Title

Semi-parametric graphical computation approach using loss-based estimation to estimate exposure effects: applications on infant developmental outcomes

Author

Aguilar Schall, Raul Eduardo

Publication Date

2010

Peer reviewed|Thesis/dissertation

Semi-parametric graphical computation approach using loss-based estimation to estimate exposure effects: applications on infant developmental outcomes.

by

Raul Eduardo Aguilar Schall

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Biostatistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Alan Hubbard, Chair
Professor Sandrine Dudoit
Professor Brenda Eskenazi

Fall 2010

Semi-parametric graphical computation approach using loss-based estimation to estimate exposure effects: applications on infant developmental outcomes.

Abstract

Semi-parametric graphical computation approach using loss-based estimation to estimate exposure effects: applications on infant developmental outcomes.

by

Raul Eduardo Aguilar Schall

Doctor of Philosophy in Biostatistics

University of California, Berkeley

Associate Professor Alan Hubbard, Chair

In epidemiology, researchers try to answer questions about exposures and their effects (associations) on a variety of outcomes of interest. Most times, the collected data comes from observational studies, meaning that the researcher did not control the exposure to which each subject under study was exposed, like it is done in clinical trials. Additionally, researches collect information on other variables which could act as potential confounders of exposure. Estimation of adjusted associations under these conditions, if not reliant on arbitrary and thus biased parametric models, suffers from the curse of dimensionality. This dissertation describes semi-parametric statistical approaches to address the correct estimation of the parameter of interest using targeted maximum likelihood estimation (TMLE) methodology, which optimally adapts estimates of the data-generating distribution for estimation of the association of interest. The process optimally relies on machine learning techniques and is a modification of the likelihood-based algorithm where the parameter is defined by the so-called G-computation formula.

Chapter 2 provides the estimation of direct effects, adjusting for the possible indirect effects through intermediate variables. TMLE is used, with the help of model selection using the SuperLearner algorithm,[74] to obtain estimators for the direct effect. General methods on how to estimate the natural and controlled direct effects using TMLE controlling for the intermediate variables are implemented. These techniques are then used to examine the direct effect of maternal depression on cognitive and language development in 350 Mexican-American children in the CHAMACOS birth cohort study. Children of mothers with depressive symptoms scored significantly lower (-2.82 (p-value < 0.05) points in the Preschool Language Scale) on the expressive communication compared to those of non-depressed mothers after controlling for the intermediate effects of home environment and breastfeeding duration. Depression did not show a significant direct effect on auditory comprehension, mental, or psychomotor scores.

Chapter 3 present the use of TMLE and machine learning to estimate effects of organophosphate (OP) pesticides during infant stages of child growth. Many papers have been published about the adverse effects of *in utero* pesticide exposure and the effects on fetal growth.[19, 18] All the previous literature has used traditional analyses, while we implement a TMLE approach. The goal is to obtain estimates of the effects of exposure to OP pesticides not only *in utero* but later, when the child is exposed directly and how this affects its physical growth at different ages: 6, 12, 24 months, 3.5, and 5 years. Pesticides are widely used in the Salinas Valley, CA where the population under study resides. We identify several statistically significant negative effects of the exposure to OP pesticides on child's growth.

Chapter 4 presents the longitudinal analysis of the intervention effect through the use of machine learning techniques and G-computation, as well as TMLE. There are no available studies about the longitudinal effects of organophosphate (OP) pesticides on child growth measured by child weight. This is a first attempt to estimate the effects of continuous exposure to OP pesticides in children living in the agricultural region of the Salinas Valley, CA. Without a control group, we estimated the effects of an intervention where exposure was controlled and fixed to the lowest level possible, and compared the estimated child weights from this scenario with the actual weights at 3.5 years of age. We used an *ad hoc*, but still double robust, targeting step on the outcome distribution estimate in conjunction with simulation based on the G-computation formula. Our results show a negative effect of OP pesticide exposure on the mean child weight at 3.5 years, however none of them reached significance.

Chapter 5 concludes with a summary of the preceding chapters and a discussion of future research directions.

To Eva and Matthias.

Contents

List of Figures

List of Tables

Acknowledgments

First of all I would like to thank the Consejo Nacional de Ciencia y Tecnología (CONACYT) in Mexico and UC Mexus for having sponsored me for five years. Without their financial support, this long journey would not have even started.

I want to thank my advisor, Alan Hubbard, whose mentorship and help have been invaluable in the elaboration of this dissertation. I particularly appreciate his patience when my son Matthias was born and Dr. Hubbard granted me the time to enjoy my son's first year of life. I also appreciate the much needed encouragement to return to my work when I was finally ready to complete it.

I would also like to thank everyone in the Biostatistics Group for their valuable lessons and contributions to my intellectual growth: Professor Sandrine Dudoit, who showed me the wonders and complications of computational biology for the first time and who also accepted my request that she serve as a member of my dissertation committee; Professor Nick Jewell, who asked us to call him Nick the first class, because a PhD does not change who you are; Professor Steve Selvin, who introduced me to R which I would never stop using; and Professor Mark van der Laan, whose classes were conceptually rich and very challenging.

I also want to thank the Center for the Children's Environmental Health Research, or as I like to refer to it, Eskenazi Research Group. In particular, I am thankful to Professor Brenda Eskenazi for providing context and perspective to my statistical research, and who also honored me by serving as a member of my dissertation committee. I would also like to thank Kim Harley, with whom it has been a privilege to work in the Center for the Health Assessment of the Mothers and Children of Salinas (CHAMACOS) Study.

There have been so many people who helped me through all these years, making it an unforgettable experience: Bonnie Hutchings, who gave me the good news that I had been accepted into the program and helped me complete the paperwork for my sponsor; Sharon Norris, who has continued to provide support through all of the administrative hurdles students encounter; and Burke Bundy, who was always a great help with any computer questions.

I want to thank all those who I did not know before Berkeley and now are among my closest friends: Alejandro de la Fuente, Armando Solar, and Raúl Howe. Your friendship has been a key factor in the beauty and complexity of my whole experience as a PhD student.

I want to thank my fellow doctoral students, with whom I shared disappointments, laughter and a beer from time to time, in particular Kelly Moore.

I thank the infinite support of my family, which has been invaluable for the completion of this dissertation. I thank my parents, Irene Schall and Raúl Aguilar, for their endless love and words of encouragement. I thank my aunt Delia for always having been there for us. Finally, I thank my wife Eva and my son Matthias, for their patience in this long journey, for her confidence in my ability to achieve my goals and for their unconditional love.

Chapter 1

Introduction

In epidemiology, researchers try to answer questions about exposures and their effects (associations) on a variety of outcomes of interest. Most times the collected data comes from observational studies. Thus the researcher cannot control the exposure to which each subject under study was exposed, which introduces confounding. It has been a concern, for several years now, that many of the published research findings are false.[33] This problem could be due to a number of different reason. Perhaps the most important of them being bias, where a combination of design, data, analysis, and presentation factors produce findings where there should be none. There is strong evidence that selective analysis bias reporting, is a common problem even for randomized trials.[14] Another problem is that modern epidemiology is increasingly trying to find smaller effect sizes, with mode complex and often messy data.[69] The problem is that over the past 50 years, epidemiologists have succeeded in identifying the most obvious determinants of non-infectious diseases. Smoking, the greatest culprit, can increase the risk of developing lung cancer by as much as 3000%. Now there are only more subtle associations to look for between environmental causes and disease or health effects, and thus the challenge to tease out "causa" associations from a large list of potential culprits.

Causal effects are generally defined on the basis of counterfactual outcomes, i.e., outcomes that would have been observed on a subject had the exposure, possibly contrary to fact, been set at a particular level. However, in an actual study, we only observe a single counterfactual outcome for each subject, $Y = Y_A$ corresponding to the exposure that the subject actually received, say for example, the outcome Y_1 is not observed for subjects that were not exposed ($A = 0$), where A is the exposure. Since the counterfactual outcome is missing for these subjects, we cannot directly estimate $E(Y_1)$. From the observed data, $E(Y \mid A = 1)$ can be estimated, that is the mean of the counterfactual outcome for exposed subjects ($a = 1$) among those subjects who were actually observed at that exposure level ($A = 1$). When the group under study is truly a random sample of the population, $E(Y \mid A = 1)$ is equal to $E(Y_1)$ since the group of subjects in the observed sample with $A = 1$ is indeed representative of the entire study population.

Now, suppose we are interested in the marginal effect of A on Y and define the parameter of interest as $\psi = E(Y_1) - E(Y_0)$. Under the assumption that we have a random sample from the population under study, the natural estimate of the parameter of interest if given by,

$$\widehat{\psi} = \widehat{\mu}_1 - \widehat{\mu}_0$$

where $\widehat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n} I(A_i = 1)Y_i$ and $\widehat{\mu}_0 = \frac{1}{n_0}\sum_{i=1}^{n} I(A_i = 0)Y_i$, where n_1 is the number of exposed subjects, n_0 is the number of unexposed subjects, and $n = n_1 + n_0$. Consider the conditional expectation of the outcome given treatment and covariates W, denoted by $Q(A, W) = E(Y \mid A, W)$. This function can be estimated with a linear regression model such as $\widehat{Q}(A, W) = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 W$ In this setting, $\widehat{\beta}_1$ coincides with the unadjusted estimate $\widehat{\psi}$. However, when $Q(A, W)$ is estimated as

$$\widehat{Q}(A, W) = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 W + \widehat{\beta}_3 AW,$$

then $\widehat{\beta}_1$ no longer coincides with $\widehat{\psi}_1$. In this case, to obtain the marginal effect, one must integrate out or average over the covariate(s) W. The G-computation estimator, introduced by Robins,[61, 62] is a maximum likelihood plug-in estimator that does provide an adjusted marginal effect,

$$\widehat{\psi}_{Gcomp} = \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(1, W_i) - \widehat{Q}(0, W_i)\right]$$

When $\widehat{Q}(A, W)$ is estimated with a linear model, and it does not contain any interaction terms, then $\widehat{\psi}_{Gcomp} = \widehat{\beta}_1$. The G-computation estimator is not limited to a linear model for $Q(A, W)$ when estimating the exposure effect, for example, when the outcome is binary, one could use a logistic regression model to estimate $Q(A, W)$ and use the G-computation formula to obtain the estimated risk difference.[43] The goal here is to not depend on any restrictions of the model for Q, and thus provide an adjusted marginal association that respects what we typically truly know about the data-generating distribution: almost nothing.

This dissertation proposes using targeted maximum likelihood estimation (TMLE), originally introduced in van der Laan and Rubin (2006).[76] This estimation procedure is a new approach to statistical learning that can be applied to many estimation problems. In short, TMLE is an estimation procedure that carries out a bias reduction specifically targeted for the parameter of interest. This is in contrast to traditional maximum likelihood G-computation estimation which aims for a bias variance trade-off for the whole density of the observed data, rather than a specific parameter of it. TMLE is a type of likelihood based estimator which provides an estimate that has the so-called efficient influence curve. Due to this latter fact, it thereby inherits the properties of the solution of the efficient influence curve estimating equation, including asymptotic linearity and local efficiency.[75] The ad-

vantages of this methodology over traditional methodologies are discussed for each of the applications in each chapter.

Chapter 2 of this dissertation present the framework for the proposed techniques in the context of direct and direct effects and uses it to explore the effect of maternal depression, measure using the Center for Epidemiological Studies Depression Scale (CES-D) [59] on infant neurodevelopment, evaluated using the Bayley Scales of Infant Development, Second Edition (BSID) [6] and the Pre-School Language Scale, Third Edition (PLS).[78]

Chapters 3 and 4 explore the effects of exposure to organophosphate (OP) pesticides on infant growth. The main difference between the two chapters is the approach to the search of potential effects. In Chapter 3 we perform an extensive series of cross-sectional analyses where exposure is measured in the mother and in the child, and the outcome is evaluated through four measures of infant growth - weight, length, body mass index (BMI), and waist circumference at five different ages - 6, 12, and 24 months, 3.5 and 5 years. In Chapter 4, the analysis is longitudinal and the outcome is reduced to the child weight at age 3.5 years considering a series of three OP pesticide exposures in the child.

In summary, this dissertation provides a new approach to exposure analysis through the application of targeted maximum likelihood estimation with extensive use of machine learning techniques. Each of the chapters provides concrete examples of the potential of the proposed methods to reduce bias in the search of potential effects from exposures of concern for society. In Chapter 5, the dissertation concludes with a summary of the finding from each of the applied analyses of the preceding chapters and a possible directions for future research.

Chapter 2

A general graphical computation approach using loss-based estimation to estimate direct effects: an application to maternal depression on infant neurodevelopment

2.1 Introduction

Many times the exposure of interest not only affects the outcome directly, but also through intermediate pathways. In this paper, we combine machine learning tools within a G-computation estimation framework to provide an easy-to-implement method to obtain direct effect estimations.

In particular, we will analyze the effect of postpartum maternal depression on infant neurodevelopment at 12 months postpartum. In our analysis, we measure two potential intermediate pathways through which maternal depression could affect infant neurodevelopment: 1) the Home Observation for the Measurement of the Environment (HOME) [12] and 2) breastfeeding duration. There is evidence that maternal depression negatively affects the HOME-scale reducing its score and also shortens the duration of breastfeeding. Additionally, there is evidence that a lower score in the HOME-scale as well as a shorter breastfeeding period is negatively associated with infant neurodevelopment. [2, 70, 56]

The paper is presented in the following way: first we present the background of the estimation of direct effects and review the existing estimators of direct and causal effects, their necessary conditions and assumptions. Then, we briefly describe the use of machine

learning as a semi parametric approach for model selection. In this context, we review the Super Learner procedure which combines models from different candidates (machine learning algorithms). In Section 2.3 we present the data structure and two detailed algorithms to perform the calculations for the estimation of direct effects. Finally, in Section 2.4 we conduct a complete data analysis of the effects of maternal depression on infant neurodevelopment. We present the results of these analyses and then, in the last section, we discuss the advantages and limitations of the proposed methods.

2.2 Background

2.2.1 Counterfactual framework

The counterfactual framework was first introduced by Neyman [49] and further developed by Rubin [65] and Robins [61, 62]. Suppose we observer n independent and identically distributed observations of the random vector $O = (Y, A, W) \sim p_0$, where Y is the outcome of interest, A is the exposure of interest, W is a vector of baseline covariates, and p_0 denotes the density of O. Direct effects are based on a hypothetical full data structure $X = ((Y_{az} : a\epsilon\mathcal{A}), W)$ containing the entire collection of counterfactual or potential outcomes Y_a for all possible treatments A. The observed data structure O only contains a single counterfactual outcome $Y = Y_A$ corresponding to the actual exposure of the subject. The observed data $O = (Y \equiv Y_A, A, W)$ is thus a missing data structure. The randomization assumption or coarsening at random assumption states that A is conditionally independent of the full data X given W; $P(A \mid X) = P(A \mid W)$.

Another related framework that is particularly helpful when estimating effects, direct and indirect, as well as causal structures, is the graphical representation of the problem in a directed acyclic graph (DAG). [50] Figure 2.1 represents a simple case in which the exposure A has direct and indirect effects on the outcome Y. The direct effect which is usually the question of research is represented by the solid arrow going out from A into Y. In the same DAG we see that there is another way going from A through Z to Y ; the indirect effect. It shown by the pair of dashed arrows going out of A into Z and out of Z into Y.

2.2.2 Direct effect

The standard approach in estimating effects of exposure on outcome is to run a multivariate regression of the outcome on the exposure and a group of believed relevant covariates. In the presence of intermediate variables, the multivariate regression is only adequate to estimate the direct effect if strong assumptions are made. Regardless, several minimal assump-

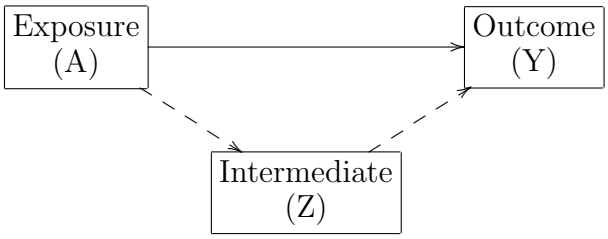


Figure 2.1: Simple directed acyclic graph (DAG) representing the direct (solid arrow) and indirect effects (dashed arrows) of exposure A on outcome Y. The indirect effect goes through the intermediate variable Z.

tions need to be made additionally to the main and standard assumption of no unmeasured confounders. It means that all covariates that affect the exposure and the outcome have been recorded and are in the model. To obtain a consistent estimator of the direct effect we also need to assume that there are no unmeasured confounders between the intermediate variable and the outcome. In Figure 2.2 these two assumptions are represented by the absence of unmeasured confounding between exposure and outcome, $U1$, and between the intermediate variable and the outcome, $U2$.

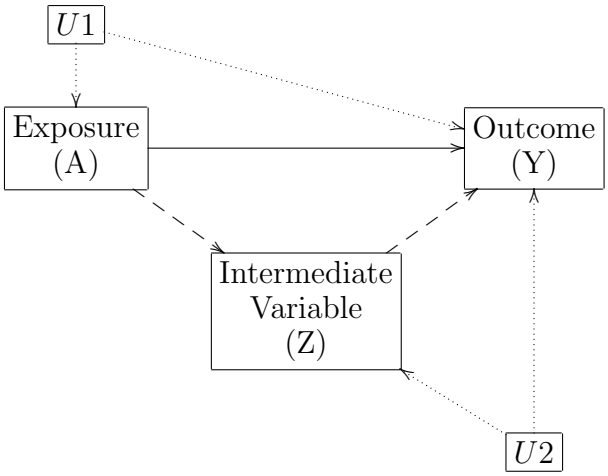


Figure 2.2: Causal diagram for main effect A on outcome Y with direct effect and indirect effect through the intermediate variable Z. No unmeasured confounders between main exposure nor intermediate variable and outcome, U1 and U2, respectively.

Petersen and van der Laan [53] present in detail the differences and assumptions required to estimate the different direct effects. They also summarize the work of Robins and Greenland [63] and Pearl [51]. In addition to the basic assumptions already stated, the different

authors make additional assumptions that are presented in the order they were published.

First, Robins and Greenland [63] show identifiability of direct effects, this is that one can measure it, assuming no interaction between the exposure and the intermediate variable at the individual level. Second, Pearl [51] proposes an alternative identifying assumption which states that within subgroups defined by the baseline covariates, the individual's outcome under a fixed level of exposure and intermediate variable is independent of the intermediate counterfactual's (or no exposure) level. This assumption can be formulated as:

$$Y_{az} \perp Z_0 \mid W \tag{2.1}$$

Third, Petersen and van der Laan's [53] show identifiability if no unmeasured confounding between the exposure variable and the intermediate variable. All three assumptions of no unmeasured confounders can be presented as the following conditions:

$$A \perp Y_{az} \mid W \tag{2.2}$$

$$Z \perp Y_{az} \mid A, W \tag{2.3}$$

$$A \perp Z_a \mid W \tag{2.4}$$

They further assume that within subgroups defined by covariates included in the multivariate regression model, the level of the intermediate variable in the absence of exposure is not informative about the expected magnitude of the exposure's effect at a fixed level of the intermediate variable. They refer to this as the "direct effect assumption", which can be formally stated as:

$$E[Y_{az} - Y_{0z} \mid Z_0 = z, W] = E[Y_{az} - Y_{0z} \mid W] \tag{2.5}$$

Two possible parameters measuring direct effect are:

- Natural Direct Effect: is equivalent to the counterfactual effect of exposure $A = a$ on outcome Y when the intermediate variable is set at a counterfactual value Z_0 that would have been observed had the individual been exposed to $A = 0$.

$$NDE = E[Y_{aZ_0} - Y_{0Z_0}] \tag{2.6}$$

- Controlled Direct Effect: is equivalent to the counterfactual effect of exposure $A = a$ on outcome Y when the intermediate variable Z is held constant at a level specified by the investigator $Z = z$ - treating az as a joint treatment.

$$CDE = E[Y_{az} - Y_{0z}] \tag{2.7}$$

Petersen and van der Laan use a different notation, calling the former Type 2 direct effect and the latter Type 1 direct effect. We use the Natural/Controlled nomenclature because we believe it is more intuitive to distinguish them.

In the fields of psychology and psychiatry there is a similar construction that tries to explain the direct effect of an exposure or stimulus on a response variable. In those fields, however, they refer to intermediate variables as mediators. The most cited paper is the one by Baron.[3] They also talk about moderators, third variables that affect the relationship between an exposure X and an outcome Y but in a different way. This literature often contrasts the definition of mediators and moderators. Consider the variable Z which does have an effect over the relationship between X and Y either as 1) a confounding variable when it causes both the exposure and the outcome, 2) a covariate when information about it improves the prediction of Y by X but does not substantially alter their original relationship, and 3) a moderator or interaction effect, when it modifies the relation of X to Y such that it depends of the value of Z.[41]

Estimators

There are two problems with traditional approaches when defining direct effect estimators. The first one is that they assume arbitrary models that may or not be correct, and additionally cannot be verified for their validity. The second problem is that they fail to explicitly define a parameter separate from the arbitrarily chosen regression model. We present in developing order four different classes of estimators that can be used to calculate direct effects . For simplicity we present estimators corresponding to the mean counterfactual outcome $\psi = E[Y_{a,z}]$.

The first class is known as G-computation estimation and was developed by Robins. [61, 62] The G-computation estimator of ψ, under the basic assumptions, can be identified by the observed data as

$$\psi(a) = E\left[Y_{a,z}\right] = E_W\Big[E[Y \mid A = a, Z = z, W]\Big] \tag{2.8}$$

A substitution estimator is used; based on the marginal distribution of W, $P(W)$, and the conditional distribution of Y given A, Z and W, $P(Y \mid A, Z, W)$. The first distribution can be estimated non-parametrically by the empirical distribution of W in the sample. Estimation of $P(Y \mid A, Z, W)$ will usually require specification of a parametric model or a large sieve. An estimate Q_n of the regression $Q(A, Z, W) = E[Y \mid A, Z, W]$ on an appropriate model defines the desired conditional distribution. The corresponding substitution estimator for ψ

is given then by

$$\psi_n^{G-comp}(a) = \frac{1}{n}\sum_{i=1}^{n} Q_n(a, z, W_i) \tag{2.9}$$

The second and third classes of estimators are closely related. They are known as inverse probability of treatment weighting ($IPTW$), and double robust $IPTW$ ($DR-IPTW$). These are based on the general estimating function methodology described in van der Laan and Robins. [75] We will describe the $IPTW$ estimator first and then explain how to build the DR version of it. Recalling the DAG framework, we are interested in the arrow that goes directly from A into Y, but we want to exclude the effects (arrows) of the covariates, confounders and/or intermediated variables. All these work against our effort of disentangling the true effect of A on Y. Using the counterfactual framework, we would like to know the outcome of each individual at all possible levels of exposure, $\{Y_{a,z} : a\epsilon\mathcal{A}, z\epsilon\mathcal{Z}\}$. In the case of a binary exposure this simplifies to two levels, $\{Y_a : a\epsilon\{0,1\}\}$. Given the lack of a real counterfactual full data set we estimate the probability that each individual received the exposure that he did. We perform a weighted analysis of the exposure on the outcome. The weights are defined by the inverse of a new function $g(.)$ which explains the observed exposures through the observed covariates $g(A, Z \mid W) \equiv P(A, Z \mid W)$. Most times this weights are stabilized by using the empirical distribution of the exposure in the numerator, $P(A, Z)$. The conditional distribution of the exposure given the covariates, $P(A, Z \mid W)$, can vary greatly when W and (A, Z) are strongly associated, creating also extremely large weights for a few subjects in the population under study. [64] We omit the details of the estimating function and the solution of the corresponding estimating equation; these can be reviewed in Bembom and van der Laan. [7] The IPTW estimator of ψ is

$$\psi_n^{IPTW}(a) = \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i = a, Z_i = z)}{g_n(A_i, Z_i \mid W_i)} Y_i \tag{2.10}$$

Additionally to the basic assumption, the model for g_n is assumed to be correctly specified. The IPTW estimator has to fulfill two additional assumptions in order to be an unbiased estimator. First, the expectation of the estimating function has to be well defined, $E[D^{IPTW}(O \mid g, \psi)] < \infty$. Second, there is no deterministic assignment of $A = a$ for given covariates; this is known as the "Experimental Treatment Assignment" (ETA) assumption. In case of violating this last assumption, $g(A \mid W)$ will not provide a consistent estimate even if it is correctly specified. [47]

The double robust version of the IPTW estimator, DR-IPTW, is robust against models' misspecifications. For the DR-IPTW estimator we need to specify two models, $Q_0(Y \mid A, Z, W)$ and $g_0(A, Z \mid W)$, but only one of them needs to be correctly specified to obtain a consistent estimate of our parameter of interest when the ETA assumption holds. [47] This means that in case of a violation of the ETA assumption, we can still obtain a consistent

estimate of our parameter as long as $Q_0(Y \mid A, Z, W)$ is correctly specified. This robustness comes with a high price of a more complex estimating equations and resulting estimator:

$$\psi_n^{DR}(a) = \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i = a, Z_i = z)}{g_n(A_i, Z_i \mid W_i)} \left[Y_i - Q_n(A_i, Z_i, W_i)\right] + Q_n(a, z, W_i). \tag{2.11}$$

The forth class of estimators is targeted Maximum Likelihood Estimation (TMLE), which is a sort of marriage of the G-comp and DR estimators. It carries out a targeted bias reduction specifically for the parameter of interest and not for the whole density of the observed data; hence its name. Consider a model $\mathcal{M}$ where the true distribution of the data is p_0. Consider and initial estimator $\widehat{p}$ of p_0. The parameter of interest is given by $\psi = \psi(p_0)$. TMLE has two goals. First, it aims to find an optimal density $\widehat{p}^* \epsilon \mathcal{M}$ that solves the efficient influence curve for the estimating equation for the parameter of interest. This results in a bias reduction compared with the maximum likelihood estimate of $\psi(\widehat{p})$. Second, the algorithm also requires that $\widehat{p}^*$ achieves an increase in the log-likelihood relative to $\widehat{p}$. The resulting substitution estimator $\psi(\widehat{p}^*)$ is a familiar type of likelihood-based estimator and due to the fact that it solves the efficient influence curve estimating equation it thereby inherits its properties including asymptotic linearity, and local efficiency.[44] For complete technical and theoretical details about this general estimating approach we refer readers to the seminal paper by van der Laan and Rubin.[76]

For a basic and applied introduction to the TMLE procedure we refer the interested reader to the "Gentle Introduction" by Gruber. [25] We now present the basic steps involved in TMLE:

1. Estimate the conditional expectation of Y given A, Z and W; denoted by $Q_n^0(A, Z, W)$.

2. Estimate the conditional distribution of the exposure given covariates; denoted by $g_n^0(A, Z \mid W)$

3. Calculate a specific covariate for each individual based on the subject's observed values A, Z, W and the estimate $g_n^0(A, Z \mid W)$. This new covariate, whose form depends on the parameter of interest and the model of $Y \mid A, Z, W$, is denoted by $h(A, Z, W)$ and sometimes referred to as a "clever" covariate.

