Hodgkinson et al. 2014), a particularly concerning consequence of stigma may be exclusion from timely and effective health care. Teen mothers often experience and anticipate discrimination from health care providers and may fear that revealing mental health concerns risks inviting punitive intervention from child protection authorities, further undermining the quality of care they receive (McArthur and Winkworth 2018;

Recto and Champion 2018; SmithBattle 2013; Yardley 2008). Negative stereotypes may also be accepted by teen mothers. Qualitative accounts describe teen mothers' efforts to resist stigma and reframe their experiences of motherhood positively, however, this is often accomplished not by challenging stereotypes but rather by distancing themselves (as good mothers) from "other" adolescent mothers whom they describe as *conforming to* the stereotype (Jones et al. 2019; Yardley 2008). These accounts suggest, at minimum, a degree of vulnerability to internalized stigma among teen mothers.

Third, the teen years encompass a period of rapid neurological change (particularly in the prefrontal cortex) and coincide with the onset of many mental health disorders (Blakemore and Mills 2014; Larsen and Luna 2018). This suggests that adolescence may be a *sensitive period* for mental health, when the developing brain is particularly vulnerable to stress and isolation, and insults to neurological development may persist over long stretches of the life course. The transition to parenthood, whenever it occurs, is often accompanied by an array of challenges, including disrupted sleep, high care demands, shifting social relationships, and mental health problems (Nomaguchi and Milkie 2020). Exacerbated by limited resources and stigma, these challenges may be both common and particularly consequential for adolescent mothers' mental health (Hodgkinson et al. 2014). Adolescence is also characterized by heightened sensitivity to peer judgments (Blakemore and Mills 2014), potentially compounding the effects of stigma and social exclusion on teen mothers' well-being.

### Selection and Confounding

While credible evidence and theory support causal effects of teen motherhood on mental health, it is also clear that young motherhood is *caused by* early-life disadvantage. It is therefore possible that associations between early motherhood and later-life mental health status may represent failure to adequately control for selection into young motherhood. For instance, adverse childhood experiences (Hillis et al. 2004), socioeconomic disadvantage (Penman-Aguilar et al. 2013), family instability (Fomby and Bosick 2013), and poor adolescent mental health (Kalucza 2018; Mollborn and Morningstar 2009) are all factors that have been identified as causes of young motherhood that are also plausibly causes of later-life mental health. Moreover, qualitative studies often highlight how, in opposition to prevailing stereotypes, young women identify parenthood as a source of positive identity, meaning, and motivation in circumstances where they might not otherwise have access to normative career and family formation pathways (Brubaker and Wright 2006; Edin and Kefalas 2005; Jones et al. 2019; Yardley 2008). Contrary to the dominant narrative of young motherhood as a contributory factor to poor mental health, these accounts suggest that associations between young motherhood and mental health may simply reflect background disadvantage, and young motherhood may carry positive consequences in some instances. Many quantitative studies also found no effect of teen motherhood on later mental health after controlling for preparenthood disadvantage (Hillis et al. 2004; Kalil and Kunz 2002; Mollborn and Morningstar 2009; Patel and Sen 2012; Xavier et al. 2018; Xavier et al. 2017).

## Heterogeneous Effects of Teen Motherhood

To date, most studies have focused on adjudicating between causality and confounding as competing explanations for the mental health outcomes of adolescent mothers. While clearly important, the attention devoted to this problem has perhaps detracted from our understanding of how women may be *differentially affected* by teen motherhood. This is an important omission, highlighted by Mollborn (2017) as a key direction for future research on teen mothers. One of the few studies to directly investigate variation in the mental health effects of teen motherhood is that of Whitworth (2017). She concluded that there were minimal effects of teen motherhood among those that expressed the most negative preparenthood attitudes toward teen pregnancy, and *better* mental health among teen mothers who expressed the most positive attitudes. Studies have also investigated heterogeneity in the effects of teen motherhood on mental health across countries and generations. For example, in the UK, the United States, and Australia, teen motherhood has been found to have substantially larger effects on mental health among younger cohorts (Aitken et al. 2016; Grundy et al. 2020; Güneş 2016), although it is unclear if this reflects age or cohort differences.

Effects of teen motherhood may vary because the intervening *mechanisms* operate differently depending on resources, social context, and life stage. Studies of potential links in the stress proliferation process provide mixed evidence, but broadly suggest that effects of teen motherhood may be greatest in circumstances where teen motherhood is uncommon. Diaz and Fiel (2016) analyzed data from the child and young adult cohorts of the National Longitudinal Survey of Youth 1979 and concluded that disadvantaged teens (who are more likely to become pregnant) experience few negative consequences of teen pregnancy, while teens from more advantaged backgrounds suffer larger reductions in education and earnings in early adulthood. Gorry (2019) examined heterogeneity in the effects of teen motherhood on earnings, education, and welfare receipt over socioeconomic and racial groups and found similarly that non-Hispanic Whites and those from advantaged neighborhoods are most negatively affected. Cross-national evidence in Grundy and Foverskov (2016) indicates that countries where teen motherhood is a comparatively *normative* life course event (e.g., Eastern Europe) show the weakest associations between teen parenthood and long-term physical health. Güneş (2016) found larger detrimental effects of teen motherhood on chronic conditions, physical activity, and preventative health care use among younger women (born 1960–1970).

## Challenges of Estimating Causal Effects From Observational Data

Obtaining valid causal estimates is a perennial problem across the social sciences. Although there are a range of alternative approaches, practical and ethical considerations mean that many studies continue to rely on some form of adjustment for observed confounders (i.e., regression or propensity score approaches). Formally, these approaches require an assumption of *ignorability*, expressed as  $Y(0), Y(1) \perp\!\!\!\perp A \mid \mathbf{X}$ . The ignorability assumption stipulates that (binary) treatment  $A$  is unrelated to the potential outcomes ( $Y(0), Y(1)$ ), conditional on observed confounders  $\mathbf{X}$ —in other words, there is no residual confounding that has not been measured and adjusted for.

This is generally regarded as a very strong assumption that warrants a high degree of scepticism. Incorporating a richer set of potential confounders may, however, render the ignorability assumption more plausible, suggesting that it is often desirable to extend the set of variables included in  $\mathbf{X}$  to minimize residual confounding bias.