4. Update the initial regression $Q_n^0(A, W)$ by adding the clever covariate $h(A, W)$ and estimating the corresponding coefficient by simple maximum likelihood, holding the remaining coefficient estimates fixed at their initial values. This can be done with an offset equal to $m_n^0(A, Z, W)$. The updated regression is denoted $\gamma(Q^1) = \gamma(Q^0 + \epsilon h)$, where γ is the link function.

5. Evaluate the updated regression at the exposure levels of interest, $A = a$. Taking the mean across the population you obtain a TMLE of the mean outcome at the specified exposure level.

The parameter of interest is not restricted to the difference between exposed and unexposed. It could easily be defined as the relative difference (ratio) or even odds ratio. Hence, next we only present the key component of the TMLE:

$$\psi_n^{TMLE}(a) = \frac{1}{n}\sum_{i=1}^{n} Q_n^1(a, z, W_i). \tag{2.12}$$

Like the DR-IPTW estimator, the TMLE is consistent if at least one of the two functions we have to estimate, g and Q, is correctly specified. Additionally, the estimator is locally efficient in the sense that it is efficient if both models are estimated consistently.

2.2.3 Machine learning

We need to estimate models that predict the outcome Y given the exposure A, the intermediate variable Z, and a set of covariates W, $P(Y \mid A, Z, W)$, as well as models to predict the exposure A on the set of covariates W, $P(A \mid W)$. We mentioned that the different estimation procedures rely on their models being correctly specified. Therefore, in order to avoid a misspecified model as much as possible, we will not use parametric models specified *a priori*. We will use a data-adaptive model selection approach, a machine learning approach, which has the property of potentially approaching a non-parametric model as $n \to \infty$. There are many such procedures available, like the Deletion/Substitution/Addition (D/S/A) algorithm, Least Angle Regression [16], Random Forest [11], Support Vector Machine, Generalized Additive Models (GAM), and Polychotomous Regression. Each one of the procedures listed here returns an "optimal" model, not all of them identical, based on their own optimality criteria.

For example, the D/S/A algorithm performs data-adaptive estimation through selection of the estimators based on heavy use of cross-validation and the $L2-loss$ function. The candidate estimators will always be polynomials which comply to user-specified constrains like maximal number of terms in the polynomial, maximum power of polynomial terms and maximum order of interaction.[68] Another example is GAM, which replaces the linear form for the covariates ($\sum \beta_i X_i$) by a sum of smooth functions ($\sum s_i(X_i)$), where the functions $s_i(\cdot)$ are unspecified and are estimated using an iterative procedure called local scoring algorithm, resulting in a nonparametric regression method.[29] A third example is Polychotomous Regression, which fits a regression model using linear splines and their tensor products.[21, 35]

Super Learner

Above we mentioned a list of model selctors which we'll refer to as "candidate learners". We can use them to define the models we need for our estimations. Van der Laan et al.[74] propose an algorithm to select the best possible model using a convex combination of the resulting models from each of the candidate learners. The resulting model will be as good or better than every single model from the candidate learners.

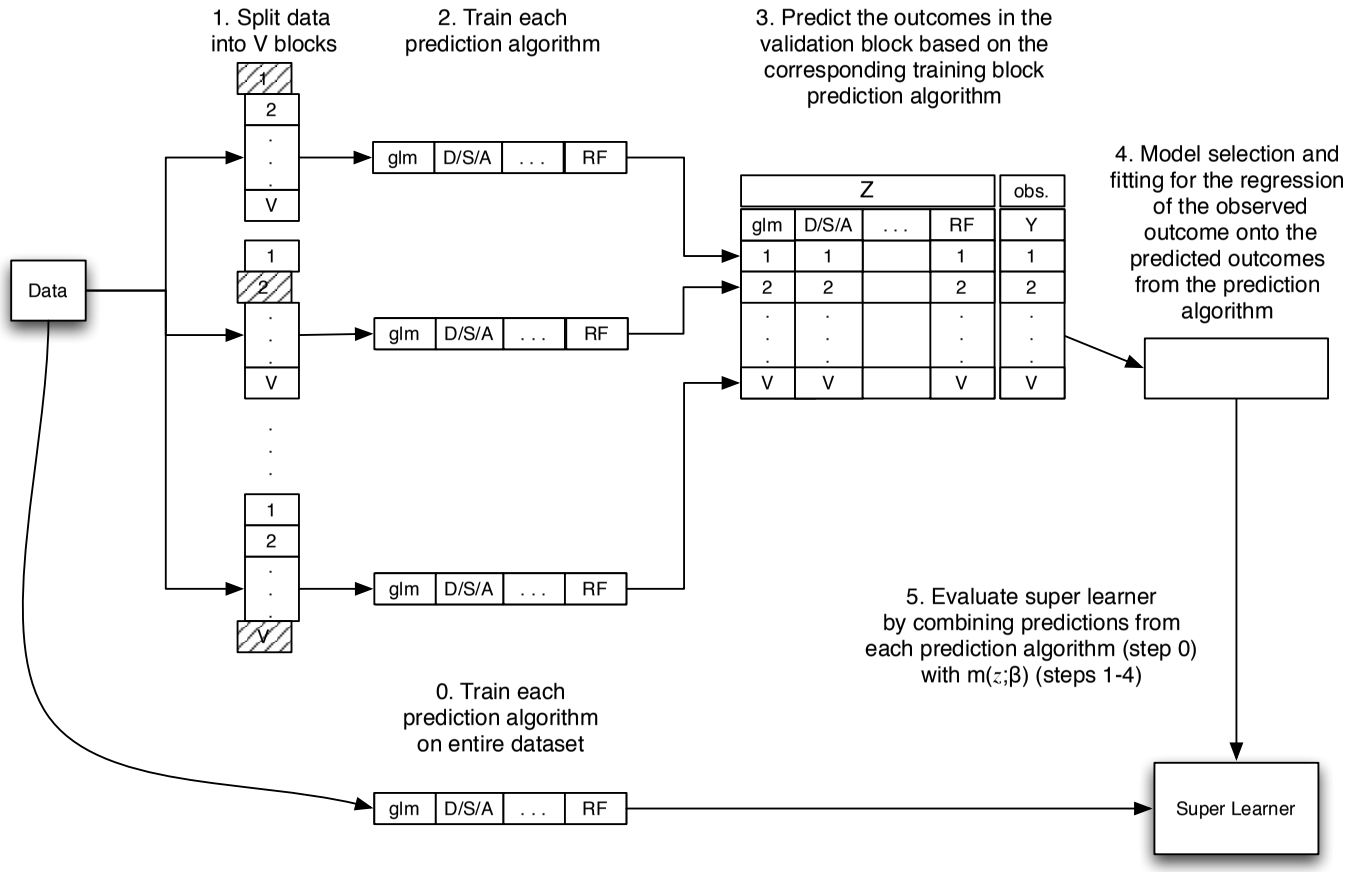


Figure 2.3: We thank Eric Polley for providing us this diagram of the Super Learner algorithm; first presented in the original paper. [74]

2.3 Methods

Consider the observed data given by n i.i.d. copies of $O = (Y, A, Z, W)$, where Y is the outcome of interest, A is a binary exposure of interest, Z is the set of intermediate variables on the causal pathway from A to Y, and W denotes the collection of measured confounders.

We intend to estimate the natural direct effect of A on Y, $E(Y_{aZ_0} - Y_{0Z_0})$, as well as the controlled direct effect of A on Y, $E(Y_{az} - Y_{0z})$. Remember from the previous definition of these two types of direct effect that the first one allows for the intermediate variable to fluctuate at the counterfactual level of no exposure, $Z_0 = Z(A = 0)$. The intermediate variable's level will assume the value it would have shown had there been no exposure, $A = 0$. On the other hand, the controlled direct effect will set the intermediate variable at a fix level, $Z = z$, and then perform the calculations of the effect of the exposure on the outcome at the preset level for the intermediate variable. In the controlled case we consider the intermediate variable to be binary for simplicity of calculations. If there is more than one intermediate variable and we need to consider all the possible level-combinations at which the n intermediate variables can be fixed, the total number of possibilities is 2^n. In the data analysis section we estimate both direct effects, natural and controlled, using the continuous and categorical (binary) intermediate variables, respectively.

The basic assumption we make is the one of no unmeasured confounders between any of the relations of A, Z, and Y (Figure 2.4) as defined by Petersen and van der Laan.[53] For the natural direct effect we also assume that the model we propose for $P(Y \mid A, Z, W)$ is correctly specified and will return consistent estimates; hence the need of machine learning procedures to define the model.

Finally, we will have the consistency assumption for both direct effect estimators, which states that the observed data is a missing data case from the counterfactual's full data.

2.3.1 Algorithms

We describe the guidelines for the implementation of two algorithms. Later, in the applied example, we will refer again to these algorithms in full detail, outlining the calculations for the particular data.

Natural Direct Effect

Consider the observed data given by n i.i.d copies of $O = (Y, A, Z, W)$ where Y is a binary outcome, A is a binary exposure, Z is an intermediate variable with continuous values and W is the set of all measured confounders, discrete and continuous. To calculate the natural direct effect we will have to impute the counterfactual observations for Z_0 as if the interme-

diate variable was not exposed to A. Next we present the steps to follow in order to obtain the G-computation estimate of the desired direct effect.

1. Use Super Learner with J candidate learners to estimate the optimal model for the intermediate variable given the exposure and the covariates: $Q_Z = Q(Z \mid A, W)$. Q_Z^* represent the optimal model returned by the Super Learner procedure.

2. Predict the counterfactual level for the intermediate variable Z at $A = 0$ using the previously defined model, $\widehat{Z}_0 = Q_Z^*(A = 0, W)$.

3. Use Super Learner to define the optimal model for the outcome Y given the observed exposure A, intermediate variable Z and covariates W: $Q_Y = Q(Y \mid A, Z, W)$. Q_Y^* represent the optimal model returned by the Super Learner procedure.

4. With the model defined in the previous step calculate the counterfactual values for the whole group at the exposed ($A = 1$) and unexposed ($A = 0$) levels, holding the values for the intermediate variable at the non-exposed values calculated in step 2.
$\widehat{Y}_1 = Q_Y^*(A = 1, \widehat{Z}_0, W)$ and $\widehat{Y}_0 = Q_Y^*(A = 0, \widehat{Z}_0, W)$.

5. Obtain the G-estimator of the natural direct effect as the mean over the population's difference between the two exposure levels:

$$NDE(a) = E(\widehat{Y}_{1Z_0} - \widehat{Y}_{0Z_0}) = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_{1Z_0} - \widehat{Y}_{0Z_0})$$

Controlled Direct Effect

Consider again the observed data given by n i.i.d copies of $O = (Y, A, Z, W)$. In contrast to the previous definition of Z, this time it will be a binary variable. To avoid any confusion we denote this new intermediate variable as Z'. To estimate the parameter of interest this time we will use TMLE. We define a new variable $A^* = I(A = a, Z' = z')$ for the target step.

1. The new variable A^* has different levels at which we want to estimate the controlled direct effect. There will be a total of $N_A \times N_Z = 2 \times 2 = 4$ level combinations at which we will estimate initial models for the TMLE procedure.

2. Use Super Learner to obtain an initial estimate of the model for the outcome Y on the exposure A^* and the covariates W, $Q_Y^0 = Q(Y \mid A^*, W)$.

3. Calculate the nuisance parameters $g(A^* \mid W)$ for the targeted steps. Each one of these nuisance parameters will be also estimated using Super Learner.

4. Calculation of the new covariate to adjust the initial estimate Q_Y^0 will be defined as

$$h(A, W) = \left(\frac{I(A^* = a^*)}{\widehat{g}(a^* \mid W)} \right).$$

5. Use $h(A^*, W)$ to update the initial estimate Q_Y^0. Perform a simple regression, using the initial estimates $m^0(A, W)$ as an offset, to obtain ϵ;

$$Q_Y^1(A^*, W) = Q_Y^0(A^*, W) + \epsilon h(A^*, W).$$

6. Calculate the controlled direct effect of the exposure on the outcome as

$$CDE(a^*) = E(\widehat{Q}_Y^1(a_1^*, W) - \widehat{Q}_Y^1(a_0^*, W)) = \frac{1}{n} \sum_{i=1}^{n} \Big(\widehat{Q}_Y^1(a_1^*, W_i) - \widehat{Q}_Y^1(a_0^*, W_i) \Big).$$

where we denote $a_1^* = (1, z')$ and $a_0^* = (0, z')$ as counterfactual levels for exposure holding the levels of $Z' = z'$ fixed.

2.4 Data Analysis

The question of interest is the impact of maternal depression on cognitive and language development in 350 Mexican-American children in the Center for the Health Assessment of the Mothers and Children of Salinas (CHAMACOS) birth cohort study. This project is a longitudinal birth cohort study mainly investigating the health consequences of pesticide exposure to the mothers and children who live in the predominantly Mexican, migrant, farm worker population in the Salinas Valley of California. Participants were enrolled between October 1999 and October 2000. CHAMACOS' staff recruited 601 pregnant women for participation. Women were eligible to participate if they received prenatal care at the Natividad Medical Center or at one of five affiliated clinics of the Clinica de Salud del Valle de Salinas and planned to deliver at the Natividad Medical Center, were at least 18 years old, spoke Spanish and/or English, were less than 20 weeks pregnant, and were entitled to health benefits under Medi-Cal. All the infants in the CHAMACOS study were born between January 2000 and June 2001. The study was approved by the Institutional Review Board of the University of California, Berkeley. Of the original 601 mothers enrolled, 42 were lost due to relocation, 20 miscarried, three had stillbirths, and two had neonates who died. The remaining 534 women delivered 539 infants. For this analysis we excluded 27 children of non-Mexican origin to make the population more homogeneous, 10 twins, seven infants with seizures, two infants with incomplete consents, and 30 preterm infants. Additionally, 113 dyads did not complete the assessment for both maternal depression and child neurodevelopment, leaving a total of 350 mothers and their infants.

Figure 2.4 shows the corresponding DAG to the problem in hand, were the outcome Y is any of the four outcomes of interest (Mental, Motor, Auditory, Expression). The exposure of interest A, for which we want to estimate the direct effect on each of the outcomes, is maternal depression, as classified during the 12-month postpartum interview and coded into a binary variable. The intermediate variables Z_1 and Z_2 correspond to HOME-scale score and breastfeeding duration in months, respectively. The set of measured confounders is given by maternal age, maternal education, years in the U.S., language spoken at home, marital status, presence of father at home, child's gender, number of other children in home, PPVT-III score, poverty level, infant cared for outside of home, social support score, exact age of child in months, and psychometrician performing evaluations.

Women were interviewed twice during pregnancy at 13 and 26 weeks gestation. Later they were interviewed again at delivery, at six and 12 months postpartum. These interviews were designed to collect information on demographic characteristics, behaviors, medical, occupational, and family history.

Breastfeeding duration was derived from maternal report at the interviews. Mothers were asked if they continued breastfeeding their children. If the answer was negative, the duration of lactation was asked. Only 17 mothers did not breastfeed their children at all. The median duration was 6 months, but this was no symmetrical distribution; 30% of women were still breastfeeding at 12 months. At six months postpartum, the mothers' scholastic verbal abilities were assessed using the Peabody Vocabulary Test - Third Edition (PPVT-III),[15] an individually-administered test available in English and Spanish. Age-standardized scores have a mean of 100 and a standard deviation of 15.

At the 12-month postnatal visit, the study's staff evaluated the home environment using the Infant/Toddler Home Observation for the Measurement of the Environment (HOME).[12] This is an extensively used instrument that evaluates the degree to which the home environment provides emotional support and offers experiences objects that foster intellectual growth in children up to age three. The HOME contains 45 items, which cluster into six subscales: parental responsivity, acceptance of the child, organization of the environment, learning materials, parental involvement, an variety in experience. Scores were based on mothers' responses during the interview and on the psychometricians' observations of the mothers' interactions with their children. Higher HOME scores indicate a better home environment.

At approximately 12 months postpartum, the mothers were also screened for depression using the Center for Epidemiological Studies Depression Scale (CES-D),[59] a 20-item self-report scale which assesses five dimensions of depressed mood: feelings of guilt and worthlessness; feelings of helplessness and hopelessness; psychomotor retardation; loss of ap-

petite; and sleep disturbance. Participants responses were rated on a four-point Likert-type scale (0=rarely or none of the time to 3=most or all of the time) and a total score ranged from 0-60. The validity and reliability of the CES-D have been well established in general and clinical populations.[26] Women were classified as "depressed" and "non-depressed" if they had scores ≥ 16 and < 16, respectively.

The key outcome variables were mental, motor, auditory, and expression development. At age 12 months, psychometricians evaluated the infants in English and/or Spanish using the Bayley Scales of Infant Development, Second Edition (BSID) [6] and the Pre-School Language Scale, Third Edition (PLS).[78] The BSID is an individually-administered instrument used to evaluate the development of children between the ages of three and 42 months. Stimulus items designed to tap their memory, problem-solving, language, fine motor, gross motor, and personal-social abilities are clustered together in three scales - the Mental Development Index (MDI), the Psychomotor Development Index (PDI), and the Behavior Rating Scale. Within each of these scales, raw scores are converted into scale (index) scores with a mean of 100 and a standard deviation of 15. Reliability coefficients indicate that all three scales of the BSID-II are internally consistent and stable.

The PLS, available in English and Spanish, evaluates emerging language behaviors in children between the ages of two weeks and six years. The Auditory Comprehension (AC) and Expressive Communication (EC) subtests measure receptive and expressive language abilities respectively, by assessing several recognized precursors of later verbal development: attention, social communication, and vocalization skills. The PLS has a mean of 100 and a standard deviation of 15.

In order to compare the proposed methods to traditional ones, we repeat the analyses using standard parametric regression, where no model selection is performed, and move towards more flexible, data adaptive methods, like GAM and Polymars and end up with the implementation of Super Learner using all three previously listed methods as the candidate learners.

The inference for all models was obtained through boot strapping the models 1,000 times. This way we obtain a non parametric version of the confidence intervals for the parameters of interest. We performed estimations for each one of the four outcomes of interest and for each of the considered candidate learners; a total of sixteen runs for the natural and 16 runs for the controlled direct effects. We will describe our calculations in detail only once because they were analogous for each one of the cases. All the analyses were performed in R v 2.9.1 [58] using the SuperLearner-,[55] gam-, [30] and polspline-packages. [36]

In section 2.3.1 we described the general algorithms for estimating the natural and the controlled direct effects. Those algorithms should be used as a general guideline.

Natural Direct Effect

1. We used Super Learner with GLM, GAM and Polymars as the candidate learners and obtained an estimate of the optimal models for the HOME-scale (Z_1) and breastfeeding duration (Z_2) given maternal depression (A) and all the covariates (W). $\widehat{Z}_1 = Q(Z_1 \mid A, W)$ and $\widehat{Z}_2 = Q(Z_2 \mid A, W)$.

2. With the predicted models for HOME-scale and breastfeeding, we calculated their counterfactual levels had all the mothers been not depressed ($A = 0$), but holding all the covariates at their recorded values, $\widehat{Z}_{1,0} = \widehat{Z}_1(A = 0, W)$ and $\widehat{Z}_{2,0} = \widehat{Z}_1(A = 0, W)$.

3. We calculated the optimal models for the four outcome variables of interest (Y_i) corresponding to the infant neurodevelopment evaluated as Mental, Motor, Auditory, and Expression. We used Super Learner with the same candidate learners as before (GLM, GAM and Polymars). For estimating these models we used the actual recorded values. The main exposure (A) was maternal depression; the intermediate variables were HOME-scale (Z_1), and breastfeeding duration (Z_2); and all the covariates (W). We obtained four different models, one for each of the outcomes of interest. $Q^*_{Y_i} = Q(Y_i \mid A, Z_1, Z_2, W)$.

4. With the defined models we calculated the counterfactual outcomes under maternal depression and absence of it, holding the values for HOME-scale and breastfeeding duration at the "no-depression" level in both calculations. For example, in the Mental development case we calculated the Mental outcome had the mother been depressed but had this depression not affected neither the HOME-scale nor the duration of breastfeeding and then we perform the calculation again as if the mother had been not depressed, for which case we also kept HOME-scale and breastfeeding at the non-depressed level.

$$\widehat{Y}_{i,1} = Q^*_{Y_i}(1, \widehat{Z}_{1,0}, \widehat{Z}_{2,0}, W)$$

$$\widehat{Y}_{i,0} = Q^*_{Y_i}(0, \widehat{Z}_{1,0}, \widehat{Z}_{2,0}, W)$$

5. With the values for the outcomes of interest at both the two exposure levels, depressed and non-depressed, we calculated the average over the difference for each individual between the two levels. This is our estimate of the natural direct effect of maternal depression on infant neurodevelopment.

$$\widehat{NDE_i} = E\Big(\widehat{Q}_{Y_i}(1, \widehat{Z}_{1,0}, \widehat{Z}_{2,0}, W) - \widehat{Q}_{Y_i}(0, \widehat{Z}_{1,0}, \widehat{Z}_{2,0}, W)\Big)$$

where i ϵ {mental, motor, auditory, expression}

Controlled Direct Effect

We defined categorical values for HOME-scale and breastfeeding duration. In the case of the HOME-scale, we performed a median split for low and high HOME-scale values. In the case of the breastfeeding duration, the median split criteria coincides with the recommended breastfeeding duration of 6 months by the World Health Organization.[37, 24] Our new categorical variables for HOME-scale and breastfeeding duration were defined as shown in Table 2.1:

Categorical variable	0=No	1=Yes
HOME-scale	≤ 36 points	> 36 points
Breastfeeding	≤ 6 months	> 6 months

Table 2.1: Cut point value to transform HOME-scale and breastfeeding variables from continuous to binary

1. For the controlled direct effect analyses we redefined the exposure variable as the triad built by the recorded depression level (depressed vs non-depressed), the binary value for the recorded HOME-scale, and the binary value for the recorded breastfeeding duration. The combination of these three variables gave us a total of eight possible values for the new exposure variable A^*. For the targeting step, instead of creating a variable with eight levels we decided to use an indicator variable to denote if an individual belonged to the level being targeted.

A^*	Maternal depression	HOME-scale	breastfeeding
a_1^*	0	0	0
a_2^*	0	0	1
a_3^*	0	1	0
a_4^*	0	1	1
a_5^*	1	0	0
a_6^*	1	0	1
a_7^*	1	1	0
a_8^*	1	1	1

Table 2.2: New exposure variable A^* levels corresponding to the eight possible combinations of maternal depression-, HOME-scale-, and breastfeeding duration-status. Each one of this levels is targeted individually.

2. We used Super Learner with GLM, GAM and Polymars as the candidate learners to find the optimal model for each one of the four outcomes of interest given the recorded exposure and covariates. $Q^0_{Y_i} = Q^0(Y_i, A, Z_1, Z_2, W)$.

3. We estimated the nuisance parameters for each one of the level combinations of maternal depression, HOME-scale, and breastfeeding duration. In Table 2.2 we see the eight possible level combinations. We calculated the model for the particular level given the set of covariates. In our example we used only one candidate learner, step-forward. These calculations produced eight different models for g, one for each value of a^*_j where $j\epsilon\{1, 2, \ldots, 8\}$. $g_j = Q(A^* \mid W)$.

4. The target step has to be calculated assuming that $A^* = (A, Z_1, Z_2)$ is actually at the $a^*_j = (a, z_1, z_2)$ level; once for each of the eight levels. To calculate the clever covariate $h(A^*, W)$ that is used to target the initial estimate of the model, we divided the indicator variable of the new exposure at the particular level a^*_j by the estimate of the corresponding function $\widehat{g}_j(a* \mid W)$ at the same exposure level, given the observed covariates. Notice that given the fact that we are targeting a particular level of exposure for $A^* = a^*$, we assume that all the observations are at this level, turning all the values of the indicator variable in the numerator equal to one.

$$h_j(A^*, W) = \frac{1}{\widehat{g}_j(a^*_j \mid W)}.$$

5. Once we had the eight different estimates of $\widehat{h}_j$, we ran individual simple regressions with GLM for each of the four outcomes on $\widehat{h}_j$ and providing the initial estimates of the Super Learner model, $m^0_{Y_i}$ as an offset. This returned the estimate of the parameter ϵ_j.

6. The initial estimates were then updated by adding $\widehat{\epsilon}$ times $\widehat{h}$ to them. This provided the building blocks for the targeted controlled direct effect. Notice that these elements have been already target.

$$\widehat{Q}^1_{j,Y_i} = \widehat{Q}^0_{j,Y_i} + \widehat{\epsilon_j} * \widehat{h_j}(A^*, W)$$

7. The controlled direct effect is then calculated as the mean difference between the pairs of $a^*_j - a^*_k$ where these two only differ in their depression level; for example $a^*_5 - a^*_1$ in Table 2.2.

$$\widehat{CDE}_{il} = E\Big(\widehat{Q}^1_{j,Y_i}(a^*_j, W) - \widehat{Q}^1_{k,Y_i}(a^*_k, W)\Big)$$

where i corresponds to the four measures of infant neurodevelopment (mental, motor,

auditory, expression) and $l = \{(j : k) : j$ and k differ at the depression status but are identical fro HOME-scale and breastfeeding duration$\}$.

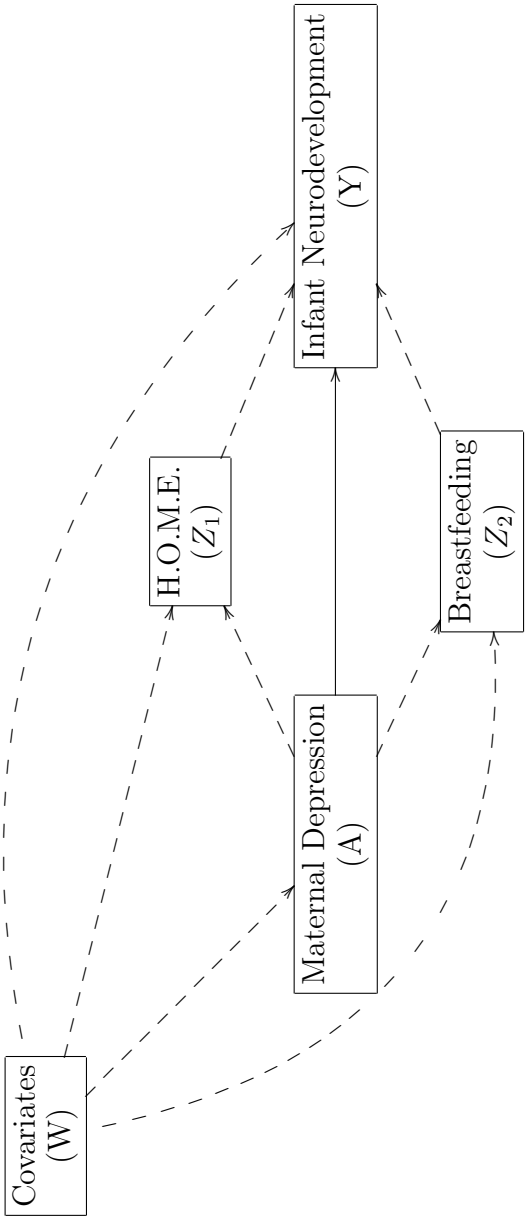


Figure 2.4: DAG for the estimation of the direct effect of Maternal Depression on Infant Neurodevelopment. Intermediate variables are the H.O.M.E-scale and duration of breastfeeding. Measured confounders are comprised by categorical and continuous variables.

Table 2.3: Demographic characteristics of CHAMACOS' study population, Salinas Valley, CA, 2000-2001 (n=350)

		Depression Score		
Independent variables	Overal (n=350)	CES-D< 16 (n=176)	CES-D≥ 16 (n=174)	p-value
Maternal age, years (mean± sd)	26.5 ± 5.3	26.6 ± 5.2	26.5 ± 5.3	0.90
Maternal education, n(%)				
≤6th grade	165 (47.1)	80(45.2)	85(49.1)	
7-12th grade	128 (36.6)	58(32.8)	70(40.5)	
Completed high school	57 (16.3)	39(22.0)	18(10.4)	0.01
Years in U.S., n(%)				
≤ 1	81(23.1)	34(19.2)	47(27.2)	
$2-5$	93(26.6)	49(27.7)	44(25.4)	
$6-10$	87(24.9)	43(24.3)	44(25.4)	
11+	50(14.3)	29(16.4)	21(12.1)	
U.S.-Born	39(11.1)	22(12.4)	17(9.8)	0.37
Language spoken at home, n(%)				
Mostly Spanish	324(92.6)	162(91.5)	162(93.6)	
Spanish & English equally	15(4.3)	7(4.0)	8 (4.6)	
Mostly English	8(2.3)	7(4.0)	1(0.6)	
Other	3(0.9)	1(0.6)	2(1.2)	0.18
Married or living as married, n(%)				
Yes	291(83.1)	146(82.5)	145 (83.8)	
No	59(16.9)	31(17.5)	28(16.2)	0.74
Father lives with mother & child, n(%)				
All of the time	290(82.9)	152(85.9)	138(79.8)	
Sometime or not at all	60(17.1)	25(14.1)	35(20.2)	0.13
Child sex, n(%)				
Female	181(51.7)	99(55.9)	82(47.4)	
Male	169(48.3)	78(44.1)	91(52.6)	0.11

Independent variables	Overal (n=350)	Depression Score CES-D< 16 (n=176)	Depression Score CES-D≥ 16 (n=174)	p-value
Number of other children in home, (mean ± sd)	2 ± 1.6	1.9 ± 1.4	2.1 ± 1.7	0.41
Breastfeeding duration, months (mean ± sd)	6.4 ± 4.6	6.2 ± 4.5	6.6 ± 4.6	0.45
PPVT-III score, points (mean ± sd)	86.0 ± 20.8	88.8 ± 20.4	83.1 ± 20.9	0.01
Within 100% of federal poverty limits, n(%)				
Yes	232(66.3)	107(60.5)	125(72.3)	
No	118(33.7)	70(39.6)	48(278)	0.02
Infant cared for outside of home > 15 hours/week, n(%)				
Yes	107(30.6)	52(29.4)	55(31.8)	
No	243(69.4)	125(70.6)	118(68.2)	0.62
HOME score, points (mean ± sd)	35.7 ± 3.0	36.1 ± 2.8	35.3 ± 3.1	0.01
Social Support score, points (mean ± sd)	3.8 ± 1.0	4.2 ± .9	3.4 ± 1.0	0.00

2.5 Results

Table 2.3 on shows the demographic characteristics of the population; overall and by depression status. Overall mean standardized scores (± SD) on the neurodevelopment assessment were 100.8 (±8.8) for Bayley MDI (Mental), 106.6 (±12.5) on the Bayley PDI (Psychomotor), 100.2 (±12.7) on the PLS-AC (Auditory) subtest and 95.5 (±13.7) on the PLS-EC (Expression) subtest. Table 2.4 shows the crude association found between maternal depression and the four measures of interest. In the simplest association analysis only the auditory and expression measures of infant neurodevelopment show significant differences between the depressed and non-depressed mothers. Infants of depressed mothers averaged 2.9 (95%CI: 0.2, 5.5) and 4.7 (95% CI: 1.8, 7.5) points lower than children of the non-depressed group in the PLS-AC and PLS-EC subtests, respectively. However, the mental and psychomotor measures of neurodevelopment do not differ significantly between the two groups of mothers, depressed and non-depressed.

Table 2.4: Differences in scores on the neurodevelopment tests of 12 month-old children by maternal postnatal depression status on the CES-D (¡16 vs. ≥16 points). CHAMACOS Study, Salinas Valley California 2000-2001.

	Overall	Maternal Depression (CES-D score)		
		< 16 points	≥ 16 points	
n	350	177	173	
Neurodevelopment	mean ± sd	mean ± sd	mean ± sd	p-value
MDI	100.8 ± 8.8	101.0 ± 9.2	100.6 ± 8.3	0.66
PDI	106.6 ±12.5	105.7 ±13.5	107.4 ±11.5	0.19
PLS-AC	100.2 ±12.7	101.6 ±12.4	98.7 ±12.9	0.04
PLS-EC	95.5 ±13.7	97.82 ±13.7	93.17 ±13.3	0.001

Table 2.5 presents all the natural direct effect estimates with their respective confidence intervals for the four candidates used and the four outcomes. The analyses' methods are presented from the most rigid to the most data adaptive; from Generalized Linear models to Super Learner, respectively. For the first of our parameters of interest, the natural direct effect, we observe that only language expression, as measured by the PLS-EC subtest, shows a significant direct effect from maternal depression. It is a negative direct effect that results in a drop of -2.82 (95%CI: (-5.61, -0.04), p-value=0.05) points compared to children of non-depressed mothers. This reduction in the PLS-EC subtest score corresponds approximately to a 20% of the observed standard deviation of the test.

Table 2.5: Natural DE estimates of maternal depression on infant neurodevelopment by the Preschool Language Scale and the Bayley Scales of Infant Development. Estimates were calculated using a G-computation estimator and confidence intervals were constructed by a bootstrap procedure. $\star$: p-value < 0.1, †: p-value < 0.05, ‡: p-value < 0.01

	Mental (MDI)		Motor (PDI)		Auditory (PLS-AC)		Expression (PLS-EC)	
Method	$\widehat{NDE}$	(95% CI)	$\widehat{NDE}$	(95% CI)	$\widehat{NDE}$	(95% CI)	$\widehat{NDE}$	(95% CI)
GLM	0.85	(-1.12, 2.81)	2.28	(-0.59, 5.14)	-2.64 *	(-5.62, 0.35)	-4.51 ‡	(-7.65, -1.37)
GAM	0.44	(-1.50, 2.38)	1.44	(-1.35, 4.23)	-1.65	(-4.24, 0.94)	$-3.72^†$	(-6.63, -0.80)
Polymars	-0.75	(-2.83, 1.33)	-0.08	(-2.53, 2.38)	-0.91	(-3.27, 1.44)	$-2.79^†$	(-5.54, -0.03)
SL	-0.75	(-2.66, 1.16)	-0.02	(-2.60, 2.57)	-0.96	(-3.18, 1.26)	$-2.82^†$	(-5.61, -0.04)

The apparent association between PLS-AC and maternal depression in Table 2.4 disappears in our estimates of the natural direct effect. We only observe a possible direct effect of maternal depression on the auditory development when using GLM as the single candidate learner and this is with a non-significant p-value of 0.08, somewhat larger than the standard 0.05. Even though no other outcome besides PLS-EC turned out to be significantly affected by maternal depression, we observe the hypothesized results of a negative effect of maternal depression on infant neurodevelopment as we move away from GLM to Super Leaner. This is, that for GLM we observe positive effects of maternal depression on mental and motor infant neurodevelopment which turn negative in the estimates from Super Learner. The negative effects are hypothesized and previously reported in the literature. [22, 45, 54] The estimates remained non significant regardless of the candidate learner employed to estimate them.

Table 2.6: Initial estimates, $Q0$, of the CDE of maternal depression on infant neurodevelopment measured in four different outcomes. Estimates shown are before targeting the parameter of interest and adjusting by the clever covariate of the TMLE procedure.

Method	Mental (MDI)	Motor (PDI)	Auditory (PLS-AC)	Expression (PLS-EC)
GLM	0.80	2.21	−2.65	−4.56
GAM	0.43	1.42	−1.69	−3.75
Polymars	−0.82	−0.13	−0.84	−2.87
SL	−0.79	−0.05	−0.94	−2.90

The controlled direct effect estimates ψ_{CDE} in Table 2.7 behave very similarly to the previous results for the NDE. The direct effect on PLS-EC is the only one that is consistently significant regardless of the candidate learner used to estimate it. We observe a negative effect of maternal depression over the child's expression development regardless of the HOME-scale and breastfeeding duration level combination. Only three estimates did not achieve statistical significance. The first one is using GLM at the low level for HOME-scale and the short breastfeeding duration. The second and third cases occur at the low level of HOME-scale but long breastfeeding duration for Polymars and the Super Learner, respectively. For the auditory language development, we mostly lack significance just like

in the natural case. Only three level combinations of HOME-scale and breastfeeding duration turned significant estimates and only when GLM was used to calculate them. For the mental and psychomotor outcomes we observe again a change in sign in the estimates. The effect of maternal depression changes from positive to negative as we move forward from the rigid methods to the data adaptive ones. These last behaviors are observed at most level combinations of the intermediate variables. The ones that do not show this change is sign show at least a reduction of the estimate towards the null of no effect. Finally, the changes that we observe from a low to a high level in the HOME-scale or between a breastfeeding duration of less than six months to a duration exceeding this threshold should be interpreted with out most caution because of the lack of significance.

We observe that the effect of the target procedure moderately modifies the initial estimates (Q^0) of the controlled direct effects at the HOME-scale and breastfeeding duration fixed levels; compare the targeted estimates in Table 2.7 to Table 2.6. This could imply that our original estimates of the controlled direct effect of maternal depression on the neurodevelopment outcomes were not markedly off target, or did not present particularly large bias.

Table 2.7: Controlled DE estimates $\widehat{CDE}$ of maternal depression on infant neurodevelopment measured in four different outcomes. Estimates were calculated using TMLE and confidence intervals were constructed using the estimated standard error by a bootstrap procedure of 1000 cycles. The paired numbers $\{00, 01, 10, 11\}$ correspond to the fixed levels of categorical HOME-scale and breastfeeding duration, respectively, as described in Table 2.1. $\star$: p-value < 0.1, †: p-value < 0.05, ‡: p-value < 0.01

		Mental (MDI)		Motor (PDI)		Auditory (PLS-AC)		Expression (PLS-EC)	
Method	Level	ψ_{CDE}	(95% CI)	ψ_{CDE}	(95% CI)	ψ_{CDE}	(95% CI)	ψ_{CDE}	(95% CI)
GLM	00	0.74	(-1.34, 2.81)	2.10	(-0.88, 5.09)	-2.42	(-5.36, 0.53)	-4.46	(-7.53, -1.39)
	01	0.89	(-1.17, 2.95)	2.23	(-0.74, 5.20)	$-2.55^{\star}$	(-5.46, 0.35)	$-4.41^{\ddagger}$	(-7.45, -1.38)
	10	0.66	(-1.39, 2.71)	2.26	(-0.71, 5.22)	$-2.87^{\star}$	(-5.78, 0.03)	$-4.60^{\ddagger}$	(-7.69, -1.51)
	11	0.74	(-1.31, 2.79)	2.13	(-0.81, 5.08)	$-2.84^{\star}$	(-5.76, 0.07)	$-4.66^{\ddagger}$	(-7.69, -1.64)
GAM	00	0.38	(-1.63, 2.39)	1.34	(-1.44, 4.12)	-1.47	(-4.15, 1.20)	$-3.66^{\dagger}$	(-6.57, -0.76)
	01	0.46	(-1.54, 2.46)	1.34	(-1.42, 4.09)	-1.55	(-4.21, 1.11)	$-3.57^{\dagger}$	(-6.46, -0.69)
	10	0.29	(-1.71, 2.29)	1.47	(-1.29, 4.23)	-1.90	(-4.56, 0.75)	$-3.80^{\dagger}$	(-6.72, -0.87)
	11	0.37	(-1.62, 2.37)	1.36	(-1.40, 4.11)	-1.80	(-4.46, 0.86)	$-3.77^{\dagger}$	(-6.64, -0.90)
Polymars	00	-0.52	(-2.90, 1.86)	-0.10	(-2.76, 2.57)	-1.09	(-3.57, 1.39)	$-3.28^{\dagger}$	(-6.17, -0.39)
	01	-0.70	(-2.93, 1.52)	-0.21	(-3.03, 2.61)	-0.51	(-3.04, 2.02)	-2.26	(-5.15, 0.63)
	10	-0.72	(-3.08, 1.64)	0.51	(-2.45, 3.47)	-1.21	(-3.72, 1.29)	$-3.15^{\dagger}$	(-6.06, -0.23)
	11	-0.72	(-3.02, 1.58)	-0.08	(-3.09, 2.93)	-0.92	(-3.45, 1.61)	$-2.74^{\star}$	(-5.65, 0.18)
SL	00	-0.50	(-2.65, 1.66)	-0.02	(-2.57, 2.57)	-1.16	(-3.55, 1.21)	$-3.30^{\dagger}$	(-6.05, -0.56)
	01	-0.67	(-2.65, 1.42)	-0.13	(-2.73, 2.54)	-0.63	(-2.96, 1.85)	-2.32	(-4.99, 0.51)
	10	-0.69	(-2.81, 1.48)	0.57	(-2.19, 3.33)	-1.30	(-3.67, 1.09)	$-3.17^{\dagger}$	(-5.98, -0.42)
	11	-0.69	(-2.65, 1.54)	0.003	(-2.75, 2.82)	-1.02	(-3.36, 1.37)	$-2.78^{\dagger}$	(-5.54, -0.05)

2.6 Discussion

The purpose of this article was to estimate the direct effect of maternal depression on infant neurodevelopment. Although our findings support the hypothesis that maternal depression affects the expression development of economically-impoverished children, we failed to confirm findings that children of depressed mothers have lower overall cognitive, motor and language development.[67, 54] This analysis has a number of limitations. One limitation is that mothers were not assessed for maternal depression until 12 months postpartum; thus, we do not know the timing or severity of their depression during the first year of the child's development. Previous studies have included older children and their cognitive assessment may have included a larger number of language-based items than the used on Bayley to assess 12-month olds. Sharp et al. (1995) reported that depressed mothers were less likely to breastfeed and that they breastfeed their male infants less than their female infants. The authors suggested that this may in part explain why maternal depression impacts boys more than girls. However, in the CHAMACOS population, where nearly all women breastfed, depressed mothers tended to breastfed longer than non-depressed mothers. Although depressed women breastfed their male infants for a slightly shorter period of time than their female infants (data not shown), there were no differences in the impact of depression on male and in female children.

Another potential limitation of this analysis is that CES-D may not have been an appropriate tool for assessing maternal depression in this Mexican-origin population. Based on the recognized cut-points for depression using the CES-D, 13% of the women in this research sample obtained scores that signaled major depression (> 30) and 37% were mildly to moderately depressed (≥ 16 and ≤ 30). Although the CES-D is a widely used tool of depression and is available in Spanish, some have suggested it may not be a valid measure of depression in an immigrant population or that a higher score should be used for defining depression [77]. Compared to infants of mothers who were not depressed, we found that infants of mothers with depression had poorer expressive language abilities (PLS-EC), but that they did not differ in general cognitive (Bayley MDI), motor(Bayley PDI) or auditory language (PLS-AC) performance when estimating the natural direct effect. The targeted controlled direct effect estimates showed the exact same significance pattern as the natural direct effect estimates.

We believe that the use of machine learning (data adaptive) techniques to determine the underlying model of the observed data provides in the end better estimates of the direct effects, both natural and controlled. When using machine learning techniques, there is no predetermined model that needs to be justify, and the risk of violating some of the assumptions in the estimation of direct effects is reduced.

The proposed algorithms should be followed just as a guideline. We performed more

analyses and evaluated other candidate learners, like support vector machine (SVM) and D/S/A. However, we found difficulties with some of them and decided to exclude them. The main problem with most data adaptive techniques used to determine the optimal model of a data set is that they do not allow the user to fix or force variables into the model. This presented a major issue in our calculations, when the exposure was not selected by the candidate learner and was simply left out of the model. We could have complicated the algorithms and forced the exposure into the optimal models after they had been selected by those candidate learners, but this would have just obscured what might already seem a convoluted procedure to the less experimented programmers.

Another issue consider when using this type of calculations is the time needed to run them. The extended duration to perform these analyses occurs in the bootstrap procedure which obtains distribution estimates. Some candidate learners are inefficient, resulting in an extended time for each run. The calculations for the natural direct effect using all three candidates in the Super Learner procedure took ten hours to run; while the ones for the controlled direct effect took over five days. In particular we were forced to drop D/S/A as one of the candidate learner because each run in the TMLE estimation took on average 26 minutes. This would have made the estimation of the inferences, a thousand bootstrap cycles for each outcome, take about 20 days to complete. The time constrains should be considered before starting a full run which might take extremely long time. On the other hand we believe that anything within the time frame of one week is reasonable if we consider how much time is invested in collecting the data; sometimes years.

Chapter 3

Semi-parametric estimation of adjusted (marginal) associations in high-dimensional studies of environmental exposures: an example from estimating the potential effects of organophosphate pesticides on infant growth

3.1 Introduction

The Environmental Protection Agency's Pesticide Program report estimated that more than 1.2 billion pounds of pesticide active ingredients are used every year in the United States, with approximately 700 million pounds used in agriculture. [34] California is the state with the largest agricultural output, and in particular the Salinas Valley in Monterey County, which is often referred to as the "Nation's Salad Bowl". There is evidence of widespread pesticide exposure for all the population in the U.S., including pregnant women and children. [1, 8, 9, 31, 40, 20] There have been few studies that have examined the association of prenatal pesticide exposure and fetal growth or gestational duration in humans, but they show inconclusive results.[60, 19, 18] Restrepo et al (1990) found that potential exposure of women to pesticides during pregnancy was associated with an increased risk of low birth weight, small for gestational age (SGA), preterm delivery, or shortened gestation, whereas Fensterd and Coye (2003) found no association. Eskenazi et al. (2004) failed to demonstrate an adverse relationship between fetal growth and any measure of *in utero* organophosphate

pesticide exposure. No studies have examined the relationship of *in utero* OP exposure and subsequent effects in children's growth nor child exposure and subsequent effects.

The purpose of the present analyses is to determine if there is evidence of an effect between the exposure to organophosphate (OP) pesticides and the child's growth at 6, 12, 24 months, 3.5 and 5 years. Exposure is assessed by measurement of urinary OP metabolites at seven different time points: prenatally (twice) and at delivery in the mother and later in the child at the same times that the outcome of interest was recorded. Growth in children was evaluated in four measures: weight, length/height, body mass index (BMI) (kg/m^2), and waist circumference. We use Targeted Maximum Likelihood Estimation with a G-comp approach to obtain the desired effect estimates of exposure on outcome. This is a semiparemetric approach that allows us to avoid defining models *a priori*.