Even if ignorability is satisfied, analysts must still fit a model—for the probability of treatment conditional on confounders, or for the outcome conditional on treatment and confounders (or both). Focusing on the outcome model, this may be written in general terms as

$$Y = f(A, \mathbf{X}) + \epsilon.$$

The challenge then becomes specifying  $f(A, \mathbf{X})$  appropriately—the function that links treatment and confounders to the outcome. Misspecification of  $f(A, \mathbf{X})$  may bias causal estimates (Ho et al. 2007) and is complicated by the usual absence of theory justifying any specific functional form. A tension also exists between the potential benefit of extending  $\mathbf{X}$  (to support the ignorability assumption) and the complexity of the model specification task, which becomes more difficult as the dimensionality of  $\mathbf{X}$  increases. In practice, researchers often fit multiple candidate models—this will often result in better in-sample fit, but creates other problems. Once data have been used to select a model, standard errors from a final analysis conducted using that same data are no longer valid without additional corrections that are not part of standard practice (Berk et al. 2013). This is because model selection is stochastic—under repeated sampling, different models would be selected based on random variation in the observed data—and this variability is not incorporated in the calculation of standard errors (Berk et al. 2013). Moreover, presented with a range of estimates, analysts may be tempted to choose those that conform to prior beliefs, exceed conventional thresholds for null-hypothesis significance testing, or are otherwise more “interesting” in some way. Propensity score approaches, while avoiding the need to model the outcome, must still correctly specify a model for the probability of treatment conditional on covariates.

Similar challenges arise regarding heterogeneous causal effects, where interest centers on how the effect of “treatment” is moderated by covariate(s). We briefly discuss these problems in the context of two predominant approaches to effect heterogeneity. Regression models including interactions between covariates and treatment represent the predominant approach to effect heterogeneity. This requires that the analyst correctly specify  $f(A, \mathbf{X})$ , and it suffers from the same problems discussed above. The common fallback option (linearity assumptions for both “main effects” and interactions) is often unreliable and may lead to fragile and model-dependent estimates of the quantities of interest. To illustrate, a recent reanalysis of papers using linear covariate-by-treatment interactions in leading political science journals concluded that the majority are unreliable because of either neglected nonlinearity or lack of common support for the moderator between treated and untreated groups (Hainmueller et al. 2019). While the substantive content of these papers is distinct from the current context, we suspect that these kinds of issues are likely common across the social sciences.

An alternative approach, proposed by Xie et al. (2012), aims to identify heterogeneity in treatment effects as a function of the propensity score, which is the

probability of treatment conditional on pretreatment covariates. Xie et al.'s proposal incorporates a series of primarily nonparametric approaches to accomplish this task and has proved useful in applied research (e.g., Diaz and Fiel 2016). This approach requires a correctly specified model relating treatment to covariates and considers only effect heterogeneity by the propensity score. In general, treatment effects may be a function of other covariates that are unrelated to the probability of treatment or may vary in response to covariates that are both positively *and* negatively associated with treatment. Therefore, the propensity score approach may obscure important, substantively interesting effect heterogeneity.

In the current study, we demonstrate the application of BART (Chipman et al. 2010; Hill 2011) to the estimation of both the familiar average treatment effect (ATE) and heterogeneous causal effects. We present a full description in the next section but note here that BART provides a highly flexible model fit by modeling the outcome as the sum of many small regression trees, which are regularized by priors that “shrink” the trees toward the null. The individual trees naturally incorporate nonlinearities and interactions (including interactions of treatment with any covariate), and consequently the overall model inherits this property without needing prior knowledge of the true functional forms. BART models avoid the issues with model selection by selecting the best predictors and interactions during the fitting process. A growing body of simulation evidence indicates that BART often outperforms traditional estimation methods (e.g., linear regression, propensity score methods) and modern competitors such as causal forests (Brand et al. 2021; Wager and Athey 2018) as an estimator for both average and heterogeneous treatment effects (Dorie et al. 2019; Hahn et al. 2020; Hill 2011; Wendling et al. 2018). BART performs well in settings where  $\mathbf{X}$  may be relatively high-dimensional or includes irrelevant covariates (Chipman et al. 2010) and is only marginally less efficient than *correctly specified* linear models (Hill 2011). These features enable us to control for a much more extensive set of potential confounders than would be ordinarily feasible.

## Data and Methods

Data for the study are drawn from the 1970 British Birth Cohort Study (BCS70). The BCS70 follows everyone living in England, Scotland, and Wales who were born in a single week of 1970 (Elliott and Shepherd 2006). The initial sample included 17,196 people, 9,842 of whom remained in the study at the age 42 wave in 2012 (74.6% of 13,189 traced and eligible cohort members) (TNS BMRB 2012). The survey covers many aspects of family circumstances, health, education, and social development from childhood, adolescence, and adulthood. Detailed information about the study can be found in the cohort profile (Elliott and Shepherd 2006) and technical report (TNS BMRB 2012).

### Dependent and Independent Variables

We define “teen motherhood” as giving birth before the age of 20. Age at first birth was calculated by subtracting the respondents' birth year from the year of reported



births, using information about all reported live births from the BCS70 waves at ages 30 and 42.

Two measures of mental health and well-being are used as outcomes. Our first outcome is the Malaise inventory (Rutter et al. 1970), measured at ages 30, 34, and 42. The inventory consists of 24 yes/no self-completion questions, which combine to measure levels of psychological distress or depression (Rutter et al. 1970). The scale has been validated for general population samples (Rodgers et al. 1999) and covers emotional disturbance and associated physical symptoms, with scores ranging from 0 to 24. Scores are dichotomized at 8+ for the full scale at age 30 and at 4+ for the nine-item scale at ages 34 and 42.

Our second outcome is the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) at age 42 (Tennant et al. 2007). WEMWBS is a 14-item scale of mental well-being covering subjective well-being and psychological functioning, including positive affect (feelings of optimism, cheerfulness, relaxation), satisfying interpersonal relationships, and positive functioning (energy, clear thinking, self-acceptance, personal development, competence, autonomy) (Tennant et al. 2007). The scale is scored by summing responses to each item on a Likert scale of 1 to 5, with a minimum score of 14 and a maximum of 70. High scores indicate better mental well-being. The WEMWBS has the advantage of being free of ceiling effects in population samples and is thus more sensitive to variation in mental well-being among the nonclinical population. Our analytic sample had a mean WEMWBS score of 48.7, slightly lower than the provisional Scottish population mean score of 50.7 (Stewart-Brown and Janmohamed 2008).

### Covariates and Missing Data

To select potential confounders, we first identified topic domains we wished our covariates to cover on the basis of previous research and theory. The 13 domains included delinquency, education attainment and aspirations, family history and family stability, health behavior, family housing situation, mental health in childhood, orientation to the world, parents' parenting strategy, peers and peer characteristics, physical health, parents' physical health, relationship with parents, and parent and grandparent education and social class. We selected 70 covariates of interest with acceptable levels of missing data, measured from ages 0 to 16. We also included an estimate of the propensity score (the probability of teen motherhood conditional on covariates, estimated using BART) based on evidence that this improves the quality of causal estimates (Hahn et al. 2020). An overview of variables and domains can be found in online appendix Table A1, and descriptive statistics in Table A2.

Missing data for background covariates were imputed using multiple imputation by chained estimation in the mice R package (van Buuren 2012). Cases with missing outcomes or fertility data were excluded. A total of 100 imputed data sets were created.