3.2 Materials and methods

Participants

The population are participants of the longitudinal birth cohort study of the Center for the Health Assessment of the Mothers and Children of Salinas (CHAMACOS), which is conducted by the Center for Children's Environmental Health Research at the University of California, Berkeley. The study focuses on the effects of pesticides and other environmental exposures on the health of pregnant women and their children living in the Salinas Valley.

Participants, pregnant women, were enrolled between October 1999 and October 2000 as they entered prenatal care at Natividad Medical Center or at any of five centers of the Clinica de Salud del Valle de Salinas. This is a low-income, mainly agricultural, migrant, and of Mexican descent population. For a detailed description of the population and the enrollment criteria see Eskenazi et al. [18] The women were interviewed twice during pregnancy and shortly after delivery. The women and their children were followed and interviewed after birth on pre-established time intervals at 6, 12, 24 months, and at 3.5, and 5 years. During these interviews, trained bilingual bicultural personnel conducted extensive and detailed questionnaires about the child's health, family status, and life events during the time since the last interview. The children biometrics were carefully measured and recorded. Biological samples including urine were also collected.

Pesticide exposure measurement

The exposure to organophosphate pesticides was assessed by measuring non-specific organophosphate dialkyl phosphate metabolites in urine. Maternal urine was collected during pregnancy (twice) and at delivery. The first and second samples during pregnancy occurred at a mean of 13 weeks gestation (range, 4-29 weeks) and at 26 weeks (range, 18-39 weeks), respectively. Child's urine was then collected at each of the subsequent interviews after birth. Spot urine samples were collected at each of the interviews mentioned above. Urine specimens were aliquoted and stored at -80C until shipment to the Center for Disease Control and Prevention (CDC; Atlanta, GA) for analysis of dialkyl phosphate metabolite levels. [18]

Six dialkyl phosphates metabolites were measured in the urine samples using gas chromatography and mass spectrometry and quantified using isotope dilution calibration.[10] The dialkyl phosphates measured were dimethylphosphate, dimethyldithiophosphate, dimethylthiophosphate, diethylphosphate, diethyldithiophosphate, and diethylthiophosphate. This six metabolites are grouped into dimethyl phosphates (DMs) and diethyl phosphates (DEs). Approximately 80% of the organophosphate pesticides used in the Salinas Valley devolve to one or more of these metabolites, which are excreted in urine. The most commonly used pesticides in this region that devolve to dialkyl phosphates are presented in Table 4.1

Table 3.1: Dialkyl phosphate pesticides metabolites and parent compounds

Marker of exposure	Parent compounds or class
Dialkyl phosphate metabolites (nmol/L)	
Dimethyl phosphates	Malathion, oxydemeton-methyl dimethoate, naled, methidathion
Diethyl phosphates	Diazinon, chlorpyrifos, disulfoton

At each time of exposure the following procedures were applied to urine samples and their metabolites measurements: given that dialkyl phosphates come from more than one OP pesticide, quantities of the six metabolites were converted to molar concentration (nanomols per liter) and summed to obtain the total concentrations of dialkyl phosphate (DAP) metabolites for each sample. We averaged the two prenatal sample concentrations. The sum of all six metabolites provided an estimate of the total organophosphate exposure for each individual at each of the time points listed in Table 3.2. The three DMs and three DEs were also added to obtain total concentrations of dimethyl and diethyl phosphate metabolites, Total DAPs.

When working with urine samples, it is a standard procedure to adjust for creatinine concentration. Creatinine gives a reference for how diluted the samples are. The concentrations were determined using a commercially available diagnostic enzyme method (Vitros CREA

slides; Ortho Clinical Diagnostics, Raritan, NJ). Samples with creatinine levels $< 10mg/dL$ were considered too dilute for accuracy of analysis, and therefore excluded. Table 3.2 shows the number of available records by time point.

Definition of outcomes

We studied the effects of the three considered exposures (Total DAPs, DMs, and DEs) on four measures of growth in children (weight, length/height, BMI, and waist circumference). Child's recumbent length or standing height in centimeters, and child's weight in kilograms were collected at each one of the considered time points. Infant birth crown-heel length and weight were obtained from hospital delivery logs and medical records. The following measurements were performed by trained staff of the CHAMACOS project. Infants length up to two years were measured using a table or board, and children two years and older using a wall mounted stadiometer. These measurements were performed three times and then averaged. Weight was recorded once using an appropriately adjusted scale. BMI, adjusted for sex and exact age in months, were calculated using the collected data starting at 24 months because this is the minimum age that the Center for Disease Control and Prevention (CDC) has established on its growth charts for this variable. Child's waist circumference in centimeters was only collected starting the 5-years interview. Given that this is the end point for the collection of data we only present this child outcome at this particular age.

Table 3.2: Sample size of the four outcomes under study at each of the considered time points.

	Exposure	Outcome			
Time of measurement	Total DAPs / Creatinine	Weight (kg)	Length (cm)	BMI (kg/m^2)	Waist Circum. (cm)
Prenatal	533[1] / 533	-	-	-	-
Delivery	493 / 493	-	-	-	-
6 months	417 / 414	425	423	ND	NA
12 months	405 / 405	415	415	ND	NA
24 months	381 / 380	389	386	386	NA
3.5 years	298 / 281	339	336	335	NA
5 years	330 / 319	332	331	331	332

1 There was an additional observation for DMs, 534.
ND: Not Defined
NA: Not Applicable

Eskenazi et al. (2004) performed a first attempt to identify adverse effects from organophosphate pesticides on fetal growth in this same population.[18] In their analysis, exposure was collected on the mothers and then birth length, weight, head circumference and gestational age were the outcomes under analysis. They used multivariate regression, adjusting for selected covariates, and those were the parametric models presented in their results.

Parameter of interest

Our parameter of interest is the marginal adjusted effect difference of the exposure A on the outcome Y. A is a discrete exposure with three possible levels: low, medium, high. Given a full unobserved (counterfactual) dataset $X_{full} = (Y_a, W, a \in \mathcal{A})$, where Y_a corresponds to the observed outcome under exposure $A = a$, we define our parameter of interest as $\psi_{a,0} = E(Y_a) - E(Y_0)$. We define (for each contribution of the risk factor of interest, A, and outcome, Y) the observed data as $O = (Y_A, A, W) \sim P$. Specifically, under identifiability assumptions discussed below, we can estimate ψ as function of the observed data, $\psi(P)$.

$$\begin{aligned} \psi_{a,0} &= E\left[Y_a - Y_0\right] \\ &\underset{=}{\overset{\text{U.A.}}{}} E_W\left[E(Y \mid A = 1, W) - E(Y \mid A = 0, W)\right] \qquad \text{U.A.} = \text{under assumptions} \end{aligned}$$

The estimation will only be valid for the observed dataset if the following assumptions hold.

The first assumption is the consistency assumption which states that the observed data, $O = (Y_A, A, W)$, is only a missing data case from the counterfactual's full data, or $\psi_{a,0}(P) = \psi_{a,0}$, and the potential outcome for any particular unit is stable in the sense that it would take the same value under its exposure, independently of what other units get for an exposure. This is also know as the stable unit treatment value assumption (SUTVA).[66]

The second assumption is known as coarsening at random (CAR) and it states that the treatment assignment is independent of the outcome conditional on the measured covariates. Consider the dataset $O = (Y, A, W)$, where Y is the outcome, A is the exposure or treatment assignment which can take values $a \in \mathcal{A}$, and W is the set of covariates. Under the CAR assumption we have: $A \perp Y_a, \forall a \in \mathcal{A} \mid W$.

The third assumption is known as the experimental treatment assignment (ETA) assumption. It requires that the probability of receiving treatment conditional on the covariates is not fully determined. In practical terms, the conditional probability needs to be away from 0 and 1. Observations within strata of the covariates W must have a probability greater than 0 of receiving treatment at all possible levels of the treatment assignment, $\forall a \in \mathcal{A}, P(A = a \mid W) > 0$. When the ETA assumption does not hold it is said that there is a "theoretical ETA violation". On the other hand, there are also "practical ETA violations".

These occur when $\exists a \in \mathcal{A}$ such that $P(A = a \mid W) < \epsilon_n$ or $P(A = a \mid W) > 1 - \epsilon_n$, for a small ϵ_n (e.g. $\epsilon_n = 0.1$) and n is the sample size.[47]

Given the defined levels of exposure, we can estimate the adjusted marginal effect at each one of them before calculating the effect difference between them, as suggested in Moore and van der Laan (2008).[44] We refer to them as the elements of our parameter of interest, and denote them as $\delta_a = E_W\left[E(Y \mid A = a, W)\right]$.

Estimators

We estimated the effects of organophosphate pesticides on child's growth using targeted maximum likelihood estimation (TMLE).[76] TMLE carries out a targeted bias reduction specifically for the parameter of interest, hence its name. Consider a model $\mathcal{M}$ where the true distribution of the data is p_0. Consider and initial estimator $\widehat{p}$ of p_0. The parameter of interest is given by $\psi = \psi(p_0)$. TMLE has two goals. First, it aims to find an optimal density $\widehat{p}^* \in \mathcal{M}$ such that the plug-in estimator of the parameter of interest based on this density is asymptotically equivalent to an estimator that solves an estimation equation based on the so-called efficient influence curve. This results in a bias reduction compared to the estimate of $\psi(\widehat{p}_{MLE})$. Second, the algorithm also requires that $\widehat{p}^*$ achieves an increase in the log-likelihood relative to $\widehat{p}$. The resulting substitution estimator $\psi(\widehat{p}^*)$ is a familiar type of likelihood-based estimator, where the parameter of interest is based on the G-computation formula, but the original MLE density estimate is modified such that the resulting estimate is asymptotically linear with the desired efficient influence curve. [44] For complete technical and theoretical details about this general estimating approach we refer readers to the seminal paper by van der Laan and Rubin.[76]. Next, we present the basic steps involved in TMLE of δ_a:

1. Estimate the conditional expectation of Y given A, and W; denoted by $E\left[Y \mid A, W\right] = \gamma(Q(A, W)) = Q_n^0(A, W)$, where γ is the link function.

2. Estimate the conditional distribution of the exposure A given covariates W; denoted by $g_n^0(A \mid W)$

3. Calculate a specific covariate for each individual target based on the subjects observed values A, W and the estimate $g_n^0(A \mid W)$. This new covariate, whose form depends on the parameter of interest and the model of $Y \mid A, W$, is denoted by $h(A, W)$ and sometimes referred to as a "clever" covariate.

4. Update the initial regression $Q_n^0(A, W)$ by adding the clever covariate $h(A, W)$ and estimating its corresponding coefficient by simple maximum likelihood, holding $Q_n^0(A, W)$

fixed at their initial values by using it as offset. The updated regression is denoted $\gamma(Q^1) = \gamma(Q^0 + \epsilon h)$.

5. Implement the estimator by using Q^1 to predict with all $A = a$ and keeping the original values for W.

Specifically, the TMLE estimate of δ_a:

$$\delta_a^{TMLE} = \frac{1}{n} \sum_{i=1}^{n} Q_n^1(a, W_i). \tag{3.1}$$

The TMLE is consistent if at least one of the two, Q and g, is consistently estimated. Additionally, the estimator is locally efficient in the sense that it is efficient if both models are estimated consistently.

We need to estimate Q_{n0} and g_{n0} and the different estimation procedures rely on these models being correctly specified. Therefore, in order to avoid misspecified models, we will not use parametric models specified *a priori*. We will use a data-adaptive model selection approach, a machine learning approach, which has the property of potentially approaching a non-parametric model as $n \to \infty$. There are many such procedures available, like the Deletion/Substitution/Addition (D/S/A) algorithm, Least Angle Regression [16], Random Forest [11], Support Vector Machine, Generalized Additive Models (GAM), and Polychotomous Regression[35]. Each one of the procedures listed here returns an "optimal" model, not all of them identical, based on their own optimality criteria.

For example, the D/S/A algorithm performs data-adaptive estimation through selection of the estimators based on heavy use of cross-validation and the $L2 - loss$ function. The candidate estimators will always be polynomials which comply to user-specified constraints like maximum number of terms in the polynomial, maximum power of polynomial terms and maximum order of interaction.[68] Another example is gam, which replaces the linear form of the covariates ($\sum \beta_i X_i$) by a sum of smooth functions ($\sum s_i(X_i)$), where the functions $s_i(\cdot)$ are unspecified and are estimated using an iterative procedure called local scoring algorithm, resulting in a nonparametric regression method.[29] A third example is Polychotomous Regression, which fits a regression model using linear splines and their tensor products.[21, 35]

Above, we mentioned a list of model selectors which we will refer to as "candidate learners". We can use them to define the models we need for our estimations. Van der Laan et al.[74] propose an algorithm to select the "best possible model" using a convex linear combination of the resulting models from each of the candidate learners. The resulting model will be as good or better than every single model from the candidate learners.

Data analysis

In our analyses, we ran all possible combinations of exposure on outcomes as long as the time of exposure occurred prior or concurrent to the time of the outcome. The three exposures considered were the total sum of dialkyl phosphate metabolites (Total DAPs), the sum of dimethyl phosphates (DMs), and the sum of diethyl phosphates (DEs). We already presented the outcomes and their frequency can be seen in Table 3.2. We have that weight and length were measured at 5 time points; BMI was measured at 3 time points; and waist circumference only measured at one time point. Given the temporal constraint defined above for times of exposure and times of outcome, we ran 75 cross sectional analyses for weight and length; 54 cross sectional analyses for BMI; and 21 cross sectional analyses for waist circumference. For example, these last 21 cross sectional analyses correspond to the three different exposures (Total DAPs, DEs, and DMs), and the 7 valid times of exposure (Prenatal, Delivery, 6M, 12M, 24M, 3.5Y, and 5Y), prior or concurrent to the waist circumference measurement at 5 years.

An important consideration in all these analyses was that only complete pairs of exposure-outcome could be included. In those cases where one of the six metabolites was not readable because of analytic interference, the missing value was imputed using simple regression analysis on the other metabolites within the same group (i.e. diethyl or dimethyl phosphates, respectively) at the corresponding time point. This imputation was justified because of the high correlation of the metabolites within groups at the same time point. Metabolites which were missing because their levels were below the limit of detection (LOD) were given the value of the LOD divided by the square root of two. [32]

Among the complete exposure-outcome pairs, some of their covariates had missing values. We decided to impute these covariates to avoid further loss of observations. The imputations were performed in R using the multiple imputations package "mi".[58, 23] The strength of this package is that it performs an iterative regression imputation of the missing values until approximate convergence is achieved. In our case, the matrix W of covariates could be split into those covariates with missing observations, M with columns $M(1), \ldots, M(K)$, and those covariates with complete observations, C. First, the missing values of M are imputed using a crude approach (for example, imputing by randomly selecting from the observed outcomes for that variable). Then the algorithm continues imputing $M(1)$ given $M(2), \ldots, M(K)$ and C; imputing $M(2)$ given $M(1), \ldots, M(K)$ and C; and so forth, randomly imputing each variable and looping until approximate convergence.

It is important to mention that the covariates changed depending on the time of the outcome under consideration. This means that not all the same covariates were considered for all time points. Some of them were baseline covariates that did not change and others were time depending covariates that became relevant at a later age of the child. For example,

the variable corresponding to the child being breastfed at the time of the interview did not make sense after two years of age; preschool was not considered until 3.5 years when it is seemed plausible that children started attending preschool. For a complete list of covariates considered at each time point see Table 4.2 . Only some covariates were included in the optimal model.

Table 3.3: Covariates used in the multiple analyses

Covariate	6M	12M	24M	3.5Y	5Y
Sex	✓	✓	✓	✓	✓
Age in months	✓	✓	✓	✓	✓
Birthweight	✓	✓	✓	✓	✓
Infant's chest circumference at delivery	✓	✓	✓	✓	✓
Gestational age	✓	✓	✓	✓	✓
Breastfeeding status	✓	✓	✓		
Breastfeeding length	✓	✓	✓	✓	✓
Husband's agricultural worker	✓	✓	✓	✓	✓
Maternal smoking status	✓	✓	✓	✓	✓
Child around smokers	✓	✓	✓	✓	✓
WIC	✓	✓	✓	✓	✓
Child attended child care	✓	✓	✓	✓	✓
Mother's height	✓	✓	✓	✓	✓
Mother's pre pregnancy weight	✓	✓	✓	✓	✓
Mother's pre pregnancy BMI	✓	✓	✓	✓	✓
Poverty status at baseline	✓	✓	✓	✓	✓
Poverty status at the time	✓	✓	✓	✓	✓
Maternal education	✓	✓	✓	✓	✓
Paternal education	✓	✓	✓	✓	✓
Years in the US at baseline	✓	✓	✓	✓	✓
Maternal calories consumption at 26wks	✓	✓	✓	✓	✓
Diet Quality Index for Pregnancy at 26wks	✓	✓	✓	✓	✓
Child attended preschool at 42M				✓	✓
TV hours/weekday				✓	✓
TV hours/weekend day				✓	✓
Hours/day child played outside in past week				✓	✓

For simplicity of calculations, all exposures of interest were categorized into tertiles within their specific time point. This simplification makes the targeted step of our estimation much simpler. Hence, we calculated the targeted mean outcome value for each of the outcomes of interest (weight, length, BMI, and waist circumference) at each tertile of exposure. Using these estimates, it was easy to calculate our parameters of interest, the targeted mean difference between tertiles of exposure. We then calculated all three possible differences between the tertiles.

The inference for all models was obtained using the Influence Curve (IC). Assuming Q and g are known (that is, variability introduced by modeling uncertainty is a second order phenomena), the IC for our final estimate, $\widehat{\psi}_{a,0}^{TMLE} = \widehat{\delta}_a^{TMLE} - \widehat{\delta}_0^{TMLE}$ is given by:

$$IC(O; Q_n^1; g_n^1; \widehat{\psi}_{a,0}) = h(A = a, W) * (Y - Q_n^1(A, W)) + Q_n^1(A = a, W) - \psi \tag{3.2}$$

so

$$SE(\widehat{\psi}_{a,0}) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} \widehat{IC}^2(O_i)}}{\sqrt{n}} \tag{3.3}$$

The IC is very sensitive to violations of the ETA assumption, and when there is a practical violation, the estimate can be extremely variable. Whenever there was evidence of an ETA violation, an *ad hoc* procedure was implemented to reduce the set of covariates to correct the problem. Covariates that did not show any variability within the discrete exposure levels were eliminated since these would create a deterministic assignment of the value of the exposure given the covariate. Clearly this is an *ad hoc* procedure and we note that it also could introduce bias if these variables are true confounders. Another robust alternative to deal with this problem is the use of Collaborative TMLE, suggested by van der Laan and Gruber (2009) where the selection of the treatment mechanism is based on an algorithm that maximizes, over candidate estimators of the treatment mechanism, the log-likelihood of the corresponding candidate targeted maximum likelihood of the relevant factor.[73]

We describe our calculations in detail only once even though we did the same calculation for each one of the four outcomes of interest at each of the tertiles of exposure for each of the three exposures of interest. All the analyses were performed in R v 2.9.1 [58] using the SuperLearner-,[55] gam-, [30], polspline-packages (polymars),[36] and D/S/A.[48] See Appendix A for computational details.

We performed multiple tests searching for effects of the exposures on the outcomes. Therefore, the risks of a false positive discovery increased and we adjusted the significance of our results using Bonferroni. This way, we reduced the probability of false discovery and were more confident of the validity of the discoveries that remain significant after the adjust-

ment. The four different outcomes had a different number of tests for each type of exposure. We adjusted individually for that number; weight and length were adjusted by a factor of 75, BMI was adjusted by a factor of 54, and waist circumference only by a factor of 21. The multiple testing adjustment was performed on the Wald Test p-values obtained for each analysis.

Comparison of traditional methods

We also ran traditional analyses, crude- and adjusted linear regressions, to compare the results we would have got against the proposed TMLE approach returns. We only ran the traditional analyses on a subset small subset of all the analyses. The subset was one where we observed significant results. The adjusted linear regression used the same covariates that were provided to SuperLearner at the corresponding time point. Thus, in the end we compared the three estimates: ψ_{a_1,a_2}^{UNADJ}, ψ_{a_1,a_2}^{ADJUST}, ψ_{a_1,a_2}^{TMLE}.

For the linear regression, the standard output for categorical exposures is the coefficients for the change from baseline level to the corresponding level of exposure. Therefore the compared results only correspond to the differences between baseline exposure, lowest level of exposure, against the middle ($\psi_{2,1}$) and high exposure levels ($\psi_{3,1}$), respectively. For the estimation of the standard errors and corresponding p-values, we used the robust estimates. In the TMLE case we used the already defined IC and the Wald Test p-value. In the case of the linear regression, we used the "robcov" function from the Design package in R which uses the Huber-White method to adjust the variance-covariance matrix of a fit from maximum likelihood. In our particular case, the corrected covariance matrix returned is the "sandwich" estimator.[28, 39]

3.3 Results

We split the analyses into two groups, creatinine unadjusted and creatinine adjusted. In both groups we observe a small number of significant results, even before adjusting for the multiple testing. For each of the two result groups, we ran a total of 225 models and tested 3 differences within each one of them, leaving us with 675 tests. The full results are in Appendix B, where Tables B.1 - B.16 show the results for the analyses without adjusting the urine samples for creatinine concentration and Tables B.17 - B.32 present the results for the analyses after having adjusted the exposures for creatinine.

For convenience, we present in Table 3.4 only those effects that remained significant after adjusting for multiple testing. In the cases of the creatinine unadjusted analyses we obtained 54 significant effects, but only 12 remained significant after adjusting for the multiple testing. In the creatinine adjusted models we observed close to twice as many significant

results; 90 before adjusting for multiple testing and 28 after this adjustment. We will go over the individual results by outcome, starting with child weight, followed by child length, child BMI, and child waist circumference.

Table 3.4: Significant effects, after adjusting for multiple testing, of organophosphate (OP) pesticides on mean child outcome by tertiles of exposure. OP exposure is measured by Total DAPs, DEs, and DMs.

Creatinine unadjusted

Outcome		**Exposure**		**TMLE Differences**				**Wald Test p-values**		
Source	Time	Source	Time	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$		$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Weight	24M	DAPs	24M	-0.511	-0.969	-0.457		0.00 ‡	0.00 ‡	0.01
Weight	24M	DEs	24M	-0.464	-0.965	-0.501		0.01	0.00 ‡	0.03
Weight	24M	DMs	12M	-0.321	-0.623	-0.302		0.22	0.00 ‡	0.29
Weight	24M	DMs	24M	-0.408	-0.858	-0.450		0.01	0.00 ‡	0.00
Length	24M	DAPs	24M	-0.74	-1.31	-0.56		0.00 ‡	0.00 ‡	0.04
Length	24M	DEs	24M	-0.46	-1.24	-0.78		0.42	0.00 ‡	0.19
Length	24M	DMs	24M	-0.60	-1.26	-0.66		0.02	0.00 ‡	0.01
BMI	24M	DAPs	24M	-0.37	-0.71	-0.34		0.00	0.00 ‡	0.06
BMI	24M	DEs	24M	-0.39	-0.79	-0.39		0.09	0.00 ‡	0.18
BMI	24M	DMs	24M	-0.31	-0.62	-0.31		0.06	0.00 ‡	0.06

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Before adjusting for multiple testing, we observed 55 effect differences by tertiles of exposure with Wald Test p-values < 0.05; after the adjustment only 17 weight differences remained significant (see Table 3.4). The creatinine unadjusted results showed fewer significant results than the creatinine adjusted analyses, five against twelve, respectively. We describe the creatinine unadjusted significant results first. For Total DAPs exposure we only observe significant differences at the 24 months outcome-exposure combination. The weight differences between the first and second tertiles and the first and third tertiles are -0.511 kg (adjusted p-value= 0.003), and -0.969 (adjusted p-value$<$ 0.001), respectively. The difference between the second and third tertiles was similar in size, -0.457kg, but lost significance after adjusting for the multiple testing. The exposure to DEs at 24 months also shows significant effects on weight at 24 months between the third and first tertiles of exposure; $-0.965kg$ (adjusted p-value$<$ 0.001). The other two differences between tertiles of DEs exposure lost

significance after adjusting for multiple testing. These differences were still the second and third largest differences by tertile, for all DEs exposures on child weight: $-0.464kg$ between second and first tertiles and $-0.501kg$ between third and second tertiles, respectively.

Table 3.5: Significant effects, after adjusting for multiple testing, of organophosphate (OP) pesticides on mean child outcome by tertiles of exposure. OP exposure is measured by Total DAPs, DEs, and DMs.

Creatinine adjusted

Outcome		**Exposure**		**TMLE Differences**				**Wald Test p-values**		
Source	Time	Source	Time	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$		$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Weight	12M	DAPs	6M	-0.134	-0.296	-0.162		0.01	0.00 ‡	0.03
Weight	12M	DAPs	12M	-0.081	-0.224	-0.144		0.55	0.00 †	0.21
Weight	24M	DAPs	DL	-0.182	-0.363	-0.181		0.60	0.00 ⋆	0.58
Weight	24M	DAPs	6M	-0.241	-0.476	-0.235		0.00	0.00 ‡	0.07
Weight	24M	DAPs	12M	-0.464	-0.636	-0.172		0.01	0.00 ⋆	0.22
Weight	24M	DAPs	24M	-0.420	-0.822	-0.402		0.04	0.00 ‡	0.01
Weight	24M	DEs	24M	-0.462	-0.922	-0.460		0.00 ⋆	0.00 ‡	0.00 †
Weight	42M	DEs	24M	-0.411	-0.954	-0.544		0.33	0.00 ‡	0.15
Weight	24M	DMs	12M	-0.278	-0.501	-0.223		0.05	0.00 ‡	0.04
Weight	42M	DMs	12M	-0.694	-0.457	0.238		0.00 ‡	0.01	0.17
Length	12M	DAPs	6M	-0.27	-0.51	-0.24		0.01	0.00 ‡	0.13
Length	12M	DAPs	12M	-0.16	-0.49	-0.33		0.57	0.00 †	0.19
Length	24M	DAPs	24M	-0.87	-1.54	-0.67		0.00	0.00 ‡	0.00
Length	42M	DAPs	12M	-0.74	-1.25	-0.51		0.03	0.00 †	0.07
Length	12M	DEs	6M	-0.40	-0.84	-0.43		0.15	0.00 ‡	0.02
Length	24M	DEs	24M	-0.55	-1.25	-0.70		0.17	0.00 ‡	0.08
Length	42M	DEs	24M	-0.48	-0.90	-0.42		0.28	0.00 ‡	0.33
Length	24M	DMs	24M	-0.59	-1.20	-0.60		0.04	0.00 ‡	0.04
BMI	24M	DAPs	6M	-0.29	-0.55	-0.26		0.00	0.00 ‡	0.07
BMI	24M	DEs	24M	-0.36	-0.70	-0.34		0.04	0.00 ‡	0.03
BMI	42M	DEs	24M	-0.26	-0.55	-0.29		0.25	0.00 ‡	0.16
BMI	24M	DMs	DL	-0.20	-0.37	-0.17		0.54	0.00 ⋆	0.56
BMI	24M	DMs	12M	-0.24	-0.42	-0.17		0.12	0.00 ⋆	0.14
BMI	42M	DMs	12M	-0.66	-0.12	0.54		0.00 ‡	0.42	0.00 ‡

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

DMs exposure at 12 and 24 months showed significant differences in child weight at 12 and 24 months, but only between the first and third tertiles of exposure. With 12 months DMs, only the difference between extreme tertiles of exposure was significant even before the multiple testing adjustment, $-0.623kg$ (adjusted p-value= 0.007). For 24 months DMs, all three differences were significant before adjusting for multiple testing, but only the one between the lowest and highest levels remained significant, $-0.858kg$ (adjusted p-value$<$ 0.001). The other two differences were $-0.408kg$ between the first and second tertiles and $-0.450kg$ between the second and third tertiles, respectively. The creatinine adjusted analyses show significant differences on child weight at more time points than 24 months only. Total DAPs at 6 and 12 months, between the first and third tertiles of exposure, have a significant negative effect on child weight at 12 months. The effect from 6 months Total DAPs is $-0.296kg$ (adjusted p-value$<$ 0.001) and from 12 months exposure is -0.224 (adjusted p-value$==$ 0.022). Total DAPs exposure, between the lowest and highest levels, also show negative effects on child weight measured at 24 months. Delivery and 12 months exposures showed nominal significant effects of $-0.363kg$ (adjusted p-value= 0.081) and $-0.636kg$ (adjusted p-value= 0.091), respectively. Total DAPs exposures at 6 and 24 months showed negative significant effects on child weight of $-0.476kg$ (adjusted p-value= 0.007) and $-0.822kg$ (adjusted p-value$<$ 0.001), respectively. Effects from DEs exposure at 24 months on weight measured also at 24 months, was the only group of effects that remained, at least nominally, significant for all exposure level differences after adjusting for multiple testing. The effects between first and second tertiles were $-0.462kg$ (adjusted p-value= 0.081); between first and third tertiles were $-0.922kg$ (adjusted p-value$<$ 0.001); and between second and third tertiles were $-0.460kg$ (adjusted p-value= 0.040). Additionally, we observed a significant child weight difference at 42 months between the lowest and highest levels exposure from DEs at 24 months; $-0.954kg$ (adjusted p-value$<$ 0.001). Exposure to DMs only showed two significant weight differences after adjusting for multiple testing. Both significant differences corresponded to 12 months exposure, but weight measured at 24 and 42 months. The former corresponded to the difference between lowest and highest tertiles of exposure, $-0.501kg$ (adjusted p-value= 0.003) and the latter to the difference between the first and second tertiles of exposure, $-0.694kg$ (adjusted p-value= 0.002). At 42 months outcome, the weight difference between the first and third tertiles of exposure, $-0.457kg$, lost significance after adjusting for multiple testing. Additionally, this particular exposure-outcome combination presented a "U" shape effect with respect to exposure, showing a decrease in weight between the first and second tertiles of exposure, but an increase when DEs exposure increased from the second to the third tertile.

For the length outcome, before adjusting for multiple testing, we had 58 statistical significant differences between tertiles of exposure for all three exposure measures: Total DAPs, DEs, and DMs. After adjusting for multiple testing, only twelve differences remained significant. Eleven of these corresponded to the difference between the highest and the lowest tertiles of exposure. In the creatinine unadjusted analyses, all four significant differences

corresponded to exposure at 24 months and child length also at 24 months of age. We observe the only significant difference between the first and second tertiles of exposure under Total DAPs exposure; all others correspond to differences between extreme tertiles of exposure. Child length differences between low-middle and low-high Total DAPs exposure were $-0.74cm$ (adjusted p-value= 0.003) and $-1.31cm$ (adjusted p-value$<$ 0.001), respectively. The effects of DEs and DMs exposures at 24 months, between lowest and highest exposure levels, were $-1.24cm$ (adjusted p-value$<$ 0.001) and -1.26cm (adjusted p-value= 0.002), respectively. The creatinine adjusted analyses presented also significant effect differences with 24 month exposure on 24 month child length. The effects are $-1.54cm$ (adjusted p-value$<$ 0.001) for Total DAPs, $-1.25cm$ (adjusted p-value$<$ 0.001) for DEs, and $-1.20cm$ (adjusted p-value= 0.001) for DMs. The latter was the only significant effect for DMs exposure. However, Total DAPs and DEs also showed significant effects on child length measured at 12 months and 42 months. Both exposures showed a significant effect from 6 months exposure on 12 months outcome with negative effect of $-0.51cm$ (adjusted p-value$<$ 0.001) for Total DAPs and $-0.84cm$ (adjusted p-value= 0.009) for DEs. Total DAPs exposure at 12 months showed a significant negative difference in child length at 12 months, between the lowest and highest tertiles, of $-0.49cm$ (p-value= 0.01). The last significant effect from Total DAPs occurred at the 12 months exposure and child length measure at 42 months, with a difference of $-1.25cm$ (adjusted p-value= 0.019) between the two extremest levels of exposure. DEs exposure at 24 months showed a significant effect child length at the 42 months with a decrease in length of $-0.90cm$ (adjusted p-value$<$ 0.001). All other TMLE length differences between tertiles of exposure are close the null, being less than one unit (1 cm), but show almost consistently a negative effect on exposure's increase.

For BMI, we observed a total of 26 significant effects before adjusting for multiple testing; Wald Test p-values < 0.05. Table 3.4 shows the eight effects that remained statistically significant, and two more were nominally significant with an adjusted p-value < 0.1, after adjusting for the multiple testing. In the creatinine unadjusted analyses, we only observe significant effects with 24 months exposure and child BMI measure at 24 months. The significant effects only showed between the highest and lowest tertiles of exposure as expected. The differences are $-0.71kg/m^2$ (adjusted p-value= 0.003), $-0.79kg/m^2$ (adjusted p-value= 0.013), and $-0.62kg/m^2$ (adjusted p-value= 0.036) for Total DAPs, DEs, and DMs, respectively. For the creatinine adjusted analyses we have more significant differences, but only DEs exposure coincided with the previous results at 24 month exposure and 24 month outcome, with an effect difference between extreme tertiles of exposure of $-0.70kg/m^2$ (adjusted p-value$<$ 0.001). This exposure-outcome effect was also significant for Total DAPs, but only before adjusting for multiple testing. Most significant differences occur between the highest and lowest tertiles of exposure. However, we also have two highly significant differences between the first and second tertiles ($-0.66kg/M^2$ (adjusted p-value$<$ 0.000)), and between the second and third tertiles ($0.54kg/M^2$ (adjusted p-value$<$ 0.000)) for DMs exposure at the 12 months and child BMI at 42 months. These results suggest a "U" shape

effect from DMs exposure on the child's BMI. Except for this case, we see increasing negative effects of the exposure on child BMI.

Child's waist circumference measured at five years of age only showed five significant differences between tertiles of exposure before adjusting for multiple testing; this was for both, creatinine- unadjusted and adjusted, analyses. After adjusting for multiple testing all findings lost their significance. In Tables B.16 and B.32 we observe that the differences in child's waist circumference for the tertiles of exposure did not show a consistent direction relative to exposure increase. The largest effect difference that we observe is only 2.3cm. A waist circumference this size is less than a 4% difference given the range of waist circumference measured in our population.

Table 3.6: Estimated effect differences from tertiles of creatinine adjusted **Total DAPs** exposure on child's weight at 24 months; p-values shown inside parentheses. For the crude and adjusted regression, the estimated values correspond to the coefficients from the linear models with exposure defined as factors corresponding to the tertiles of exposure, and the first tertile of exposure defined as baseline. Difference between second and third tertiles of exposure is the difference of the coefficients and hence no p-value is provided. TMLE estimates were extracted from Table **??**.

Total DAPs on Child Weight (kg) @ 24 months

Unadjusted Regression				
Exposure Time	$\psi_{2,1}^{UNADJ}$		$\psi_{3,1}^{UNADJ}$	
PN	-0.131	(0.597)	-0.038	(0.879)
DL	-0.290	(0.249)	-0.419	(0.100)
6M	-0.209	(0.420)	-0.574	(0.027)
12M	-0.667	(0.008)	-0.645	(0.011)
24M	-0.121	(0.628)	-0.669	(0.008)

Adjusted[1] Regression				
Exposure Time	$\psi_{2,1}^{ADJUST}$		$\psi_{3,1}^{ADJUST}$	
PN	-0.156	(0.524)	-0.469	(0.017)
DL	-0.213	(0.376)	-0.523	(0.009)
6M	-0.199	(0.421)	-0.390	(0.060)
12M	-0.674	(0.005)	-0.444	(0.025)
24M	-0.412	(0.086)	-0.481	(0.014)

TMLE				
Exposure Time	$\psi_{2,1}^{TMLE}$		$\psi_{3,1}^{TMLE}$	
PN	-0.040	(0.868)	-0.027	(0.854)
DL	-0.182	(0.604)	-0.363	(0.001) ⋆
6M	-0.241	(0.003)	-0.476	(0.000) ‡
12M	-0.464	(0.008)	-0.636	(0.001) ⋆
24M	-0.420	(0.040)	-0.822	(0.000) ‡

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months
1: Covariates are defined in Table 4.2 in the column for 24M outcome.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table 3.6 presents the comparative results between unadjusted linear regression, adjusted linear regression and the TMLE G-computation estimation we proposed. As we mentioned before, this comparison only considered a small extract from all the analyses. Given that the creatinine adjusted analyses of Total DAPs on child weight were the ones that showed the most significant effect differences, we decided to use these for our comparisons. In particular we only replicated the analyses for child weight measured at 24 months. A similitude between all three classes of analyses is that child weight is negatively affected by increasing exposure in almost all cases. However, if we had performed the unadjusted linear regression or the adjusted linear regression for all the cases we performed the TMLE, none of the results for the first two analyses would have remained significant after adjusting for multiple testing. The effect differences from TMLE are not always larger than the estimates from the other two methods, what could explain the statistical significance even after adjusting for multiple testing. Actually, the effect difference for Total DAPs at 6 months on child weight is smaller in TMLE than in the unadjusted linear regression, $-0.476kg$ and $-0.574kg$, respectively.

3.4 Discussion

In our analyses we used TMLE and a G-computation approach to estimate the effects of organophosphate pesticides, assessed by urinary dialkyl phosphate metabolites, on child growth because we believe this is the best way to provide unbiased and accurate estimates of our parameter of interest. We only found few significant effects, and all of them corresponded to outcomes measured at 12, 24, and 42 months. Outcomes measured at 6 months or 5 years did not show significant effects from exposure at any time point. All significant effects supported the hypothesis that exposure to organophosphate pesticides has a negative effect on child's growth measured as weight, length and BMI. Our analyses failed to show any negative effect of OP pesticide exposure on child waist circumference measured at 5 years of age.

There are no results available about the effects of organophosphate pesticide exposure on child growth after delivery. Therefore, it is impossible to compare our results with any other study. All the studies we could find analyzing OP pesticides exposure focused on maternal exposure and fetal growth.[18, 19, 52, 60] A recent study by Barr et al. (2010), again only presents results of exposure to pesticides, measured in maternal and umbilical cord sera, and their relation to birth outcomes. [4]

A possible weakness of our analyses is the source of the exposure measurement, urine. Other studies of the effects of organophosphate pesticide exposure measure the pesticides in blood instead of measuring the metabolites in urine. [52] One advantage of the blood measurements is that they provide a direct measure of exposure to parent compounds and may more accurately reflect the dose.[5, 46] However, the dialkyl phosphate metabolites reflect

exposure to about 80% of the organophosphate pesticides used in the Salinas Valley [17], even if not all organophosphate pesticides, like acephate, devolve into these urinary metabolites. Although the source of exposure is not traceable by measurement of the non specific metabolites, they represent an excellent and integrated measure of exposure to a whole class of pesticides. Another important reason for measuring these non specific metabolites of organophosphate pesticides in urine was that at the time of the initial study by Eskenazi et al. [18] there were no analytical methods to measure specific exposure to many important organophosphate pesticides in urine or in blood. For consistency and ease, we considered that measurement of dialkyl phosphate metabolites was an adequate measure to characterize and integrate exposure to multiple organophosphate pesticides that come from different sources.

We find intriguing that for the creatinine unadjusted exposure, only exposure measured at 24 months showed significant effects on child's growth, measured also at 24 months. At this time, we ignore why these effects are only present at 24 months of age. According to our analyses, 24-months exposure lacked significant in effects on later outcomes, either at 3.5 or 5 years. Given the characteristics of our population, it is very unlikely that there will be another study with a similar set up in order to verify our results. It is also worth noting that the majority of the significant effects that we observed were not just borderline significant, but highly significant even after adjusting by a stringent method like Bonferroni's adjustment.

All significant effects were in the *a priori* hypothesized direction; this is that an increase in OP pesticides exposure harms child growth. However, our population is very different from the average children population at age 3.5 years. For example, the mean weight for girls and boys in our group was 17.878kg and 17.267kg, respectively. These mean weights are above the 95th percentile for girls and above the 75th percentile for boys, compared to the national US weight distribution for children the same age. At age five years, the mean weights for girls and boys in our cohort are 21.928kg and 21.997kg, respectively; both values just below the 90th percentile of the national distribution.[13] On the other hand, the height of the children, boys and girls, is just around the national mean. Therefore, the mean BMI for girls at age 24 months is $17.5kg/m^2$ (above the 75th percentile); at age 3.5 years is $17.8kg/m^2$ (close to the 95th percentile); and at age 5 years is again $17.8kg/m^2$ (close to the 95th percentile). For boys the situation is similar, with a mean BMI at age 24 months of $17.2kg/m^2$ (below the 75th percentile); at age 3.5 years it is $17.4kg/m^2$ (close to the 90th percentile); and at age 5 years it is $18.0kg/m^2$ (just above the 95th percentile). Given that we are dealing with an extreme population, where obesity is the norm and not the exception, the true effects of OP pesticides on child growth could be actually underestimated.

Relative to the comparison of the traditional methods, crude and adjusted, against the TMLE we propose, we believe this is the best example of how traditional methods return

biased results and are inefficient. This latter fact is shown by the lac of significant effects after adjusting for multiple testing. While some of the effects were also detected by the traditional methods, their significance disappeared, while TMLE remained significant even after the stringent adjustment using Bonferroni's method.

Finally, we expect that future data from the same cohort study is analyzed for the same effects of OP pesticides exposure on child growth as the CHAMACOS study continues interviewing and evaluating its children as they grow older. Data should be analyzed as it becomes available using the same techniques presented in this document. A possible modification to the analyses could be the way we simplified the exposure from a continuous variable to tertiles of exposure. It would be feasible to extend the categories of exposure, maybe to five or ten. We considered that any truly significant effects would be noticeable with the split defined as low, medium and high exposures. The strength of this study is the use of machine learning techniques with the emphasis of implementing TMLE to focus on the effects of OP pesticides on the outcome of interest only, without getting lost on the definition of a model. Considering the number of models estimated, 450, it would have been illogical to assume that all models had the same form *a priori* and performing a model selection by hand for each explored association would not have been possible either. Machine learning for semi-parametric estimation in such high-dimensional studies is the best way to proceed.

Chapter 4

Semi-parametric estimation of longitudinal effects of environmental exposures: simulation of intervention effects on infant weight at 3.5 years of age

4.1 Introduction

The Environmental Protection Agency's Pesticide Program reported that more than 1.2 billion pounds of pesticide active ingredients are used every year in the United States, with approximately 700 million pounds used in agriculture. [34] California is the State with the largest agricultural output. One major agricultural area of the State is the Salinas Valley in Monterey County, which is often referred to as the "Nation's Salad Bowl". There is evidence of widespread pesticide exposure to all the population in the U.S., including pregnant women and children. [1, 8, 9, 31, 40, 20] There have been several studies which have analyzed the effects of *in utero* OP pesticide exposure and fetal growth. [19, 52, 60, 18] However, these studies have shown conflicting results. For example, Berkowitz et al. (2004) found that maternal levels of chlorpyrifos above the limit of detection coupled with low maternal PON1 activity were associated with a significant but small reduction in head circumference. The pesticide metabolite levels alone were not associated with any of the fetal growth indices. Perera et al. (2003) found that, in residents of upper Manhattan, New York, increasing levels of OP pesticide chlorpyrifos in umbilical cord blood were associated with decreased birth weight and birth length but not with head circumference. Eskenazi et al. (2004) failed to demonstrate an adverse relationship between fetal growth and any measure of *in utero* OP pesticide exposure as measured by dialkyl phosphate metabolites in urine. Furthermore,

they found an increase in body length and head circumference associated with some exposure measures.

Although there are not studies in humans, there is some evidence of postnatal weight gain in rats that were chronically exposed to OP pesticides. [38, 42] Meggs and Brewer (2007) found that rats in the exposed group were significantly heavier than those in the control group and the increase occur concretely in adipose tissue. Lassiter and Brimijoin (2007) found that male rats, exposed between gestational day 7 and till the end of lactation, showed excess weight gain and larger volumes than control rats 45 days after birth and that the differences increased with time.

The purpose of the present study is to investigate the longitudinal effect of OP pesticides exposure on child's weight at 3.5 years of age. These children are participants in a longitudinal birth cohort where information about their exposure to organophosphate pesticides, weight, and other covariates were measured at several time points during this period. Our particular question of interest is to estimate the impact on the mean weight of this population that an hypothetical intervention, where exposure was reduced to its minimum at all time points, would have compared to the mean weight given the actual levels of exposure. The estimation of this particular parameter of the intervention-specific counterfactual distribution is identified by the so called G-computation formula. The parameter we estimated was simple, but at the same time it is considered the gold standard of any intervention. It is unreasonable to ask what would have been the effect between the highest and the lowest levels of exposure because we will not try to expose the children to the highest levels ever, for health and ethical reasons, but we will certainly try to lower their exposure to the lowest levels possible. For the targeted estimation, we considered a degenerate initial estimator of the intermediate conditional distribution. This way we only needed to target the final step at the outcome, child weight at 3.5 years, given the history of covariates and history of exposures.[72]

4.2 Materials and methods

Participants

The population under analysis participated in the longitudinal birth cohort study of the Center for the Health Assessment of the Mothers and Children of Salinas (CHAMACOS), which is conducted by the Center for Children's Environmental Health Research at the University of California, Berkeley. The study focuses on the effects of pesticides and other environmental exposures on the health of pregnant women and their children living in the Salinas Valley, CA. The population and its characteristics have been previously described in

detail by Eskenazi et al. [18].

The women and their children have been followed and interviewed after birth on pre-established time intervals at 6, 12, 24 months, and 3.5 years. During these interviews, trained, bilingual and bicultural personnel conducted extensive and detailed questionnaires about the child's health, family status, and life events during the time since the last interview. The children biometrics were carefully measured and recorded. Biological samples, blood and urine, were also collected.

Pesticide exposure measurement

The exposure to organophosphate (OP) pesticides was assessed by measuring non-specific organophosphate dialkyl phosphate metabolites in urine. Maternal urine was collected during pregnancy (twice) and at delivery. The first and second samples during pregnancy occurred at a mean of 13 weeks gestation (range, 4-29 weeks) and at 26 weeks (range, 18-39 weeks), respectively. Spot urine samples were collected from the children at each of the subsequent interviews after birth. Urine specimens were aliquoted and stored at -80° C until shipment to the Center for Disease Control and Prevention (CDC; Atlanta, GA) for analysis of dialkyl phosphate metabolite levels. [18]

Six dialkyl phosphates metabolites were measured in the urine samples using gas chromatography and mass spectrometry and quantified using isotope dilution calibration.[10] The dialkyl phosphates measured were dimethylphosphate, dimethyldithiophosphate, dimethylthiophosphate, diethylphosphate, diethyldithiophosphate, and diethylthiophosphate. The six metabolites are grouped into dimethyl phosphates and diethyl phosphates. Approximately 80% of the organophosphate pesticides used in the Salinas Valley devolve to one or more of these metabolites, which are excreted in urine. The most commonly used pesticides in this region that devolve to dialkyl phosphates are presented in Table 4.1

Table 4.1: Dialkyl phosphate pesticides metabolites and parent compounds

Marker of exposure	Parent compounds or class
Dialkyl phosphate metabolites (nmol/L)	
Dimethyl phosphates	Malathion, oxydemeton-methyl dimethoate, naled, methidathion
Diethyl phosphates	Diazinon, chlorpyrifos, disulfoton

Quantities of the six metabolites were converted to molar concentration (nmols per liter)

allowing us to add them. From these elements we defined our three measures of exposure: total concentrations of dialkyl phosphate metabolites (Total DAPs), total concentration of diethyl phosphate metabolites (DEs), and total dimethyl phosphate metabolites (DMs). The two maternal prenatal sample were averaged to create a single prenatal exposure value.

It is a standard procedure to adjust for creatinine concentration when exposure is measured in urine samples. Creatinine gives a reference on how diluted the samples are. The concentrations were determined using a commercially available diagnostic enzyme method (Vitros CREA slides; Ortho Clinical Diagnostics, Raritan, NJ). Samples with creatinine levels $< 10 mg/dL$ were considered too diluted for accuracy of analysis and therefore excluded.