### Bayesian Additive Regression Trees

BART (Chipman et al. 2010) is a tree-based regression model notable for flexibility and parsimonious modeling of many variables. BART can capture high-order

variable interactions and account for uncertainty by representing model parameters probabilistically in the Bayesian framework. This section will first describe the general structure of BART, in relation to standard regression, and follow with a description of the tree structure employed by BART.

BART models generally have the same structure as other regression models, that is, for continuous outcome  $Y$ , with treatment  $A$  and covariates  $\mathbf{X} = [X_1, X_2, \dots, X_p]$ , the overall regression model is described by

$$Y = f(A, \mathbf{X}) + \epsilon,$$

where  $f(A, \mathbf{X})$  is a model of the mean of the outcome conditional on treatment and covariates, and  $\epsilon$  is zero-mean normally distributed noise with variance  $\sigma^2$ . In the case of BART,  $f(A, \mathbf{X})$  is a sum of many regression trees (see the following for details). In comparison, standard linear regression simply uses the linear function. The extension of BART to noncontinuous outcomes, such as binary responses, is analogous to generalized linear models (Nelder and Wedderburn 1972) in that  $f(A, \mathbf{X})$  is connected to the mean of the outcome through the link equation

$$E(Y) = h^{-1}[f(A, \mathbf{X})],$$

where  $h$  is the link function. We use the probit link for the binary Malaise outcomes.

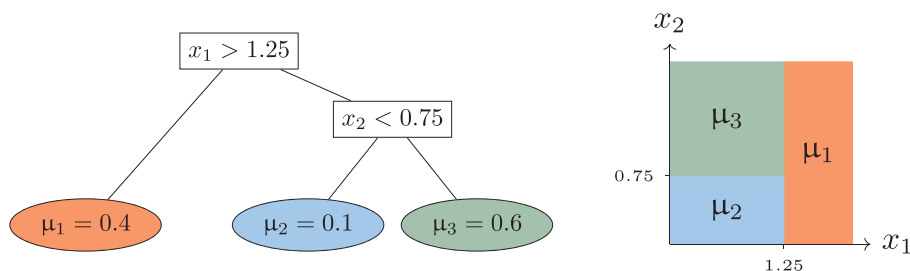
BART assumes a sum of trees structure for the mean model  $f$ . Each tree is a regression tree—a binary tree  $\mathbf{T}$  consisting of successive nodes of decision rules and a layer of terminal nodes with mean values  $\mathbf{M} = [\mu_1, \mu_2, \dots, \mu_m]$ , where  $m$  is the number of terminal nodes. Each tree recursively partitions the data into several subgroups with a particular mean value. This process is denoted by the function  $g(A, \mathbf{X}; \mathbf{T}, \mathbf{M})$ , which takes inputs of treatment and covariates, tree structure with decisions, and mean values at terminal nodes. An illustrative example of a single tree is given in Figure 1.

Rather than a single large tree (which would be unstable and have high variance), BART uses the sum of many smaller trees (commonly 200). Each tree is constrained by priors to be a weak learner, which retains flexibility but penalizes overfitting (Chipman et al. 2010). The sum of trees structure is described mathematically as

$$f(A, \mathbf{X}) = \sum_{k=1}^K g(A, \mathbf{X}; \mathbf{T}_k, \mathbf{M}_k),$$

where  $K$  is the number of trees.

The model is fit using Markov chain Monte Carlo (MCMC) as described in Chipman et al. (2010). Model summaries and fitted values are defined in terms of the Bayesian posterior samples from the MCMC procedure (Gelman et al. 2014). Models were estimated separately for each imputed data set, with the first 1,000 MCMC iterations discarded as “burn-in” and 1,000 posterior samples retained. Final estimates are constructed by pooling the retained posterior samples (Gelman et al. 2014), totaling 100,000 samples. We used the BART R package (McCulloch et al. 2021) to estimate all models, and the tidyreatment package (Bon 2021) to extract causal estimates.



**Fig. 1** Decision tree example: A single decision tree (left) with three terminal nodes,  $\mu_1, \mu_2$ , and  $\mu_3$ . Left branches indicate that the preceding nodes' condition is true, while right branches indicate that they are false. The tree function  $f(x_1, x_2)$  assigns values to the parameter space based on the tree. For this tree, any  $(x_1, x_2)$  with  $x_1 > 1.25$  will have  $f(x_1, x_2) = 0.4$ . The nodes  $\mu_2$  and  $\mu_3$  are interaction effects as they depend on both  $x_1$  and  $x_2$ . A second representation of  $f(x_1, x_2)$  is given on the right.

### BART for Estimation of Causal Effects

Causal inference with BART relies on assumptions of ignorability and positivity (Hill 2011). Ignorability ( $Y(0), Y(1) \perp\!\!\!\perp A \mid \mathbf{X}$ ) stipulates that (binary) treatment  $A$  is unrelated to the potential outcomes ( $Y(0), Y(1)$ ), conditional on the observed confounders  $\mathbf{X}$ . Positivity ( $0 < p(A=1 \mid \mathbf{X}) < 1$ ) requires that there is a nonzero probability of receiving every level of the treatment for all values of the confounders. Given these assumptions, the conditional average effect of treatment (CATE) for subjects with observed confounder values  $\mathbf{X}=\mathbf{x}$  is  $E(Y \mid A=1, \mathbf{X}=\mathbf{x}) - E(Y \mid A=0, \mathbf{X}=\mathbf{x})$  (Hill 2011). Estimation of causal effects therefore resolves into the problem of estimating the response surfaces under treatment and no treatment.

Sample-average effects for any subgroup of interest (including the ATE or Average Treatment Effect on the Treated, ATT) may be calculated by averaging the CATE over subjects in the relevant group. For example, the ATT is calculated as the average of the CATE for women in the sample who became young mothers. Uncertainty is quantified through variation in estimates over MCMC iterations. Specifically, we report 95% credible intervals based on a normal approximation, based on evidence in Carnegie (2019) that this type of credible interval exhibits better nominal coverage. In keeping with best-practice reporting advice (Amrhein et al. 2019), we discuss throughout the range of parameter values that are compatible with our estimates (high and low) in addition to point estimates.

## Results

Table 1 presents sample descriptive statistics for age at first birth and the four outcome variables. Teen mothers fare notably worse on all outcomes. For the three Malaise items, teen parents are roughly 10 to 12 percentage points more likely to experience elevated levels of distress than women who became mothers at an older age at each time point. For the WEMWBS, teen parents average 2.1 points lower, equivalent to roughly .23 of the sample standard deviation. Sample sizes differ by outcome, ranging from 2,538 to 2,777. To check the positivity assumption, we plot

**Table 1** Sample summary statistics by outcome and young motherhood

Outcome Variable	Normative-Age Mothers (20–30)	Young Mothers (<20)	Total	Unadjusted Difference
<b>Malaise Inventory (age 30)</b>				
Low (0–7)	1,993 (87.2%)	366 (75.9%)	2,359 (85.3%)	
High (8–24)	292 (12.8%)	116 (24.1%)	408 (14.8%)	11.3%
Total	2,285	482	2,767	
<b>Malaise Inventory (age 34)</b>				
Low (0–3)	1,729 (81.6%)	292 (69.7%)	2,021 (79.6%)	
High (4–9)	390 (18.4%)	127 (30.3%)	517 (20.4%)	11.7%
Total	2,119	419	2,538	
<b>Malaise Inventory (age 42)</b>				
Low (0–3)	1,837 (80.3%)	342 (70%)	2,179 (78.5%)	
High (4–9)	451 (19.7%)	147 (30%)	598 (21.5%)	10.3%
Total	2,288	489	2,777	
<b>WEMWBS (age 42)</b>				
Mean (SD)	49.1 (8.5)	47.0 (9.5)	48.8 (8.7)	–2.1
Total	2,132	446	2,578	

Note: WEMWBS = Warwick-Edinburgh Mental Well-being Scale.

the estimated propensity score separately for teen mothers and older mothers in Figure 2. The plot shows sufficient overlap, albeit with potential problems at very low propensity scores.

Table 2 presents the estimated sample ATE for each of the mental health outcome measures: dichotomous indicators of depressive symptoms based on the Malaise inventory at ages 30, 34, and 42, and the WEMWBS at age 42. Malaise outcomes are modeled using the probit link, and the WEMWBS is modeled as continuous. All models control for the full set of covariates described in online appendix Tables A1 and A2. Estimates for the binary Malaise outcomes are expressed as percentage-point differences, and estimates for the continuous WEMWBS outcome are expressed in raw (unstandardized) units. In all cases, our estimates show that the ATEs are small in magnitude. We find weak evidence of an increased risk of poor mental health for teen mothers at age 30 (3.6 percentage points; CI: 0.01–7.2 percentage points) after adjusting for controls, although the lower bound of the credible interval is very close to 0. Otherwise, although the upper bounds of the credible intervals are consistent with small detrimental effects, the interval overlaps 0 for all dependent variables, indicating a slight possibility that the true effects of young motherhood on later-life mental health are in fact protective. Moreover, this indicates that our analysis does not provide any strong evidence in support of deleterious average causal effects of young motherhood beyond age 30, and we note a slight decline in the magnitude of point estimates for the Malaise outcomes at older ages. With respect to the WEMWBS scores, we note that the 95% credible interval is consistent with, *at most*, a quite small detrimental effect equal to less than one sixth of a standard deviation (–1.2). The point estimate for the ATE of young parenthood on the WEMWBS scores corresponds to an effect of less than 4% of a standard deviation. In contrast, for the Malaise outcomes, the upper bounds of the credible intervals correspond to an increase in the risk of psychological distress of roughly seven

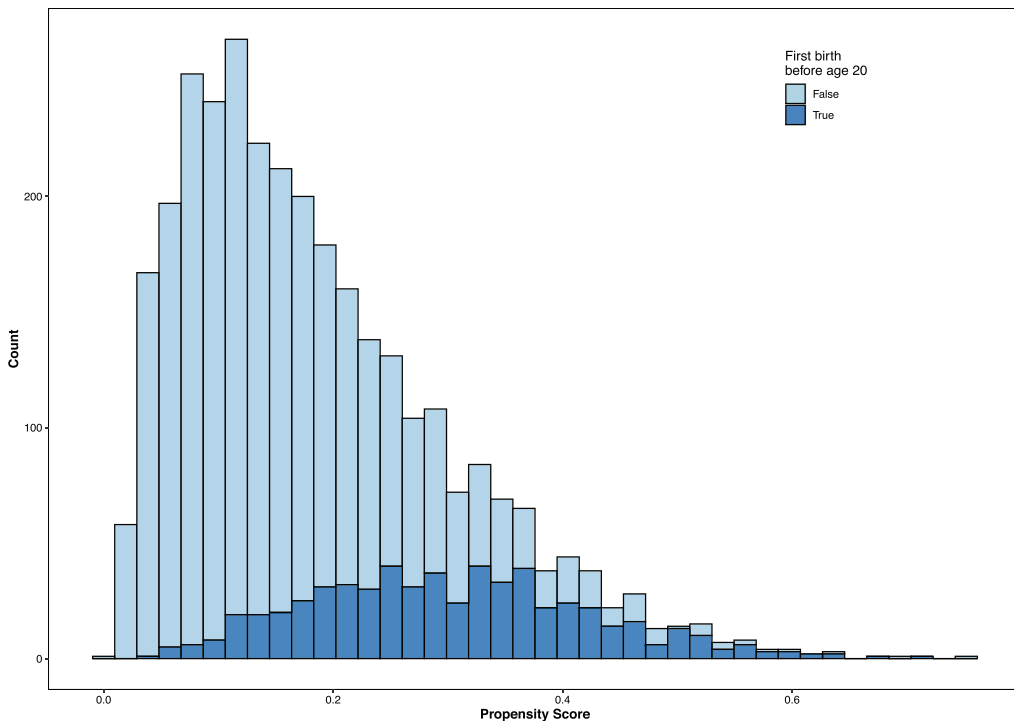


Fig. 2 Propensity score overlap by young motherhood status

Table 2 Average effects of young motherhood by outcome

Outcome Variable	ATE Point Estimate (95% credible interval)
Malaise (age 30)	3.6 (0.01, 7.2) <sup>a</sup>
Malaise (age 34)	3.4 (-0.9, 7.7)
Malaise (age 42)	3.1 (-1.1, 7.3)
WEMWBS (age 42)	-0.3 (-1.2, 0.5)

Note: WEMWBS = Warwick-Edinburgh Mental Well-being Scale.

<sup>a</sup> Denotes 95% credible interval that excludes zero.

to eight percentage points. Given that the base prevalence of these outcomes ranges from 15% to 22%, an increase of seven to eight percentage points would represent an important effect. Overall, the credible intervals strongly suggest very small effects of young motherhood on the WEMWBS, while leaving open the (statistical) possibility of substantively meaningful detrimental effects on the Malaise outcomes.

### Sensitivity Analysis for Average Effects

The choice of cutoffs for “young” motherhood is admittedly arbitrary, although consistent with previous literature. To address this issue, we conducted a series of

**Table 3** Average effects of young motherhood using alternative cut points and comparison groups

Outcome Variable	ATE Point Estimate (95% credible interval)	
Alternative Cut Points	<18 vs. 18–30	<22 vs. 22–30
Malaise (age 30)	5.1 (−0.3, 10.4)	4.7 (1.8, 7.6) <sup>a</sup>
Malaise (age 34)	0.2 (−4.3, 4.6)	3.3 (−0.3, 6.9)
Malaise (age 42)	4.9 (−1.4, 11.1)	2.8 (−0.7, 6.3)
WEMWBS (age 42)	−0.1 (−1.3, 1.0)	−0.4 (−1.2, 0.4)
Alternative Comparison Group	<20 vs. 20–24	<20 vs. 25–30
Malaise (age 30)	2.0 (−1.9, 6.0)	5.0 (1.2, 8.8) <sup>a</sup>
Malaise (age 34)	1.9 (−2.5, 6.4)	4.5 (−0.2, 9.3)
Malaise (age 42)	2.3 (−2.2, 6.8)	4.1 (−0.6, 8.8)
WEMWBS (age 42)	−0.1 (−1.0, 0.7)	−0.8 (−1.9, 0.3)

Note: WEMWBS = Warwick-Edinburgh Mental Well-being Scale.