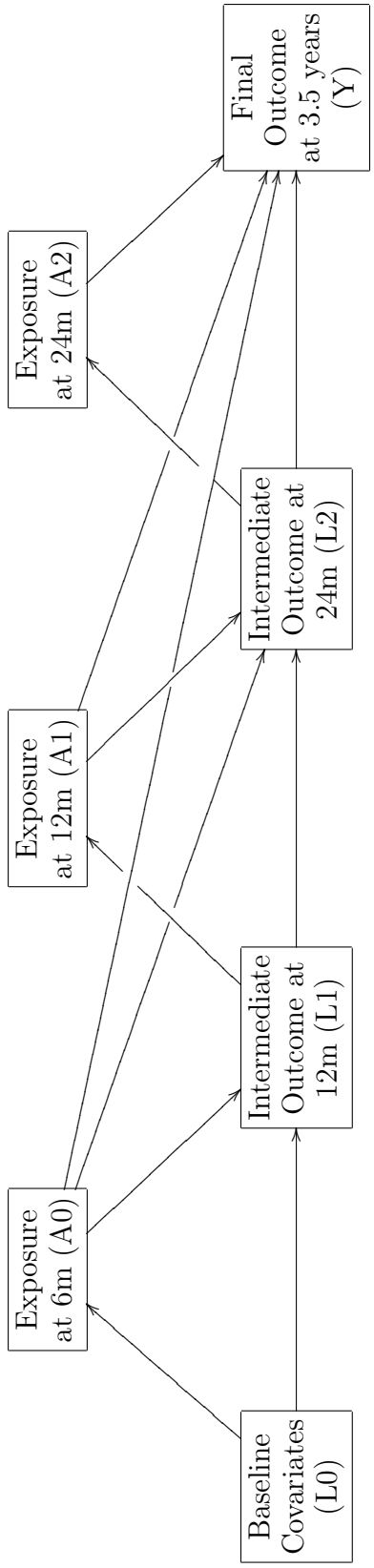


Figure 4.1: Causal diagram for longitudinal effects of organophosphate (OP) pesticides on child weight at 3.5 years of age. Exposure is measured in child at times 0, 1, and 2 corresponding to 6, 12, and 24 months of age. Intermediate measures of weight ($L(t)$) are also recorded at 12 and 24 months of age.

4.2.1 Parameter of interest

Figure 4.1 shows the hypothesized directed acyclic graph (DAG) for our analysis.[62] The main outcome is child's weight at 3.5 years of age. However, child weight was also collected at each of the preceding time points: delivery, 6 months, 12 months, and 24 months.

The complete technical details supporting our estimations can be found in van der Laan (2010).[71, 72] Here, we only present the basic elements needed to perform this estimation. We define the observed data as $O = \big(L(0), A(0), L(1), A(1), L(2), A(2), Y = L(3) \big) = \big(\overline{A}, \overline{L}(\overline{A}) \big) \sim P$. $\overline{L}(\overline{A})$ is the observed covariate process over all time points up to the outcome Y. It is assumed that $L(t)$ occurs before $A(t)$, and we are interested in the effect of interventions on the A-nodes of this graph. In particular, our parameter of interest is the marginal effect difference between the child exposure fixed at its lowest level at each time point, $\overline{A} = \overline{0}$, and the actual exposure levels on the observed outcome at 3.5 years, Y .

$$\psi(\overline{0}) = E\big[Y\big] - E\big[Y(\overline{0})\big] \tag{4.1}$$

For simplicity, we used the definition in van der Laan (2010), to estimate the G-computation at a degenerate initial estimator. This method assumes that the initial estimator provides deterministic predictions of the intermediate time-dependent covariates, so that the clever covariate for all intermediate factors equals zero. Therefore, the targeted Maximum Likelihood Estimation (TMLE) only involves updating the conditional distribution of the final node Y, given its predecessors.[72]

Estimators

We targeted our initial estimate of the OP pesticides effect on child's weight holding exposures at their lowest level. TMLE carries out a targeted bias reduction specifically for the parameter of interest. For complete technical and theoretical details about this general estimating approach we refer readers to the seminal paper by van der Laan and Rubin.[76]

Below we present the targeted element for the parameter of interest. Remember that the other element necessary to calculate our parameter of interest does not need to be targeted since it is the mean child weight of the observed data at the actual levels of exposure. We will only target the weight of the children holding the exposure level low at all three time points. TMLE steps:

1. Estimate the conditional expectation of Y given $\overline{A}$, and $\overline{L}(\overline{A})$; denoted by

$$E\big[Y \mid \overline{A}, \overline{L}(\overline{A})\big] = \gamma\Big(Q\big(\overline{A}, \overline{L}(\overline{A})\big)\Big) = Q_n^0\Big(\overline{A}, \overline{L}(\overline{A})\Big), \tag{4.2}$$

   where γ is the link function.

2. Estimate the conditional distribution of the exposure $\overline{A}$ given covariates $\overline{L}(K)$ and previous exposures $\overline{A}(K-1)$, where K is the end point. In order to estimate $g(\overline{A} \mid \overline{L}(2))$, we factorize it into three treatment times and fit each of them separately. For instance:

$$\begin{aligned} g(\overline{A} \mid \overline{L}(2)) &= \prod_{t=0}^{2} g\left(A(t) \mid \overline{L}(t), \overline{A}(t-1)\right) \\ &= g\left(A(0) \mid L(0)\right) \times g\left(A(1) \mid A(0), L(0), L(1)\right) \times g\left(A(2) \mid A(0), A(1), L(0), L(1), L(2)\right) \end{aligned} \tag{4.3}$$

3. Calculate a specific covariate for each individual target based on the subjects observed values $\overline{L}(\overline{A})$, $\overline{A}$ and the estimate $g_n^0\big(A(K) \mid \overline{L}(K), \overline{A}(K-1)\big)$. This new covariate, whose form depends on the parameter of interest and the model of $Y \mid \overline{A}, \overline{L}(\overline{A})$, is denoted by $h(\overline{A}, \overline{L})$ and sometimes referred to as a "clever" covariate. In our case

$$h(\overline{A}, \overline{L}(\overline{A})) = \frac{I\big(\overline{A} = (0,0,0)\big)}{g\big(\overline{A} \mid \overline{L}(2)\big)} \tag{4.4}$$

4. Update the initial regression $Q_n^0\big(\overline{A}, \overline{L}(\overline{A})\big)$ by adding the clever covariate $h(\overline{A}, \overline{L})$ and estimating its corresponding coefficient by simple maximum likelihood, holding $Q_n^0\big(\overline{A}, \overline{L}(\overline{A})\big)$ fixed at their initial values by using it as offset. The updated regression is denoted $\gamma(Q^1) = \gamma(Q^0 + \epsilon h)$.

5. Implement the estimator by using Q^1 to predict with all $\overline{A} = \overline{a}$ and keeping the original values for $\overline{L}\big(\overline{a} = (0,0,0)\big)$.

Specifically, the TMLE of $\delta(\overline{0})$:

$$\delta^{TMLE}(\overline{0}) = \frac{1}{n} \sum_{i=1}^{n} Q_n^1(\overline{0}, \overline{L}_i). \tag{4.5}$$

TMLE is consistent if at least one of the two functions, Q and g, is consistently estimated. Additionally, the estimator is locally efficient in the sense that it is efficient if both models are estimated consistently.
We need to estimate both, Q_n^0 and g_n^0, and their models need to be correctly specified for the estimating procedure. We will not use parametric models specified *a priori*. Instead, we will use a data-adaptive model selection approach, a machine learning approach, which has the property of potentially approaching a non-parametric model as $n \to \infty$. There are many such procedures available, like the Deletion/Substitution/Addition (D/S/A) algorithm, Least Angle Regression [16], Random Forest [11], Support Vector Machine, Generalized Additive Models (GAM), and Polychotomous Regression[35]. Each one of these

procedures returns an "optimal" model, not all of them identical, based on their own optimality criteria.

For example, the D/S/A algorithm performs data-adaptive estimation through selection of the estimators based on heavy use of cross-validation and the $L2 - loss$ function. The candidate estimators will always be polynomials which comply to user-specified constraints like maximum number of terms in the polynomial, maximum power of polynomial terms and maximum order of interaction.[68] Another example is gam, which replaces the linear form of the covariates ($\sum \beta_i X_i$) by a sum of smooth functions ($\sum s_i(X_i)$), where the functions $s_i(\cdot)$ are unspecified and are estimated using an iterative procedure called local scoring algorithm, resulting in a nonparametric regression method.[29] A third example is Polychotomous Regression, which fits a regression model using linear splines and their tensor products.[21, 35]

Above we mentioned a list of model selectors which we will refer to as "candidate learners". We can use them to define the models we need for our estimations. Van der Laan et al. (2007) propose an algorithm, the SuperLearner, to select the best possible model using a convex combination of the resulting models from each of the candidate learners. The resulting model will be as good or better than every single model from the candidate learners.[74]

4.3 Data analysis

We limited our analysis to those participants with complete records. This is, those whose exposure had been evaluated and their weight measured at each of the considered time points. Our final sample size was 265 children. In those cases were one of the six metabolites was not readable because of analytic interference, the missing value was imputed using simple regression analysis on the other metabolites within the same group (i.e. diethyl or dimethyl phosphates, respectively) and the same time point. This imputation was justified because of the high correlation of the metabolites within groups by time point. Metabolites missing because their levels were below the limit of detection (LOD) were given the value of the LOD divided by the square root of two. [32]

We measured child exposure at 6, 12, and 24 months as well as the child weight at birth, 6, 12, 24 months, and 3.5 years of age - the end point of our analysis. To estimate our parameter of interest, we needed to specify three different models. From Figure 4.1, The first model corresponded to the categorical child weight at 12 months ($L1$) given the baseline covariates ($L0$) and the child exposure at 6 months ($A0$). The second model corresponded to the categorical child weight at 24 months ($L2$) given baseline covariates, child exposures at 6 and 12 months($A1$), and the previous categorical weight at 12 months. Finally, we estimated the model for child weight, as a continuous variable, at 3.5 years (Y) given the

baseline covariates, child exposures at 6, 12, and 24 months($A2$), and the intermediate categorical weights at 12 and 24 months.

We categorized OP pesticide exposure into three levels: 0 =low, 1 =medium, and 2 =high. This procedure simplified our calculations and still allowed us to obtain the parameter of interest fixing the exposure at the desired level. We also categorized the intermediate child weights at 12 and 24 months into low-, medium- and high weight. Simulating the intermediate weights at the desired exposure level in the continuous case is a much harder problem. It implies estimating the full conditional density instead of only estimating the probability of being at each weight level given the exposure and covariates. Child weight at 3.5 years was left as a continuous variables since no simulation of it was needed for a later step.

Baseline covariates were defined as maternal exposure (measured during pregnancy and delivery), child weight (delivery and 6 months), and covariates recorded up to the 6 months interview. Both, birth and 6 months weights were also left as continuous variables because no simulation was performed with them. The full list of covariates at each time point can be found in Table 4.2. Some of the covariate were time dependent and therefore recorded at each of the interviews.

Table 4.2: Covariates used in the different model selection procedures

	Model		
Covariate	L1	L2	Y
Sex	✓	✓	✓
Age in months	✓	✓	✓
Birthweight	✓	✓	✓
Infant's chest circumference at delivery	✓	✓	✓
Gestational age	✓	✓	✓
Breastfeeding status	✓	✓	
Breastfeeding length	✓	✓	✓
Husband's agricultural worker	✓	✓	✓
Maternal smoking status	✓	✓	✓
Child around smokers	✓	✓	✓
WIC	✓	✓	✓
Poverty status at baseline	✓	✓	✓
Poverty status at the time	✓	✓	✓
Child attended child care	✓	✓	✓
Mother's height	✓	✓	✓
Mother's pre pregnancy weight	✓	✓	✓
Mother's pre pregnancy BMI	✓	✓	✓
Maternal exposure during pregnancy	✓	✓	✓
Maternal exposure at delivery	✓	✓	✓
Maternal education	✓	✓	✓
Paternal education	✓	✓	✓
Years in the US at baseline	✓	✓	✓
Maternal calories consumption at 26wks	✓	✓	✓
Diet Quality Index for Pregnancy at 26wks	✓	✓	✓
Child weight at 6M (continuous)	✓	✓	✓
Categorical child weight at 12M		✓	✓
Categorical child weight at 24M			✓
Child attended preschool at 42M			✓
TV hours/weekday			✓
TV hours/weekend day			✓
Hours/day child played outside in past week			✓

There were missing values among the covariates at the different time points. We decided to impute those values because we did not want to lose more observations and we considered that imputing them would not substantially affect or influence our results in a strong way. The imputations were performed in R using the multiple imputations package "mi".[23, 58] The strength of this package is that it performs an iterative regression imputation of the missing values until approximate convergence is achieved. In our case, the matrix of covariates could be split into those covariates with missing observations, M with columns $M(1), \ldots, M(K)$, and those covariates with complete observations, C. First, the missing values of M are imputed using a crude approach (for example, imputing by randomly selecting from the observed outcomes for that variable). Then the algorithm continues imputing $M(1)$ given $M(2), \ldots, M(K)$ and C; imputing $M(2)$ given $M(1), \ldots, M(K)$ and C; and so forth, randomly imputing each variable and looping until approximate convergence. Additionally, the fact that these covariates were considered to be potentially included in the models does not imply that the machine learning processes used actually selected them as part of their optimal model.

Table 4.3: Age range and distribution of the children at each of the planned interviews

Planned Interview		distribution of age in months						
6 months	actual age:	5	6	7	8	9	10	11
	n:	69	129	38	15	7	4	3
12 months	actual age:	11	12	13	14	15	16	17
	n:	76	122	35	14	8	7	3
24 months	actual age:	23	24	25	26	27	28	29
	n:	90	122	32	11	5	3	2
42 months	actual age:	41	42	43	44	45	46	≥ 47
	n:	54	114	38	34	7	6	12

Table 4.3 shows the age range of the children for each of the interview cycles. The data in this order violates the time-ordering assumption because, for example, the exposure measurement of a 12-month old child will clearly have occurred before the weight measurement of a 13-month old child, but we assumed in our model that the exposure measured at the 12-months interview happened after the weight measured at the same interview, for all children. Even though for most children we had the weight at the exact age of the planned interview, we could not adjust the other weights to this age. Having done so would have violated the time ordering assumption for those who were interviewed ahead of time. Therefore, we

decided to adjust the weights to the earliest age within each interview cycle.

We performed a linear interpolation procedure to set all child weights at the lowest age of the corresponding interview. The interpolated values corresponded to 5 months of age at the 6-months interview, 11 months of age at the 12-months interview, 23 months of age at the 24-months interview, and 41 months of age at the 42-months or 3.5 years interview. The interpolating points were the unadjusted weights at the immediately preceding interview and the weight record to be adjusted. This is a valid estimation since children of the ages under study follow almost a linear piece-wise pattern within short periods of time, as can be seen from the available national growth charts. [13] Finally, we remark that even though these adjustments modify our original data, and in particular our outcome variable, it was not possible to adjust the exposure values in a consistent way since those cannot easily be estimated or interpolated from their other measurements as the weights are. Dialkyl phosphate metabolites are highly variable and we could not support the validity of any interpolation between two measurements.

We mentioned above that complete analysis required estimating three different models, $L1(A_0, L_0)$, $L2(\overline{A}_1, \overline{L}_1)$, and $Y(\overline{A}_2, \overline{L}_2)$. All three models were estimated using machine learning techniques through the use of SuperLearner.[74] We used four different candidate learners for each one of the model selections: glm, gam, polymars, and D/S/A. The last two candidate learners had to be tailored for each of the model selections. We specified internal parameters to guarantee that the exposure would always be part of the end model. There was a possibility that the exposure was left out of the main model if the candidate learner did not consider it significant for the model under its selection criteria. Forcing the exposure into the models was necessary for later stages when we simulated exposure at the lowest levels.

Since the intermediate child weights were defined as categorical, with more than two levels, we had to specify multinomial models for them. The estimation of these models required us to go over a few more steps because, currently, the SuperLearner package in R only allows model selection of binary or continuous outcomes, where the family is either binomial or gaussian, respectively.[55, 58] Neither of these two families was directly appropriate for the model selection of the intermediate outcomes. We defined an alternative way to obtain the optimal model: the models for 12 and 24 months categorical child weight were estimated by weighting three individual models for each of them. These individual models corresponded to a dummy variable (Z) indicating if the weight corresponded to the low, medium, or high level, respectively. For simplicity of notation we denote the corresponding covariates for each of the two intermediate models with W, not without emphasizing that W is not the same for both time points. We estimated the redefined binary models with SuperLearner using the previously named candidate learners.

$$\begin{aligned}
\text{Model 1}: &\quad \widehat{P}_{low}(I(Z = low) \mid W) \\
\text{Model 2}: &\quad \widehat{P}_{medium}(I(Z = medium) \mid W) \\
\text{Model 3}: &\quad \widehat{P}_{high}(I(Z = high) \mid W)
\end{aligned}$$

$$S_W = \widehat{P}_{low}(I(Z = low) \mid W) + \widehat{P}_{medium}(I(Z = medium) \mid W) + \widehat{P}_{high}(I(Z = high) \mid W)$$

$$\widetilde{P}(Z = i \mid W) = \frac{\widehat{P}_i(Z = i \mid W)}{S_W} \quad \text{where: } i \in \{low, medium, high\}$$

This weighting provided us with three probabilities, which added to one, and corresponded to the chance of being at each one of the weight levels given the sets of covariates. These probabilities were used to simulate the discrete weights later, when we fixed the exposure at its lowest level.

The model selection for the end outcome was also performed using SuperLearner and the same four candidate learners. This estimation did not require additional steps since the outcome was continuous and the package was capable of accommodating it. Nevertheless, we also fine tuned polymars and D/S/A to guarantee that the exposures remained in the final model.

All the models were estimated using the original sample size of 265 observations. However, for the calculation of the parameter of interest we resampled with replacement 10,000 observations. This provided us with an estimate that should be closer to the actual population. From these 10,000 observations, it was straight forward to calculate the mean child weight under their actual exposure and covariates, $E\big[Y\big]$, the first element of our parameter of interest. Then, we conducted the simulation to obtain the child weight if the exposure levels had been low at all three times of exposure. Using the estimated model for the categorical weight at 12 months, we passed the actual covariate values but fixed the 6 month child exposure to "low" for everybody, $A(0) = 0$. This returned a matrix of 10,000 rows and three columns, corresponding to the probability of being at each of the three weight levels at 12 months of age. We used this probabilities and sampled the weight level at 12 months for each child. Using these newly simulated weight levels, the actual covariate values, and the 6- and 12-month exposures fixed at the "low" level ($\overline{A}(1) = (0, 0)$), we calculated the probabilities of being at each of the three weight levels at 24 months of age; new $10,000 \times 3$ matrix. Using the probabilities in the last matrix, we sampled the weight level for the children at 24 months. Using the simulated child weight categories at 12- and 24-months, the fixed exposure levels at the low level ($\overline{A}(2) == (0, 0, 0) = \overline{0}$), and the actual covariate levels, we calculated the weight for the 10,000 children at age 3.5 years. With the simulated weights we

calculated the mean child weight at the desired level of exposure, $\delta(\overline{0})$. The last calculation provided the second element necessary to estimate the mean weight difference between the actual level of exposure and the intervention level.

To improve the estimation of our parameter of interest, we targeted the above simulated estimate. This procedure required additional model estimations for g (see Eq.4.3) which was actually factorized into three models. These models were selected using only polymars. Nevertheless, we performed the model selection using SuperLearner with a single candidate learner. The target exposure is not a frequent level of exposure in our population; only 6% of the children showed exposure levels being low at all three time points. Once each of these models was estimated, we simulated them at the low level of exposure and obtained our clever covariate for the targeted step, $h(\overline{A}, \overline{L}(\overline{A})$. After performing the full target procedure we updated our initial estimates of the simulation at the level of interest and obtained the new targeted values, $\delta^{TMLE}(\overline{0})$. We calculated the mean targeted weight and the mean targeted difference between the child weights at the actual exposure levels, $E\big[Y\big]$, and the targeted exposure levels, $\psi(\overline{0})$.

We used bootstrap with 1,000 cycles to estimate the inference of our estimates. Each bootstrap cycle involved sampling with replacement 265 observations from our original 265 records. The newly sampled population was used to recalculate the coefficients of the optimal models provided by SuperLearner. The convex combination of the models would be the same as in the original model selection for all cycles. With the updated coefficients and sampling with replacement 10,000 observations from this new population, we re-estimated the parameter of interest and its targeted version once more, just as described above. This process allowed us to build confidence intervals for our estimates, initial and targeted.

The whole analysis was performed again using creatinine adjusted exposures. The methods used were identical and the purpose of it was only a comparative analysis of the two ways to present the exposure given that it was measured in urine samples.

See Appendix C for a simplified guideline algorithm of all the calculations.

4.4 Results

All our results showed negative effects of being exposed to higher levels of OP pesticides. We observe that the actual levels of exposure, out of which only 6% were low all the time, decrease the mean weight of the population compared to the mean weight if all individuals had been exposed to the low levels at all time points. The simulated mean child weight increased under the different conditions of our analyses: initial (before target adjustment), targeted, creatinine unadjusted, and creatinine adjusted. However, none of the differences

between the mean child weight at 3.5 years of age at the actual levels of exposure minus the mean child weight at the simulated level of interest were significant. Table 4.4 shows all the mean weight differences and their p-values; all of them are far from reaching significance.

Except for the case of the creatinine adjusted DMs exposure analysis, last row of Table 4.4, the targeted analyses showed a small shift away from the null compared to the initial estimate equivalents. However, the differences between these two estimates were not significant either (results not shown). It is interesting to notice that the creatinine adjusted analyses for Total DAPs and DMs showed also a shift towards the null. In the case of DEs exposure, the creatinine adjusted values shifted away from the null compared to their creatinine unadjusted versions. The latter are the largest differences we observe; above 0.5kg, but just reached a p-value of 0.6.

Figures 4.2-4.4 present the histograms of the 1,000 cycles of the bootstrap estimates for Total DAPs, DEs, and DMs exposure analyses, respectively. We purposely marked the histograms within each set only with the letters a-d. The intention was to show how similar the results are for the four analyses. The order of the histograms corresponds to the same order of the analyses presented in Table 4.4 by exposure: a)initial - creatinine unadjusted, b)targeted- creatinine unadjusted, c)initial - creatinine adjusted, d)targeted - creatinine adjusted. The red lines in the histograms denote the point estimates from Table 4.4.

Table 4.4: Simulated mean child weight differences in kg between the observed weights at the actual levels of exposure and the weights at the simulated levels, holding exposure low at all time points of exposure. (Walt Test's p-values shown in parenthesis.)

Exposure	**Creatinine**	**Initial estimate** $\psi(\overline{0})$	**Targeted estimate** $\psi^{TMLE}(\overline{0})$
Total DAPs	unadjusted	-0.246 (0.745)	-0.273 (0.734)
	adjusted	-0.117 (0.786)	-0.146 (0.779)
DEs	unadjusted	-0.039 (0.797)	-0.179 (0.772)
	adjusted	-0.512 (0.604)	-0.566 (0.566)
DMs	unadjusted	-0.339 (0.714)	-0.356 (0.706)
	adjusted	-0.169 (0.771)	-0.150 (0.776)

Figure 4.2: Histogram for the mean child weight difference in kg between the actual levels of **Total DAPs** exposure and the level of interest, holding Total DAPs exposure at its low level at all time points. Left and right columns correspond to the initial and targeted estimates, respectively. Upper row and lower row correspond to the creatinine unadjusted and creatinine adjusted exposures, respectively. The purpose of having labeled the four variants of our analyses only with letters a, b, c, and d was to emphasize how all of them show the same consistent result. The red line corresponds to the point estimate and the bars are the result of a 1,000 cycles of bootstrap. The blue dashed line corresponds to the mean of the bootstrap cycles.