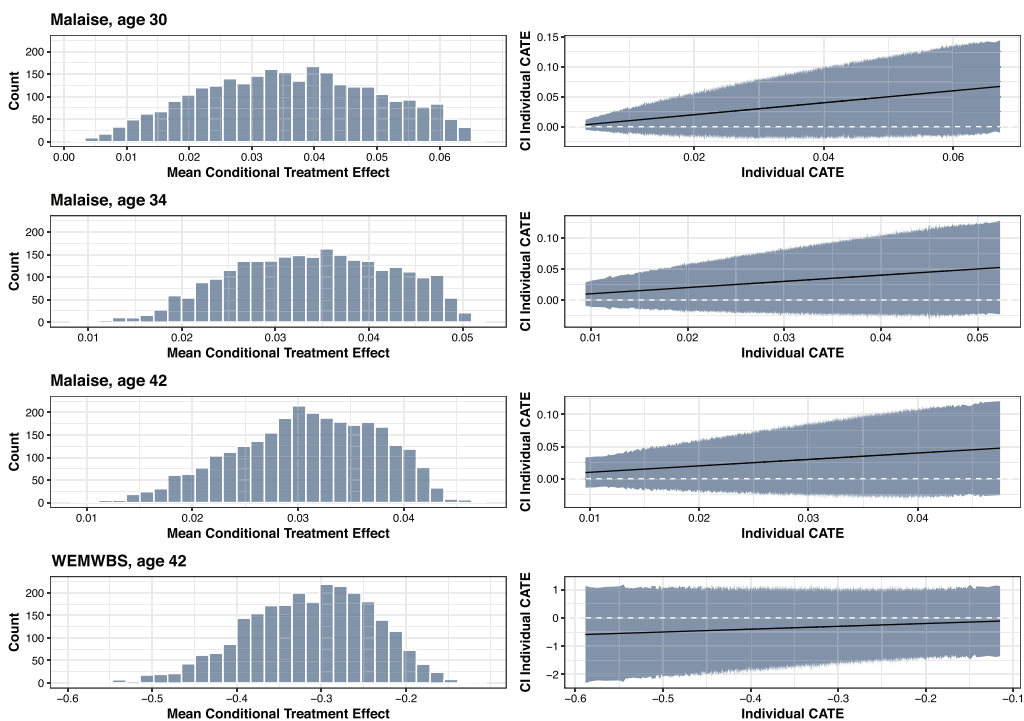
<sup>a</sup> Denotes 95% credible intervals that exclude zero.

sensitivity analyses. First, we varied the definition of “young motherhood” to be either 17 or younger (compared with 18–30) or 21 or younger (compared with 22–30). Results from this analysis are shown in [Table 3](#). In all cases, we arrive at conclusions that are substantively similar to those in the main analysis, although the credible intervals are notably wider for the <18 versus 18–30 comparison owing to the small number of first births occurring at age 18 or younger. The only notable difference in comparison to the analyses reported in [Table 2](#) is that the ATE for <18 versus 18–30 at age 30 overlaps 0 (although the point estimate is in fact larger than the corresponding value for the <22 versus 22–30 comparison).

Second, in the main analysis we assume that variation in ages at first birth within the “young” and “normative” groups is inconsequential for mental health. For the latter group, this encompasses a wide age range (20–30), during which time partnerships and human capital are typically established. We therefore conducted a series of additional analyses using more restrictive definitions of the “normative” comparison group, as either 20–24 or 25–30. The results from these analyses are also shown in [Table 3](#). In substantive terms, we find larger (roughly double) effects for the analyses of Malaise outcomes that use 25–30 as the comparison group relative to analyses that use 20–24 as the comparison group. The credible interval for the ATE for motherhood before age 20 versus motherhood at ages 25–30 on Malaise at age 30 excludes 0, whereas the corresponding credible interval for the <20 versus 20–24 comparison does not.

### Heterogeneous Effects

We next present evidence regarding possible *heterogeneity* in the effects of young motherhood across individuals. As discussed earlier, it is possible that small average effects may conceal variation across groups, with some experiencing more detrimental consequences (and some potentially positive effects). In practice, however, we find little evidence to support this contention. The left-hand panel of [Figure 3](#)



**Fig. 3** Individual CATE estimates by outcome. Average individual estimates are based on 100,000 posterior samples and include 95% credible intervals (CI). WEMWBS = Warwick-Edinburgh Mental Well-being Scale.

plots the distribution of the individual CATE estimate on the four different outcomes across sample members. Point estimates for the WEMWBS CATE are concentrated between roughly  $-0.5$  and  $-0.1$  and are centered close to the ATE estimate of  $-0.3$ . Substantively, the largest individual CATE ( $-0.6$ ) point estimate is equivalent to slightly less than 7% of a standard deviation—still practically very small. Moreover, the variance in the individual CATEs is dwarfed by comparison with the degree of statistical uncertainty associated with the estimates—95% credible intervals comfortably include 0 for all cases. Thus, our analysis suggests that, in addition to no average effect of young motherhood (compared with the counterfactual of motherhood for ages 20–30) on mental well-being at age 42, there are no *subgroups* in the data for whom a meaningful effect is likely to be present. We further investigated possible heterogeneity in effects in the form of common alternative estimands (Table 4), including the average treatment effect on the treated (ATT) and the average treatment effect on the controls (ATC). These estimates were substantively identical to those in the main analysis. Last, many recent studies—for example, Diaz and Fiel (2016)—investigate effect heterogeneity as a function of the propensity score, as advocated by Xie et al. (2012). Selection into motherhood on the propensity score may be important because teens anticipate potential consequences of childbearing, potentially resulting in “positive selection” whereby women who would be most affected are least likely to become teen mothers. We therefore calculated the ATE within quintiles of the

**Table 4** Average effects of young motherhood for young mothers and normative-age mothers

Outcome Variable	ATT (95% credible interval)	ATC (95% credible interval)
	Young Mothers	Normative-Age Mothers
Malaise (age 30)	4.3 (0.2, 8.4) <sup>a</sup>	3.5 (0.03, 6.9) <sup>a</sup>
Malaise (age 34)	3.9 (-1.0, 8.7)	3.3 (-0.9, 7.5)
Malaise (age 42)	3.3 (-1.2, 8.0)	3.0 (-1.1, 7.1)
WEMWBS (age 42)	-0.3 (-1.1, 0.5)	-0.3 (-1.2, 0.5)

Note: WEMWBS = Warwick-Edinburgh Mental Well-being Scale.

<sup>a</sup> Denotes 95% credible intervals that exclude zero.

propensity score, finding no variation in the magnitude of effects of young motherhood on WEMWBS scores (Table 5).

Individual CATE estimates for the Malaise outcomes vary between a 0- to 6-percentage-point increase in risk of psychological distress at age 30 and a 1- to 5-percentage-point increase at ages 34 and 42. While the magnitude of this variation appears substantively important, we do not believe this represents evidence of effect heterogeneity, for two reasons. First, the uncertainty of the estimates dwarfs the variation in the individual CATEs. Second, there is negligible heterogeneity in effects on the underlying probit scale. This means that the apparent variation in the magnitude of effects (expressed as percentage points) reflects variation in the underlying *marginal* probability of experiencing poor mental health, rather than the presence of interactions between young motherhood and covariates within  $f(A, \mathbf{X})$ . In practice, this means that young motherhood has larger effects *in percentage-point terms only* for women who experience an elevated risk of poor mental health as a function of other background characteristics. Subgroup analyses by young motherhood status (ATT/ATC) and by quintiles of the propensity score reflect this fact, with relatively larger percentage-point effects of young motherhood among young mothers, and among women who had a higher propensity to experience young motherhood. Otherwise, there are minimal substantive differences in the findings for these subgroups compared with the main analysis of Malaise outcomes.

If analysis had identified meaningful variation in effects across individuals, we would ordinarily proceed to explore subgroups whose mental health appears to be more or less strongly affected by young motherhood. Because our analysis found no evidence of individual effect heterogeneity for the WEMWBS outcome, nor (on the probit scale) for the Malaise outcomes, we did not proceed to this analysis. Rather, the lack of effect heterogeneity suggests that any effects of young motherhood on mental health are substantially *homogeneous*, at least as a function of the (extensive) set of background covariates included in our analysis.

## Discussion and Conclusion

A long history of research and public commentary links young motherhood with a host of negative outcomes for both mothers and their children (Mollborn 2017).



**Table 5** Average effects of young motherhood by propensity score quintiles

Outcome Variable	ATE Point Estimate (95% credible interval)
Malaise (age 30)	
1 (least likely to be teen mother)	2.2 (−0.2, 4.6)
2	3.2 (−0.1, 6.4)
3	3.7 (0.03, 7.3) <sup>a</sup>
4	4.3 (0.1, 8.4) <sup>a</sup>
5 (most likely to be teen mother)	5.0 (0.2, 9.9) <sup>a</sup>
Malaise (age 34)	
1 (least likely to be teen mother)	2.6 (−0.9, 6.1)
2	3.1 (−0.9, 7.2)
3	3.4 (−1.0, 7.9)
4	3.8 (−1.0, 8.6)
5 (most likely to be teen mother)	4.3 (−1.1, 9.7)
Malaise (age 42)	
1 (least likely to be teen mother)	2.5 (−1.0, 6.0)
2	2.9 (−1.1, 6.9)
3	3.2 (−1.1, 7.5)
4	3.3 (−1.2, 7.9)
5 (most likely to be teen mother)	3.6 (−1.3, 8.6)
WEMWBS (age 42)	
1 (least likely to be teen mother)	−0.4 (−1.3, 0.6)
2	−0.3 (−1.2, 0.6)
3	−0.3 (−1.2, 0.5)
4	−0.3 (−1.1, 0.5)
5 (most likely to be teen mother)	−0.3 (−1.1, 0.6)

Note: WEMWBS = Warwick-Edinburgh Mental Well-being Scale.

<sup>a</sup> Denotes 95% credible intervals that exclude zero.

However, as data and methods have become more sophisticated, it seems increasingly likely that *causal* effects of young motherhood are small or nonexistent, or in some cases confined to relatively advantaged segments of the population who are unlikely to experience young motherhood in any case (Diaz and Fiel 2016; Gorry 2019). Our results largely support the argument that poorer outcomes experienced by young mothers are primarily due to the high level of background disadvantage they experience rather than detrimental effects of young motherhood *per se*. In our primary analyses, the point estimate for the ATE was roughly 70% smaller than the bivariate differences for Malaise outcomes at ages 30, 34, and 42, and 85% smaller for the WEMWBS at age 42. Except for the Malaise outcome at age 30, 95% credible intervals for the ATE covered 0, meaning that in comparison to those who became mothers at 20 or older, our analyses provide only very limited evidence of harmful causal effects of young motherhood on later mental health. While our estimates indicate some support for an increased rate of mental health problems among young mothers at age 30, the effect is quite small and not distinguishable from 0 at older ages.

Our analysis does, however, provide some limited evidence that effects of young motherhood on mental health may depend on the counterfactual state that “young motherhood” is compared with, and the life stage of the woman. Specifically,

sensitivity analysis of Malaise outcomes at age 30 in which teen mothers were compared with those who became mothers at ages 25–30 (excluding mothers at 20–24) indicated a five-percentage-point increase in the risk of poor mental health. With the major caveat that causal interpretation of our estimates remains subject to the strong assumption that all relevant confounders have been accounted for, this suggests that there may be some benefit to delaying motherhood into the middle to late 20s in the short term. At older ages, estimated ATEs for the same contrast, although positive and only marginally smaller in magnitude, were not distinguishable from 0. Estimates for the contrast between teen motherhood and motherhood at 20–24 were uniformly not distinguishable from 0.

An important limitation of many common methods is that they provide only average effects and leave open the possibility that important variation in effects may exist between subgroups in the data. Effect heterogeneity of this kind has the potential to be both illuminating theoretically and important for policy (e.g., with respect to targeting interventions), and there is therefore considerable value in methods that can properly identify heterogeneous effects. In practice, our analyses found little evidence of effect heterogeneity, as seen in the largely homogeneous individual CATE estimates. We note that this represents an important finding: in addition to there being minimal evidence of sample-average effects of teen motherhood, our analysis further suggests that there are *no subgroups in which effects can be reliably identified*. The fact that our method allows us to arrive at a general conclusion of this nature contrasts with many alternative methods, which would permit only consideration of a limited number of prespecified subgroups and commonly require (unrealistic) prior knowledge of the correct functional forms. Because there are generally no reasonable grounds to believe that effects in social science are truly homogeneous, BART (and similar approaches such as “causal forests”; Wager and Athey 2018; see also Brand et al. 2021) has considerable potential as a tool for social science.