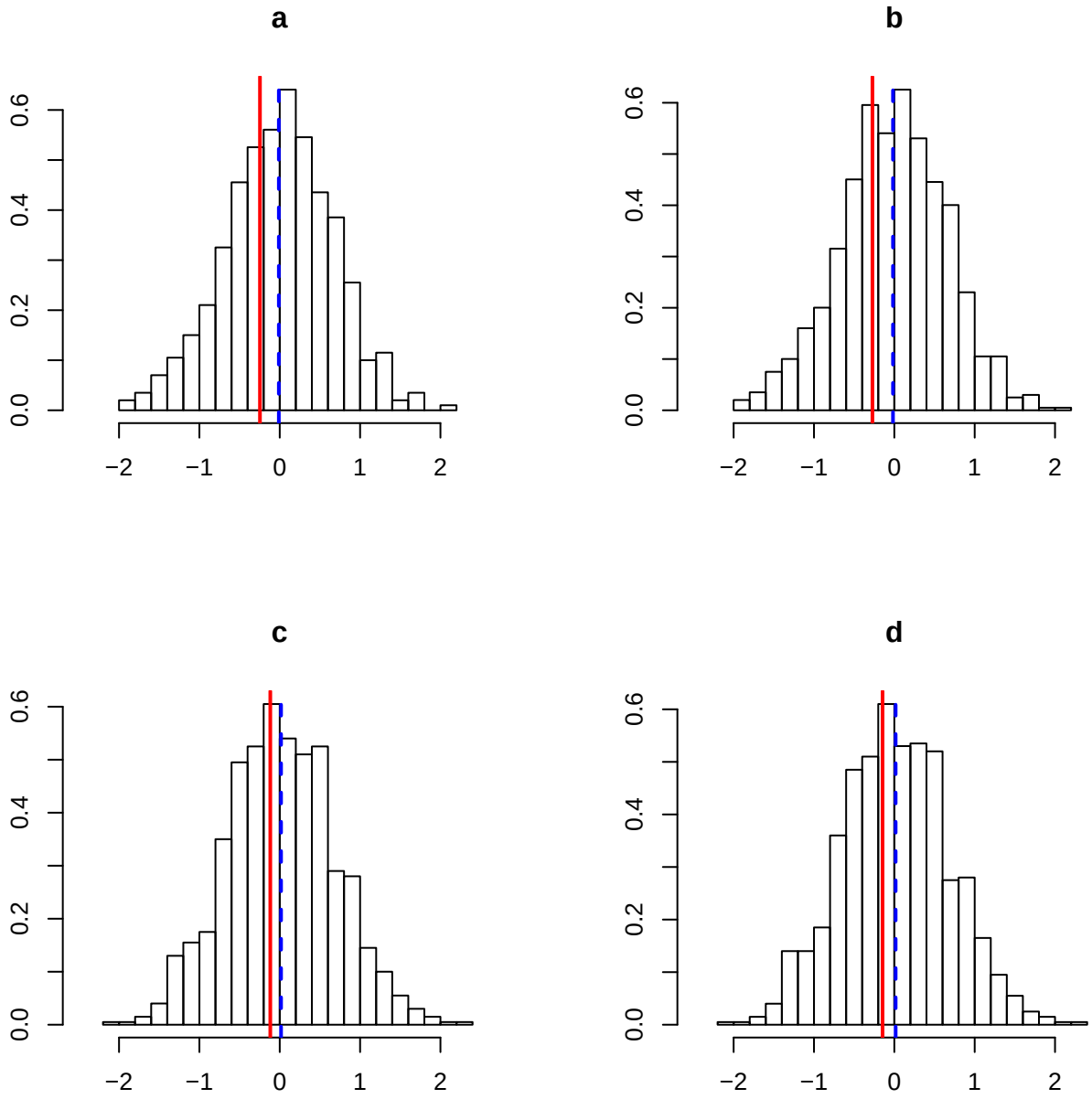


Figure 4.3: Histogram for the mean child weight difference in kg between the actual levels of **DEs** exposure and the level of interest, holding DEs exposure at its low level at all time points. Left and right columns correspond to the initial and targeted estimates, respectively. Upper row and lower row correspond to the creatinine unadjusted and creatinine adjusted exposures, respectively. The purpose of having labeled the four variants of our analyses only with letters a, b, c, and d was to emphasize how all of them show the same consistent result. The red line corresponds to the point estimate and the bars are the result of a 1,000 cycles of bootstrap. The blue dashed line corresponds to the mean of the bootstrap cycles.

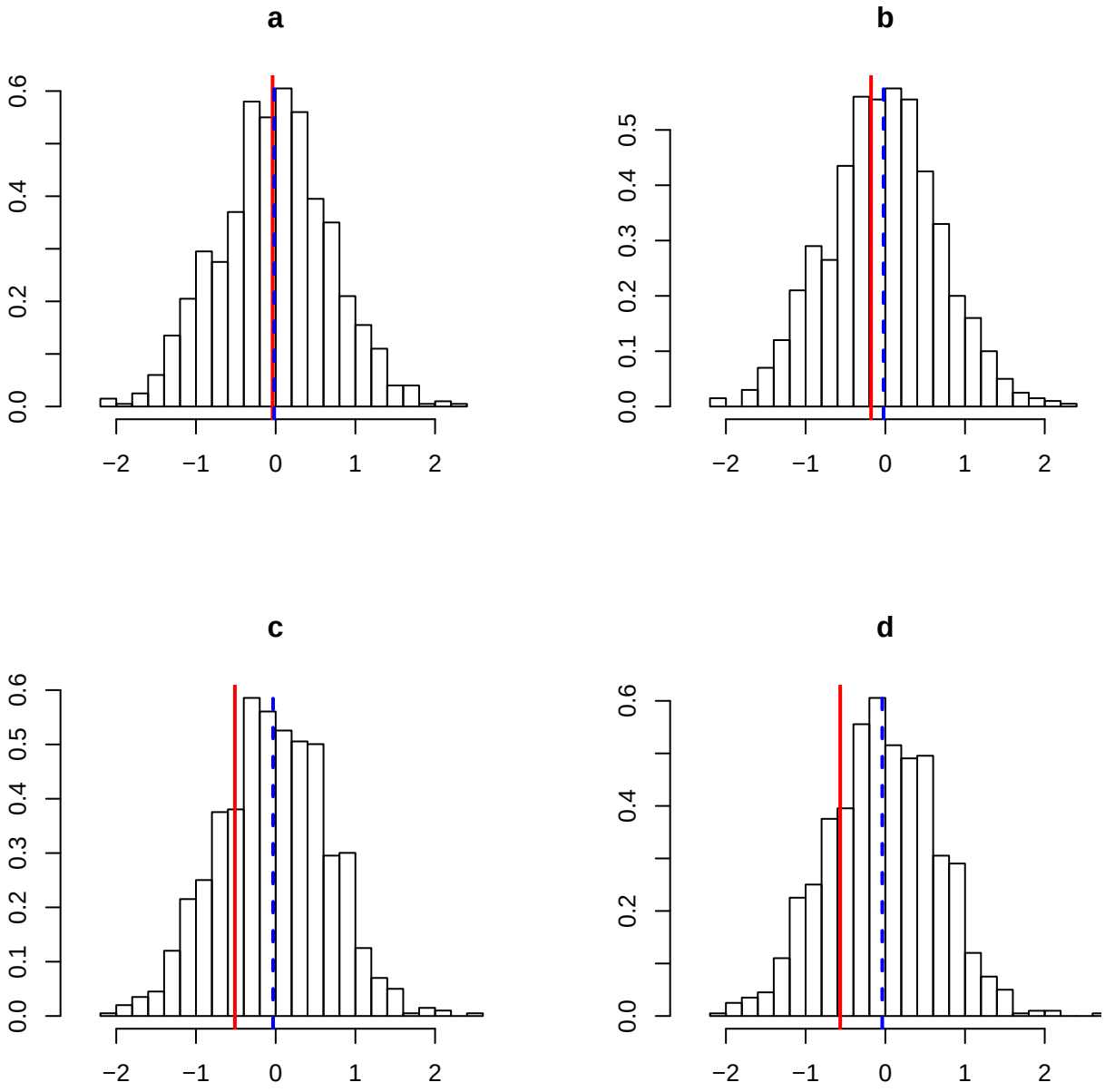
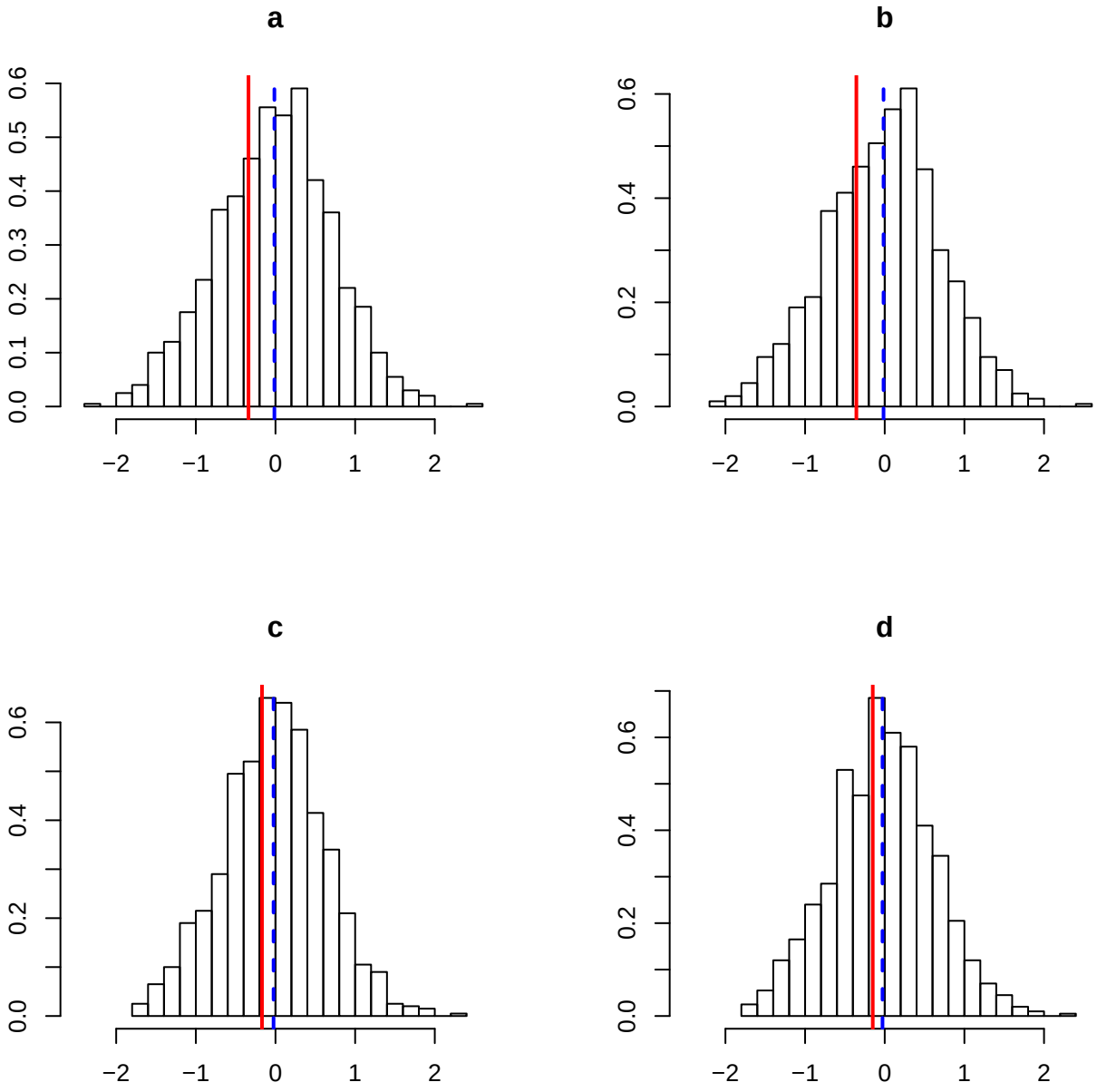


Figure 4.4: Histogram for the mean child weight difference in kg between the actual levels of **DMs** exposure and the level of interest, holding DMs exposure at its low level at all time points. Left and right columns correspond to the initial and targeted estimates, respectively. Upper row and lower row correspond to the creatinine unadjusted and creatinine adjusted exposures, respectively. The purpose of having labeled the four variants of our analyses only with letters a, b, c, and d was to emphasize how all of them show the same consistent result. The red line corresponds to the point estimate and the bars are the result of a 1,000 cycles of bootstrap. The blue dashed line corresponds to the mean of the bootstrap cycles.



4.5 Discussion

In the cross-sectional analyses of OP pesticides exposure and child weight in Chapter 3, there were several significant associations of Total DAPs, DE, and DM on child weight evaluated at 24 months. However, among numerous comparisons, we found that creatinine adjusted DM levels at 12 months and DE levels at 24 months were the only associations with child weight at 3.5 years. There was no evident reason to explain the change in results by age 3.5 years. Therefore, in an effort to understand exposure effects on a longer time period, we decided to define the outcome of interest at 3.5 years instead of 24 months in the present chapter. Looking for effects longitudinally increased the space in which we defined our parameter of interest, diminishing power to fine significant effects.

To our knowledge and given the available literature, this is the first analysis where child exposure to OP pesticides and child weight, as a measure of growth, have been analyzed longitudinally. Having found no significant effects of the current exposure levels on the children's weights at 3.5 years could be interpreted as no adverse risk to their health from this source, but a single study is clearly not conclusive. The observed distributions from the bootstrap procedures have a wide range of effects from -2kg to +2kg, making it also impossible to claim that there is truly no effect. However, we are very confident of our results given their consistency under modified conditions: adjusting for creatinine and targeting the parameter of interest. We do observe a small shift which could be interpreted as a reduction in the bias, and as an improvement from the initial estimates, but all results remained not significant.

This study has been able to collect great amounts of data over the years and it continuously monitored a consistent population for exposure and health effects. We manipulated the collected data to make sure that the exposures and outcomes followed the specified time-ordering assumption. This meant that all exposures did occur before the outcome and not later or even at the same time. For this reason, exposure measured at the 6-months interview was considered as exposure for the child weight at 12 months; the exposure measured at the 12-months interview was defined as the exposure for the child weight at 24 months; and the exposure measured during the 24-months interview was considered to affect the outcome at 3.5 years. This way, weight- and exposure-measurements were defined in a logical time order.

We are aware of some of the weakness of our study. The first one being the source of the exposure measurement. We measured nonspecific metabolites of OP pesticides in urine samples and not their parent compounds in blood. Second, DAPs are a measure of short-term exposure to OP pesticides and may not reflect the "real" exposure over time. Third, DAPs in urine may reflect exposure to preformed DAP metabolites already present in the environment in addition to OP pesticide exposure.[57] Additionally, it is not known how DAP metabolite levels and excretion patterns may differ by PON1 status, even with individ-

uals with similar exposures.[27] Another limitation of this study is that we manipulated the original data in order to comply with the set of assumptions that are needed in longitudinal analyses, e.g., the time ordering assumption.

A further limitation of the study is observational studies like CHAMACOS will always have to deal with participants not being fully compliant to the requirements of the study, that could even end in a loss to follow-up. Figure 4.1 showed the assumed time order for the interviews and the collection of samples. However, the actual times of sample collections were different (see Table 4.3). Even though visits/interviews were scheduled, people would miss their appointments and be weighted at a later time, e.g., children could have been ten months old at the time of their "6-months" interview. There were also situations in which the participants, aware of the coming interview, would go to the CHAMACOS's facilities ahead of time to be evaluated. It was impossible to refuse evaluating them at the risk of them not returning at the precise time. The population under study is a low income farm worker community with a high index of migration, making it extremely hard to follow scheduled appointments with exactitude.

Our results may not generalizable to the rest of the US population. This population had higher exposure to OPs based on DAPs that the general US population because of their close proximity to agricultural activity in the Salinas Valley.[20] The population also belongs to a very low income stratum. In addition, the children in this study were heavier than in the general US population. For example, the children's mean weight, 17.1kg, is over the 80th percentile for children their same age nationwide.[13] The CHAMACOS children had a median weight of 16.2kg, indicating that it is skewed to the right. CDC growth charts show that the nation wide median weight (3th-97th percentiles) for 3.5 year old boys and girls is 15.3kg (12.5kg-19.5kg) and 14.9kg (12.1kg-19.6kg), respectively. In our population 38 children (14%) are over 20 kg at the same age.

The biggest limitation of the implemented method is that it completely depends on the data and it might overfit the selected model. Because the whole process happens in a black box type of procedure, the quality of the output (model) will heavily depend on the quality of the input (data). We need to further investigate the difference between the point estimates and the mean of the bootstrap procedures observed in all the histograms from Figures 4.2-4.4. At this time we ignore what can be causing such difference, given that theoretically both values should coincide. We performed the target procedure to improve the initial estimates of the parameter of interest by reducing the possible bias, even if none of the results reached significance .

The assumed structure for exposures and effects, as depicted in Figure 4.1, should only be considered as a guideline for how we believe the exposure to OP pesticides affects child weight. We used it to establish the order in which we assume these variables interact. In a

traditional setting, this could be considered a strong assumption of the model for the analyses. In our case, on the contrary, this is just the minimum assumption required to perform the machine learning process and the model selection at each time point. It is important to remember that even if we forced certain variables into some of the machine learning algorithms, their coefficients could have been very close to zero when they were considered to be of no importance for the model. We learned in previous experimental runs that the inference on variables, which were not forced into the model selection ended up with a degenerate distribution peaking at zero, corresponding to every cycle where the variable of interest was dropped by the selection algorithm.

We believe that the use of machine learning techniques are adequate for the task at hand. Its strength is that we do not try to define a complete model of the data generating distribution, but instead we focus our efforts in estimating the parameter of interest. Previously, epidemiological analyses would rely on much stronger assumptions like stratification or informal model selection techniques. This erroneous approach delivers erroneous results. Now, we find ourselves at a point of convergence between semi-parametric efficiency theory, machine learning techniques, pathway graph theory, and statistical software like R. Previous double robust methods provided the needed theory for the correct estimation of parameters of interest, but those methods were too hard to implement. There is a learning curve for the implementation of the techniques proposed in this analysis, but they are at the reach of any researcher, with free software that is constantly developed and peer reviewed and tested.[58] Moreover, if for any reason there is a preconceived model that the researcher believes is true, he should include it in the Super Learner as another of the candidate learners and it will be selected if it truly provides information on the data generating distribution. This inclusion will not weaken the proposed method nor affect its efficiency.

For future directions, we believe that our analyses need to be repeated on the same population at later ages as well as on other measures of growth, like length and BMI. Lassiter et al. (2007) found that overweight effects in rats were not noticeable before onset of puberty, at least 95 days after birth in rats. Similar results were observed by Meggs, where chronic exposure to OP insecticide chlorpyrifos had bigger effects on increased weight in rats at older ages.[38, 42] In the CHAMACOS Study, data will continue to be collected and it is possible that longer periods of exposure will have a significant and different effect on children's growth. At this point, our results hint that a higher exposure to OP pesticides causes a decrease in weight, contrary to what has been found in animal studies. The relative early end point of our study, 3.5 years of age, did not allow us to see the long term effects of OP pesticides exposure. We will continue to follow these children through puberty to determine the longer term associations with weight.

Chapter 5

Conclusion

5.1 Summary

The preceding chapters presented statistical methods for the robust estimation of potential parameters of interest over the standard methods in observational studies. The methodology for targeted maximum likelihood estimation was originally proposed in van der Laan and Rubin (2006).[76] Here we have applied this general statistical approach to the estimation of exposure effects in child development. These applications represent a novel contribution to the literature for analyzing data from observational studies.

Chapter 2 provided the natural and controlled direct effect estimation of maternal depression, evaluated by the Center for Epidemiological Studies Depression Scale (CES-D), on infant neurodevelopment, evaluated by the Bayley Scales of Infant Development and the Pre-School Language Scale.[59, 6, 78] In these particular analyses, one of the assumptions was that the main exposure had two pathways in which it affected the outcome of interest, directly and indirectly. The parameter of interest was defined in two ways. First, the direct effect of exposure holding the intermediate pathways at the unexposed levels. Second, the direct effect of exposure holding the intermediate pathways at fixed levels. The results in section 2.5 indicated that maternal depression only had a negative effect on the child's expression development with an a average of -2.8 in the scale. Most important was the comparison of the traditional methods to estimate the parameter of interest to the proposed one. We observed that traditional methods tended to overestimate the effect, showing bias, by at least 50%.

Chapter 3 presented a large set of cross-sectional analyses where the effects of organophosphate pesticides on four measures of child growth were estimated. Exposure was measured at seven different time points and outcomes were recorded at five time points. In these analyses, the efficiency and bias reduction of TMLE was shown. In particular, the com-

parison of a fraction of all the test (Table 3.6) showed how traditional methods, crude and adjusted linear regression, fail to provide significant estimates of the parameter of interest. The TMLE approach also proved to be adequate for the search of potential effects in high dimensional studies like the one under analysis. However, being this the first analysis of its kind, we remind the readers that further verification of the findings is needed and that the fact that effects were mainly present when outcomes were evaluated at 24 months, raises questions about the validity of the findings.

Chapter 4 followed naturally from Chapter 3 even if presented as an independent analysis. After an extensive cross-sectional search, the next logical step, when the data is available and its structure allows it, is the longitudinal analysis of the data. Just like it occurs with standard analysis techniques, TMLE also encounters more technical challenges in its implementation in the longitudinal scenario. However, we were able to present a simplified, yet adequate, analysis using TMLE based on the technical specifications from van der Laan (2010).[72] No analyses conducted in this study resulted in significant associations. Being this the same population under study as in Chapter 3, the longitudinal structure of the data increases the sample space, reducing power to detect potential significant effects. Additionally, the extensive pre-analysis preparation of the data could have also negatively influence our capability to detect longitudinal effects from OP pesticide exposure on child weight. Finally, we repeat that further analyses are required and in particular over the same population as data becomes available.

5.2 Directions for future research

The methods discussed in this dissertation can be applied to many epidemiological studies in useful ways. Under the analysis of Ioannidis (2005), most of the published research findings are false. This problem is due to many causes, but one of them is certainly the erroneous analysis of the collected data.[33] The studies should focus and clearly identify their parameter of interest before jumping into the definition of arbitrary parametric models, which will provide a biased answer to the original question.

The expected treatment assumption (ETA), required for the estimations conducted in all chapters, states that each exposure levels has positive probability of being observed within each strata of covariates. In the present dissertation *ad hoc* methods were implemented to correct practical violations of this assumption. However, as me briefly mentioned before, collaborative targeted maximum likelihood estimation can be applied.[73] Essentially, in this latter approach, covariates are only included in the treatment (exposure) mechanism fit if they improve the targeting of the parameter of interest while not heavily affecting the mean square error (MSE).

Robust methods for estimating associations and causal effects are required since traditional methods fail in this task. The goal is not just to gain efficiency and power, but perhaps to reduce bias.

. . .

In summary, this dissertation presents a novel contribution to the literature for methods to estimate effects in observational studies using targeted maximum likelihood estimation (TMLE). The analyses presented here suggest that these methods are in fact a useful addition to the current set of tools most commonly used, and should become standard practice in the future.

Bibliography

[1] J Adgate, D Barr, C Clayton, L Eberly, N Freeman, P Lioy, L Needham, E Pellizzari, J Quackenboss, A Roy, and K Sexton. Measurement of children's exposure to pesticides: analysis of urinary metabolite levels in a probability-based sample. Environmental Health Perspectives, 109(6):583:590, June 2001.

[2] J Anderson, B Johnstone, and D Remley. Breast-feeding and cognitive development: a meta-analysis. American Journal of Clinical Nutrition, 70(4):525:535, 1999.

[3] R Baron and D Kenny. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6):1173:1182, 1986.

[4] DB Barr, CV Ananth, X Yan, S Lashley, JC Smulian, TA Ledoux, P Hore, and MG Robson. Pesticide concentrations in maternal and umbilical cord sera and their relation to birth outcomes in a population of pregnant women and newborns in new jersey. Science of the Total Environment, 408:790:795, 2010.

[5] DB Barr, JR Barr, WJ Driskell, RH Jr Hill, DL Ashley, LL Needham, SL Head, and EJ Sampson. Strategies for biological monitoring of exposure for contemporary-use pesticides. Toxicology in Health, 15(1-2):168:179, 1999.

[6] N Bayley. Bayley Scales of Infant Development. The Psychological Corporation, San Antonio, TX, second edition, 1993.

[7] O Bembom and M van der Laan. Estimating the effect of vigorous physical activity on mortality in the elderly based on realistic individualized treatment and intention-to-treat rules. U.C. Berkeley Division of Biostatistics Working Paper Series, (Working Paper 217):16, May 2007. http://www.bepress.com/ucbbiostat/paper217.

[8] G Berkowitz, J Wetmur, E Birman-Deych, J Obel, R Lapinski, J Godbold, I Holzman, and M Wolff. In utero pesticide exposure, maternal paraoxonase activity, and head circumference. Environmental Health Perspectives, 112(3):388:391, March 2004.