Our study has several implications for policy. With respect to efforts to *delay* motherhood, our findings suggest that policy would need to achieve relatively large changes in women's birth timing (on the order of six years at minimum) to realize any meaningful benefits to mental health. As the extant literature shows small or null effects of a range of teen pregnancy prevention strategies (Baxter et al. 2021; Marseille et al. 2018), it seems unlikely that intervention can achieve delays of this magnitude in practice. Long-term trends toward later parenthood are of this size, but it is likely that these shifts are driven by broader structural and cultural change rather than intervention or policy aimed at young parenthood *per se* (Lesthaeghe 2010; Ní Bhrolcháin and Beaujouan 2012). Consequently, it is unclear that there exist viable teen motherhood prevention strategies that would be reasonably expected to translate into better mental health outcomes. Furthermore, after age 30 we find no effects of teen motherhood that can be reliably distinguished from 0, suggesting that prevention efforts are unlikely to achieve substantial long-term gains in mental health. We note, however, that teen motherhood remains a strong *marker* of disadvantage, and that motherhood brings specific needs and constraints related to caring for children. There is therefore continuing potential for younger motherhood status to be used as a mechanism for targeting mental health support, particularly as teen mothers may be more engaged with health and social services during pregnancy and when children are young. Moreover, our finding of an increased rate of poor mental health among

teen mothers (compared to motherhood at 25–30) at age 30 suggests that there may be scope for intervention after birth to mitigate any negative effects on mental health. On this point, we stress the need for future work to consider more carefully the mechanisms associated with any detrimental effect of teen motherhood. Indeed, there is a tendency in much of the literature to equate negative effects of teen motherhood with some sort of deficit in the mother herself. This ignores the evidence that teen mothers are routinely stigmatized (McArthur and Winkworth 2018; Yardley 2008) and that stigmatization is strongly linked to poorer health outcomes (Hatzenbuehler et al. 2013). Thus, interventions aiming to change public *perceptions* of teen pregnancy—in lieu of targeting perceived deficits in the mother—may ameliorate any negative effects of teen pregnancy on mental health.

As with all analysis, our work is subject to limitations. First, the Malaise outcomes are geared toward clinical mental health and are consequently less sensitive to nonclinical variation in mental health. This may have limited our ability to identify subclinical mental health effects of young motherhood. The WEMWBS does, however, capture subclinical variation in mental well-being, and results were largely consistent with the Malaise outcomes. Second, many covariates of potential importance to young motherhood (particularly from the age 16 data collection) could not be included owing to high levels of missing data. While we were able to control for an extremely rich set of potential confounders in comparison with other studies, it is nevertheless likely that some residual confounding exists. This implies that our analysis cannot rule out the possibility that there are (unmeasured) subgroups of women for whom effects of teen motherhood are larger or smaller.

Our work builds upon a long history of studies that have investigated the consequences of teen motherhood (Mollborn 2017) and showcases the application of BART as a tool for the estimation of both common causal estimands as well as heterogeneous causal effects (Hill 2011). In most settings of interest to social scientists, there is no strong rationale to believe *a priori* that causal effects are truly homogeneous and no strong theory to guide model specification; BART addresses both issues, and we therefore suggest that there is considerable potential for social science to benefit from BART or similar methods. Substantively, our findings indicate that causal effects of teen motherhood on later mental health are likely to be both small and homogeneous for the cohort of women we studied. We note, however, that the absence of effects in this cohort (now middle-aged) does not rule out the possibility that such effects may arise for younger cohorts of women. In this light, we stress the need for both future research and policy aimed at preventing teen motherhood to tread lightly to avoid further reinforcing negative stereotypes or low expectations of young mothers, because to the extent that we perpetuate such stigma, we risk *creating* the problem we purport to solve. ■

**Acknowledgments** This research was supported by the Australia Research Council (ARC) Centre of Excellence for Children and Families Over the Life Course (CE140100027/CE20010025) and the Swedish Research Council for Health, Working Life and Welfare (FORTE, 2018-00861). The analysis uses data from the 1970 British Cohort Study, which is made available by the Centre for Longitudinal Studies (CLS), UCL Social Research Institute, and the UK Data Service. The views reported here are those of the authors and should not be attributed to the ARC, FORTE, CLS, or UK Data Service. The authors wish to acknowledge the valuable assistance provided by Sebastian Kalucza in preparing the data for analysis.



- Grundy, E. M. D., Read, S., & Väisänen, H. (2020). Fertility trajectories and later-life depression among parents in England. *Population Studies*, 74, 219–240.
- Güneş, P. M. (2016). The effects of teenage childbearing on long-term health in the U.S.: A twin-fixed-effects approach. *Review of Economics of the Household*, 14, 891–920.
- Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian Analysis*, 15, 965–997.
- Hainmueller, J., Mummolo, J., & Xu, Y. (2019). How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Political Analysis*, 27, 163–192.
- Hatzenbuehler, M. L., Phelan, J. C., & Link, B. G. (2013). Stigma as a fundamental cause of population health inequalities. *American Journal of Public Health*, 103, 813–821.
- Hill, J. L. (2011). Bayesian nonparametric modelling for causal inference. *Journal of Computational and Graphical Statistics*, 20, 217–240.
- Hillis, S. D., Anda, R. F., Dube, S. R., Felitti, V. J., Marchbanks, P. A., & Marks, J. S. (2004). The association between adverse childhood experiences and adolescent pregnancy, long-term psychosocial consequences, and fetal death. *Pediatrics*, 113, 320–327.
- 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.
- Hodgkinson, S., Beers, L., Southammakosane, C., & Lewin, A. (2014). Addressing the mental health needs of pregnant and parenting adolescents. *Pediatrics*, 133, 114–122.
- Jones, C., Whitfield, C., Seymour, J., & Hayter, M. (2019). ‘Other girls’: A qualitative exploration of teenage mothers’ views on teen pregnancy in contemporaries. *Sexuality & Culture*, 23, 760–773.
- Kalil, A., & Kunz, J. (2002). Teenage childbearing, marital status, and depressive symptoms in later life. *Child Development*, 72, 1748–1760.
- Kaluza, S. (2018). Mental health problems and social disadvantages as predictors of teenage parenthood: A register-based population study of Swedish boys and girls. *Longitudinal and Life Course Studies*, 9, 212–225.
- Kane, J. B., Morgan, S. P., Harris, K. M., & Guilkey, D. K. (2013). The educational consequences of teen childbearing. *Demography*, 50, 2129–2150.
- Kravdal, Ø., Grundy, E., & Skirbekk, V. (2017). Fertility history and use of antidepressant medication in late mid-life: A register-based analysis of Norwegian women and men. *Aging & Mental Health*, 21, 477–486.
- Larsen, B., & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience & Biobehavioral Reviews*, 94, 179–195.
- Lesthaeghe, R. (2010). The unfolding story of the second demographic transition. *Population and Development Review*, 36, 211–251.
- Marseille, E., Mirzazadeh, A., Biggs, M. A., Miller, A. P., Horvath, H., Lightfoot, M., . . . Kahn, J. G. (2018). Effectiveness of school-based teen pregnancy prevention programs in the USA: A systematic review and meta-analysis. *Prevention Science*, 19, 468–489.
- Maughan, B., & Lindelow, M. (1997). Secular change in psychosocial risks: The case of teenage motherhood. *Psychological Medicine*, 27, 1129–1144.
- McArthur, M., & Winkworth, G. (2018). Give them a break: How stigma impacts on younger mothers accessing early and supportive help in Australia. *British Journal of Social Work*, 48, 1277–1295.
- McCulloch, R., Sparapani, R., Spanbauer, C., Gramacy, R., & Pratola, M. (2021). *BART: Bayesian additive regression trees* (R package version 2.9) [Computer software]. The Comprehensive R Archive Network. Retrieved from <https://cran.r-project.org/package=BART>
- McMichael, C. (2013). Unplanned but not unwanted? Teen pregnancy and parenthood among young people with refugee backgrounds. *Journal of Youth Studies*, 16, 663–678.
- Mollborn, S. (2017). Teenage mothers today: What we know and how it matters. *Child Development Perspectives*, 11, 63–69.
- Mollborn, S., & Morningstar, E. (2009). Investigating the relationship between teenage childbearing and psychological distress using longitudinal evidence. *Journal of Health and Social Behaviour*, 50, 310–326.
- Nayak, A., & Kehily, M. J. (2014). ‘Chavs, chavettes and pramface girls’: Teenage mothers, marginalised young men and the management of stigma. *Journal of Youth Studies*, 17, 1330–1345.

- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*, 370–384.
- Ni Bhrolcháin, M., & Beaujouan, E. (2012). Fertility postponement is largely due to rising educational enrolment. *Population Studies*, *66*, 311–327.
- Nomaguchi, K., & Milkie, M. A. (2020). Parenthood and well-being: A decade in review. *Journal of Marriage and Family*, *82*, 198–223.
- Patel, P. H., & Sen, B. (2012). Teen motherhood and long-term health consequences. *Maternal and Child Health Journal*, *16*, 1063–1071.
- Pearlin, L. I. (1989). The sociological study of stress. *Journal of Health and Social Behaviour*, *30*, 241–256.
- Penman-Aguilar, A., Carter, M., Snead, M. C., & Kourtis, A. P. (2013). Socioeconomic disadvantage as a social determinant of teen childbearing in the U.S. *Public Health Reports*, *128*(Suppl. 1), 5–22.
- Pescosolido, B. A., & Martin, J. K. (2015). The stigma complex. *Annual Review of Sociology*, *41*, 87–116.
- Recto, P., & Champion, J. D. (2018). “We don’t want to be judged”: Perceptions about professional help and attitudes towards help-seeking among pregnant and postpartum Mexican-American adolescents. *Journal of Pediatric Nursing*, *42*, 111–117.
- Rodgers, B., Pickles, A., Power, C., Collishaw, S., & Maughan, B. (1999). Validity of the Malaise Inventory in general population samples. *Social Psychiatry and Psychiatric Epidemiology*, *34*, 333–341.
- Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health and behaviour*. London, UK: Longman.
- Sironi, M., Ploubidis, G. B., & Grundy, E. M. (2020). Fertility history and biomarkers using prospective data: Evidence from the 1958 National Child Development Study. *Demography*, *57*, 529–558.
- SmithBattle, L. I. (2013). Reducing the stigmatization of teen mothers. *MCN: The American Journal of Maternal/Child Nursing*, *38*, 235–241.
- Stewart-Brown, S., & Janmohamed, K. (2008). *Warwick-Edinburgh Mental Well-being Scale (WEMWBS): User guide, Version 1* (Report). Edinburgh, UK: NHS Health Scotland and University of Edinburgh; Coventry, UK: University of Warwick. Retrieved from <http://www.ocagingservicescollaborative.org/wp-content/uploads/2014/07/WEMWBS-User-Guide-Version-1-June-2008.pdf>
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., . . . Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, *5*, 63. <https://doi.org/10.1186/1477-7525-5-63>
- TNS BMRB. (2012). *Technical report of the 1970 British Cohort Study: Age 42 survey (2012 – 2013)*. Retrieved from [http://doc.ukdataservice.ac.uk/doc/7473/mrdoc/pdf/bcs70\\_2012\\_technical\\_report.pdf](http://doc.ukdataservice.ac.uk/doc/7473/mrdoc/pdf/bcs70_2012_technical_report.pdf)
- Tyler, I. (2008). “Chav mum chav scum”: Class disgust in contemporary Britain. *Feminist Media Studies*, *8*, 17–34.
- van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC Press.
- Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*, 1228–1242.
- Webbink, D., Martin, N. G., & Visscher, P. M. (2008). Does teenage childbearing increase smoking, drinking and body size? *Journal of Health Economics*, *27*, 888–903.
- Wendling, T., Jung, K., Callhan, A., Schuler, A., Shah, N. H., & Gallego, B. (2018). Comparing methods for estimation of heterogeneous treatment effects using observational data from health care databases. *Statistics in Medicine*, *37*, 3309–3324.
- Whitworth, T. R. (2017). Teen childbearing and depression: Do pregnancy attitudes matter? *Journal of Marriage and Family*, *79*, 390–404.
- Wiemann, C. M., Rickert, V. I., Berenson, A. B., & Volk, R. J. (2005). Are pregnant adolescents stigmatized by pregnancy? *Journal of Adolescent Health*, *36*, 352.e1–352.e7. <https://doi.org/10.1016/j.jadohealth.2004.06.006>
- Wolfe, J. D. (2009). Age at first birth and alcohol use. *Journal of Health and Social Behavior*, *50*, 395–409.
- Xavier, C., Benoit, A., & Brown, H. K. (2018). Teenage pregnancy and mental health beyond the postpartum period: A systematic review. *Journal of Epidemiology and Community Health*, *72*, 451–457.
- Xavier, C. G., Brown, H. K., & Benoit, A. C. (2017). Teenage pregnancy and long-term mental health outcomes among Indigenous women in Canada. *Archives of Women's Mental Health*, *21*, 333–340.
- Xie, Y., Brand, J. E., & Jann, B. (2012). Estimating heterogeneous treatment effects with observational data. *Sociological Methodology*, *42*, 314–347.
- Yardley, E. (2008). Teenage mothers’ experiences of stigma. *Journal of Youth Studies*, *11*, 671–684.

Martin O'Flaherty (corresponding author)

[m.oflaherty@uq.edu.au](mailto:m.oflaherty@uq.edu.au)

*O'Flaherty* • Institute for Social Science Research and Australian Research Council Centre of Excellence for Children and Families Over the Life Course, The University of Queensland, Brisbane, Australia; <https://orcid.org/0000-0003-0788-2598>

*Kalucza* • Department of Sociology and Centre for Demographic and Ageing Research (CEDAR), Umeå University, Umeå, Sweden; <https://orcid.org/0000-0003-2696-9517>

*Bon* • School of Mathematical Sciences and Centre for Data Science, Queensland University of Technology, Brisbane, Australia; <https://orcid.org/0000-0003-2313-2949>