[9] M Bradman, M Harnly, W Draper, S Seidel, S Teran, D Wakeham, and R Neutra. Pesticide exposures to children from california's central valley: results of a pilot study. Journal of exposure analysis and environmental epidemiology, 7(2):217:234, 1997.

[10] R Bravo, W Driskell, R Whitehead Jr., L Needham, and D Barr. Quantitation of dialkyl phosphate metabolites of organophosphate pesticides in human urine using gcmsms with isotopic internal standards. Analytical Toxicology, 26:245:252, 2002.

[11] L Breiman. Random forest. Machine Learning, 45:5:32, 2001.

[12] B M Caldwell and R Bradley. Home observation for the measurement of the environment. Little Rock, Arkansas: University of Arkansas at Little Rock, 1984.

[13] Center for Disease Control and Prevention, 1600 Clifton Rd. Atlanta, GA 30333, USA. Individual Growth Charts. http://www.cdc.gov/growthcharts/charts.htm.

[14] AW Chan, A Hrobjartsson, MT Haahr, PC Gotzschw, and DG Altman. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. Journal of American Medical Association, 291(20):2457:2465, September 2004.

[15] L M Dunn and L M Dunn. Norms booklet for the peabody picture vocabulary test: Third edition. Circle Pines, MN: American Guidance Service, 1997.

[16] B Efron, T Hastie, I Johnstone, and R Tibshirani. Least angle regression. Annals of Statistics, 32(2):407:451, 2004.

[17] California EPA. Pesticide use reporting - 2008 summary data. Technical report, CA: Department of Pesticide Regulation, January 2010. [accessed 19 July 2010].

[18] B Eskenazi, K Harley, A Bradman, E Weltzien, N Jewell, D Barr, C Furlong, and N Holland. Association of in utero organophosphate pesticide exposure and fetal growth and length of gestation in an agricultural population. Environ Health Perspect, 112(10):1116:1124, 2004.

[19] L Fenster and M Coye. Birthweight of infants born to hispanic women employed in agriculture. Archives of Environmental Health, 45:46:52, 1990.

[20] National Center for Environmental Health. Fourth national report on human exposure to environmental chemicals. Technical report, National Center for Environmental Health, 2009.

[21] J Friedman. Multivariate adaptive regression splines. The Annals of Statistics, 19(1):1:141, 1991.

[22] J Galler, R Harrison, F Ramsey, V Forde, and S Butler. Maternal depressive symptoms affect infant cognitive development in barbados. Journal of Child Psychology and Psychiatry, 41(6):747:443, 2000.

[23] Andrew Gelman, Jennifer Hill, Yu-Sung Su, Masanao Yajima, and Maria Grazia Pittau. Missing Data Imputation and Model Checking. Columbia University, June 2010. http://www.stat.columbia.edu/ gelman/.

[24] World Health Organization Geneva. Global strategy for infant and young child feeding. World Health Organization, 20 Avenue Appia, 1211, Geneva 27, Switzerland, 2003.

[25] S Gruber. Target maximum likelihood estimation: A gentle introduction. U.C. Berkeley Division of Biostatistics Working Paper Series, page 15, 2009. http://www.bepress.com/ucbbiostat/paper252.

[26] D Hann, K Winter, and P Jacobsen. Measurement of depressive symptoms in cancer patients: Evaluation of the center for epidemiological studies depression scale (ces-d). Journal of Psychosomatic Research, 46(5):437:443, 1999.

[27] K Harley, K Huen, R Aguilar Schall, N Holland, A Bradman, D Barr, and B Eskenazi. Association of pesticide exposure and paraoxonase with birth outcome in mexican-american women. Environmental Health Perspectives, Submitted.

[28] F Harrell. Design Methods and Generic Functions. Department of Biostatistics, Vanderbilt University. version 2.3-0.

[29] T Hastie and R Tibshirani. Generalized additive models. Statistical Science, 1(3):297:318, 1986.

[30] Trevor Hastie. gam: Generalized Additive Models, 2008. R package version 1.0.

[31] R Hill, S Head, S Baker, M Gregg, DB Shealy, SL Bailey, C Williams, E Sampson, and L Needham. Pesticide residues in urine of adults living in the united states: reference range concentrations. Environmental Research, 71:99:108, 1995.

[32] RW Hornung and LD Reed. Estimation of average concentration in the presence of nondetectable values. Applied Occupational Environmental Hygiene, 5:46:51, 1990.

[33] J Ioannidis. Why most published research findings are false. PLoS Medicine, 2(8):696:701, August 2005.

[34] T Kiely, D Donaldson, and A Grube. Pesticide industry sales and usage 2000 and 2001 market estimates. Technical report, Biological and Economic Analysis Division Office of Pesticide Programs, Office of Prevention, Pesticides, and Toxic Substances U.S. Environmental Protection Agency Washington, DC 20460, May 2004.

[35] C Kooperberg, S Bose, and C Stone. Polychotomous regression. Journal of the American Statistical Association, 92(437):117:127, 1997.

[36] Charles Kooperberg. polspline: Polynomial spline routines, 2009. R package version 1.1.3.

[37] Michael S Kramer and Ritsuko Kakuma. The optimal duration of exclusive breastfeeding: a systematic review. World Health Organization, 20 Avenue Appia, 1211, Geneva 27, Switzerland, 2002.

[38] T Lassiter and S Brimijoin. Rats gain excess weight after developmental exposure to the organophosphorothionate pesticide, chlorpyrifos. Neurotoxicology and Teratology, 30:125:130, 2008.

[39] DY Lin and LY Wei. The robust inference for the cox proportional hazards model. Journal of the American Statistical Association, 84(408):1074:1078, 1989.

[40] C Lu, RA Fenske, NJ Simcox, and D Kalman. Pesticide exposure of children in an agricultural community: evidence of household proximity to farmland and take home exposure pathways. Environmental Research, 84:290:302, 2000.

[41] D MacKinnon, A Fairchild, and M Fritz. Mediation analysis. Annual Review of Psychology, 58:593:614, 2007.

[42] W J Meggs and K L Brewer. Weight gain associated with chronic exposure to chlorpyrifos in rats. Toxicology Investigations, 3(3):89:93, 2007.

[43] K Moore. Targeted Maximum Likelihood Estimation of Treatment Effects in Randomized Controlled Trials and Drug Safety Analysis. PhD thesis, University of California, Berkeley, 2009.

[44] K Moore and M van der Laan. Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. Statistics in Medicine, 28:39:64, 2009.

[45] L Murray. The impact of postnatal depression on infant development. Child Psychology and Psychiatry, 33(3):543:561, 1992.

[46] LL Needham, DL Ashley, and DG Jr Patterson. Case studies of the use of biomarkers to assess exposures. Toxicology Letters, 82-83:373:378, 1995.

[47] R Neugebauer and M van der Laan. Why prefer double robust estimates? illustration with causal point treatment studies. U.C. Berkeley Division of Biostatistics Working Paper Series, (Working Paper 115):23, 2002. http://www.bepress.com/ucbbiostat/paper115.

[48] Romain Neugebauer and James Bullard. DSA: Data-Adaptive Estimation with Cross-Validation and the D/S/A Algorithm. R package version 3.1.3.

[49] J Neyman. On the application of probability theory to agricultural experiments. Statistical Science, 5:465:480, 1990. Translated from the original paper in Polish, 1923.

[50] J Pearl. Causal diagrams fro empirical research. Biometrika, 82(4):669:710, 1995.

[51] J Pearl. Direct and indirect effects. Proceedings of the American Statistical Association Join Statistical Meetings, page 1572:1581, 2005.

[52] FP Perera, V Rauh, W Tsai, P Kinney, D Camann, D Barr, T Bernert, R Garfinkel, Y Tu, D Diaz, J Dietrich, and RM Whyatt. Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. Environmental Health Perspectives, 111(2):201:205, February 2003.

[53] M Petersen and M van der Laan. Estimation of direct causal effects. U.C. Berkeley Division of Biostatistics Working Paper Series, page 27, 2005. http://www.bepress.com/ucbbiostat/paper190.

[54] S Petterson and A Albers. Effects of poverty and maternal depression on early child development. Child Development, 72(6):1794:1813, 2001.

[55] Eric Polley. SuperLearner: Super Learner Prediction, 2009. R package version 1.0.

[56] P J Quinn, M O'Callaghan, G M Williams, J M Najman, M J Andersen, and W Bor. The effect of breastfeeding on child development at 5 years: A cohort study. Journal of Pediatrics and Child Health, 37(5):465:469, 2001.

[57] L Quirs-Alcal, A Bradman, M Nishioka, M Harnly, A Hubbard, and J Ferber. Characterization of pesticide and polybrominated diphenyl etherconcentrations in dust from low-income urban and farmworker households in california. Environmental Science & Technology., Submitted.

[58] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0.

[59] L S Radloff. The ces-d scale: A self-report depression scale for research in the general population. Applied Psychological Measurement, 1:385:401, 1977.

[60] M Restrepo, N Muoz, NE Day, JE Parra, L de Romero, and X Nguyen-Dinh. Prevalence of adverse reproductive outcomes in a population occupationally exposed to pesticides in colombia. Scandinavian Journal of Work, Environment and Health, 16(4):232:238, 1990.

[61] J Robins. A new approach to causal inference in mortality sudies with a sustained exposure period - application to control of the healthy worked survivor effect. Mathematical Modelling, 7:1393:1512, 1986.

[62] J Robins. A graphical approach to the identification and estimation of casual parameters in mortality studies with sustained exposure periods. Journal of Chronic Disease, 40(Suppl. 2):1398:1618, 1987.

[63] J Robins and S Greenland. Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3:143:155, 1992.

[64] J Robins, M Hernán, and B Brumback. Marginal structural models and causal inference in epidemiology. Epidemiology, 11(5):550:560, September 2000.

[65] D Rubin. Bayesian inference for causal effects: The role of randomization. Annals of Statistics, 6:34:58, 1978.

[66] D Rubin. Discussion of "randomization analysis of experimental data in the fisher randomization test" by d. basau. American Statistical Association, 75:591:593, 1980.

[67] D Sharp, D F Hay, S Pawlby, G Schmiicker, H Allen, and R Kumar. The impact of postnatal depression on boys' intellectual development. Journal of Child Psychology and Psychiatry, 36(8):1315:1336, 1995.

[68] S Sinisi and M van der Laan. Loss-based cross-validated deletion/substitution/addition algorithms in estimation. U.C. Berkeley Division of Biostatistics, Working Paper 143:54, March 2004. http://www.bepress.com/ucbbiostat/paper143.

[69] G Taubes. Epidemiology faces its limits. Science, 269:164:169, July 1995.

[70] R Uauy and P Peirano. Breast is best: human milk is the optimal food for brain development. American Journal of Clinical Nutrition, 70(4):433:434, 1999.

[71] M van der Laan. Targeted maximum likelihood based causal inference: Part i. The International Journal of Biostatistics, 6(2):42, 2010. Article 2.

[72] M van der Laan. Targeted maximum likelihood based causal inference: Part ii. The International Journal of Biostatistics, 6(2):31, 2010. Article 3.

[73] M van der Laan and S Gruber. Collaborative double robust targeted penalized maximum likelihood estimation. Working Paper 246, U.C. Berkeley Division of Biostatistics Working Paper Series, http://www.bepress.com/ucbbiostat/paper246, April 2009.

[74] M van der Laan, E Polley, and A Hubbard. Super learner. Statistical Applications in Genetics and Molecular Biology, 6(1):21, 2007. http://www.bepress.com/sagmb/vol6/iss1/art25.

[75] M van der Laan and J Robins. Unified Methods for Censored Longitudinal Data and Causality. Series in Statistics. Springer Verlag, 2003.

[76] M van der Laan and D Rubin. Targeted maximum likelihood learning. U.C. Berkeley Division of Biostatistics Working Paper Series, (Working Paper 213):87, 2006. http://www.bepress.com/ucbbiostat/paper213.

[77] W A Vega, B Kolody, R Valle, and R Hough. Depressive symptoms and their correlates among immigrant mexican women in the unites states. Social Science & Medicine, 22(6):645:652, 1986.

[78] I L Zimmerman, V G Steiner, and R E Pond. Preschool Language Scale. The Psychological Corporation, San Antonio, TX, third edition, 1992.

Appendix A

TMLE of the mean outcome by tertile of exposure

Given the characteristics of our analyses, we wrote a single function to perform the TMLE for all allowed pairs of exposure and outcome. We only needed to pass the desired exposure and outcomes of interest as the input to our estimating function. The use of SuperLearner was critical for the estimations, but the necessary candidate learners had to be specifically defined for our analyses. While SuperLearner provides a series of ready to use candidate learners or libraries, they not need to address the particular needs of every analysis and custom libraries have to be defined. For the estimation of polymars and D/S/A we had to define libraries where the exposure of interest was specifically forced into the model selection. The way these two model selection algorithms work might well leave the main exposure out of the optimal model, leading to inference errors later. Additionally the constrains of the available D/S/A libraries within SuperLearner did not provide us with sufficient degrees of freedom on the degree of the polynomial and this had also to be customized. Next we describe the estimating procedure algorithm in detail.

1. Function parameters:

   (a) the vector of actual outcomes, only one at the time from the four of interest (weight, length, BMI, and waist circumference). The outcome also corresponded to a single time of interest.

   (b) the vector of actual main exposures, only one at the time from the three of interest (Total DAPs, DEs, DMs). The exposure had to comply to the time constrain of having occurred prior or concurrent with the outcome.

   (c) the complete matrix of recorded covariates W corresponding to the time of the outcome. The set of covariates changed depending the time of the outcome.

   (d) the candidate learners for the estimation of Q_0 (glm, gam, Polymars, and D/S/A) and g (step.forward).

2. Obtain the initial model for the outcome given the exposure and the set of covariates using Super Learner and the first set of candidate learners. $Q_0 = Q_0(Y_i \mid A, W)$.

3. Target each of the tertiles of exposure, one at the time.

   (a) With the optimal models, we calculated the counterfactual outcomes at each one of the tertiles of exposure. For example, in the case of length, we calculated the mean length outcome had all participants been exposed at the same tertile, for each level.

   (b) Estimate g using Super Learner again, but with the corresponding candidate learner.

   (c) Update the initial estimate of $\widehat{Q}_0$ through the targeted estimation. $Q_1^* = \widehat{Q}_0 + \epsilon * h$ where $h = \frac{1}{\widehat{g}(a|W)}$

   (d) Calculate the populations mean at the targeted level of exposure.

$$\widehat{E}[Y_1] = \widehat{E}[Q^*(1, W)] = \frac{1}{n}\sum_{i=1}^{n} Q^*(1, W_i)$$

$$\widehat{E}[Y_2] = \widehat{E}[Q^*(2, W)] = \frac{1}{n}\sum_{i=1}^{n} Q^*(2, W_i)$$

$$\widehat{E}[Y_3] = \widehat{E}[Q^*(3, W)] = \frac{1}{n}\sum_{i=1}^{n} Q^*(3, W_i)$$

Appendix B

Complete cross-sectional result tables

Table B.1: TMLE of mean weight by tertiles of Total DAPs exposure at the time indicated on the row. The time of the outcome corresponds to the last time within each block; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Total DAPs on Child Weight (kg) @	**6M**	PN	8.209	8.170	8.169	-0.039	-0.041	-0.001	0.573	0.657	0.984
		DL	8.231	8.164	8.185	-0.067	-0.045	0.022	0.740	0.824	0.886
		6M	8.202	8.154	8.187	-0.048	-0.015	0.033	0.740	0.885	0.826
	12M	PN	10.214	10.137	10.054	-0.077	-0.160	-0.083	0.122	0.002	0.167
		DL	10.210	10.110	10.072	-0.100	-0.138	-0.038	0.444	0.217	0.756
		6M	10.174	10.109	10.037	-0.065	-0.136	-0.072	0.538	0.180	0.588
		12M	10.187	10.127	10.072	-0.060	-0.115	-0.055	0.375	0.047	0.190
	24M	PN	13.103	12.984	12.909	-0.118	-0.194	-0.076	0.317	0.145	0.470
		DL	13.209	12.995	12.783	-0.214	-0.426	-0.212	0.449	0.095	0.563
		6M	13.084	12.965	12.862	-0.119	-0.221	-0.102	0.611	0.179	0.735
		12M	13.252	12.982	12.741	-0.270	-0.511	-0.241	0.017	0.002	0.114
		24M	13.481	12.969	12.512	-0.511	-0.969	-0.457	0.000 ‡	0.000 ‡	0.007

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.2: TMLE of mean weight by tertiles of Total DAPs exposure at the time indicated on the row. The time of the outcome corresponds to the last time within each block; allowing only for exposures prior or concurrent to the time of the outcome.

	Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
		δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Total DAPs on Child Weight (kg) @ 42M	PN	17.719	17.545	17.388	-0.174	-0.331	-0.157	0.513	0.419	0.659
	DL	17.703	17.628	17.466	-0.074	-0.236	-0.162	0.916	0.720	0.793
	6M	17.971	17.550	17.208	-0.421	-0.763	-0.342	0.529	0.167	0.584
	12M	17.793	17.545	17.300	-0.248	-0.493	-0.245	0.688	0.411	0.674
	24M	17.959	17.527	17.365	-0.432	-0.594	-0.162	0.242	0.327	0.761
	42M	17.846	17.515	17.301	-0.331	-0.545	-0.214	0.453	0.330	0.757
Total DAPs on Child Weight (kg) @ 60M	PN	22.508	21.938	21.511	-0.570	-0.997	-0.427	0.016	0.008	0.230
	DL	22.381	21.949	21.555	-0.433	-0.826	-0.394	0.479	0.543	0.762
	6M	22.294	21.964	21.592	-0.330	-0.702	-0.372	0.681	0.298	0.649
	12M	22.093	22.027	21.773	-0.066	-0.321	-0.255	0.935	0.752	0.712
	24M	22.282	22.115	21.793	-0.168	-0.489	-0.322	0.793	0.414	0.611
	42M	22.557	22.018	21.590	-0.539	-0.966	-0.428	0.427	0.170	0.633
	60M	22.020	22.023	21.778	0.002	-0.243	-0.245	0.997	0.724	0.694

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.3: TMLE of mean weight by tertiles of DEs exposure at the time indicated on the row. The time of the outcome corresponds to the last time within each block; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DEs on Child Weight (kg) @	**6M**	PN	8.251	8.182	8.131	-0.070	-0.120	-0.050	0.456	0.198	0.429
		DL	8.147	8.187	8.241	0.040	0.094	0.054	0.581	0.203	0.182
		6M	8.317	8.193	8.096	-0.124	-0.221	-0.097	0.623	0.376	0.009
	12M	PN	10.178	10.142	10.096	-0.036	-0.081	-0.045	0.773	0.396	0.656
		DL	10.142	10.119	10.110	-0.023	-0.031	-0.008	0.728	0.545	0.909
		6M	10.208	10.128	10.001	-0.080	-0.207	-0.127	0.541	0.227	0.542
		12M	10.126	10.112	10.122	-0.014	-0.004	0.011	0.913	0.975	0.930
	24M	PN	13.022	12.990	12.950	-0.032	-0.072	-0.040	0.909	0.803	0.823
		DL	13.029	13.031	12.972	0.002	-0.057	-0.059	0.991	0.499	0.710
		6M	13.065	12.955	12.864	-0.110	-0.201	-0.091	0.584	0.168	0.459
		12M	13.129	12.981	12.847	-0.148	-0.282	-0.133	0.681	0.425	0.500
		24M	13.477	13.013	12.512	-0.464	-0.965	-0.501	0.005	0.000 ‡	0.026

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.4: TMLE of mean weight by tertiles of DEs exposure at the time indicated on the row. The time of the outcome corresponds to the last time within each block; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DEs on Child Weight (kg) @	**42M**	PN	17.441	17.639	17.562	0.198	0.121	-0.077	0.663	0.850	0.909
		DL	17.465	17.607	17.757	0.142	0.292	0.150	0.692	0.500	0.753
		6M	18.104	17.618	17.134	-0.485	-0.970	-0.485	0.646	0.385	0.028
		12M	17.779	17.558	17.233	-0.222	-0.547	-0.325	0.748	0.391	0.637
		24M	18.002	17.524	17.231	-0.478	-0.771	-0.293	0.491	0.006	0.682
		42M	17.657	17.527	17.447	-0.129	-0.209	-0.080	0.718	0.514	0.763
	60M	PN	22.035	21.811	21.813	-0.224	-0.222	0.002	0.777	0.739	0.998
		DL	22.015	22.071	22.167	0.056	0.152	0.096	0.894	0.881	0.917
		6M	22.339	21.961	21.465	-0.378	-0.874	-0.496	0.593	0.247	0.469
		12M	22.418	22.029	21.520	-0.389	-0.899	-0.510	0.546	0.198	0.287
		24M	22.671	21.853	21.491	-0.818	-1.180	-0.362	0.147	0.003	0.627
		42M	22.305	21.971	21.887	-0.335	-0.419	-0.084	0.641	0.542	0.910
		60M	22.380	21.932	21.518	-0.448	-0.863	-0.414	0.763	0.371	0.784

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.5: TMLE of mean weight by tertiles of DMs exposure at the time indicated on the row. The time of the outcome corresponds to the last time within each block; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DMs on Child Weight (kg) @	**6M**	PN	8.187	8.158	8.176	-0.028	-0.011	0.017	0.680	0.926	0.890
		DL	8.209	8.172	8.178	-0.038	-0.031	0.006	0.587	0.820	0.968
		6M	8.119	8.182	8.250	0.063	0.131	0.068	0.325	0.119	0.340
	12M	PN	10.209	10.138	10.038	-0.071	-0.171	-0.100	0.375	0.591	0.770
		DL	10.174	10.104	10.102	-0.070	-0.072	-0.003	0.468	0.266	0.980
		6M	10.041	10.105	10.161	0.064	0.119	0.055	0.611	0.389	0.485
		12M	10.222	10.134	10.052	-0.088	-0.170	-0.083	0.451	0.008	0.382
	24M	PN	13.064	12.992	12.901	-0.073	-0.163	-0.090	0.378	0.387	0.643
		DL	13.203	12.949	12.831	-0.254	-0.372	-0.118	0.186	0.003	0.547
		6M	12.926	13.026	12.978	0.100	0.052	-0.047	0.601	0.823	0.748
		12M	13.300	12.979	12.677	-0.321	-0.623	-0.302	0.219	0.000 ‡	0.294
		24M	13.423	13.015	12.565	-0.408	-0.858	-0.450	0.010	0.000 ‡	0.002

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.6: TMLE of mean weight by tertiles of DMs exposure at the time indicated on the row. The time of the outcome corresponds to the last time within each block; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DMs on Child Weight (kg) @	**42M**	PN	17.561	17.577	17.611	0.015	0.049	0.034	0.939	0.835	0.895
		DL	17.601	17.631	17.482	0.029	-0.120	-0.149	0.961	0.815	0.823
		6M	17.641	17.527	17.514	-0.114	-0.127	-0.013	0.946	0.944	0.985
		12M	17.921	17.566	17.153	-0.355	-0.767	-0.412	0.534	0.254	0.640
		24M	18.108	17.581	17.324	-0.527	-0.784	-0.257	0.302	0.127	0.188
		42M	17.723	17.538	17.391	-0.185	-0.332	-0.147	0.646	0.518	0.822
	60M	PN	22.207	21.966	21.822	-0.240	-0.384	-0.144	0.344	0.414	0.760
		DL	22.459	21.911	21.748	-0.547	-0.710	-0.163	0.509	0.461	0.870
		6M	22.080	21.822	21.832	-0.258	-0.248	0.010	0.787	0.811	0.984
		12M	22.298	22.322	21.595	0.024	-0.703	-0.727	0.981	0.457	0.521
		24M	22.378	22.070	21.676	-0.308	-0.703	-0.395	0.405	0.110	0.119
		42M	22.549	21.837	21.739	-0.712	-0.810	-0.097	0.518	0.314	0.938
		60M	22.162	21.965	21.693	-0.197	-0.469	-0.272	0.570	0.466	0.594

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.7: TMLE of mean length by tertiles of Total DAPs exposure at the time indicated by row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Total DAPs on Child Length (cm) @	**6M**	PN	67.16	67.16	67.19	0.00	0.03	0.03	0.982	0.930	0.903
		DL	67.46	67.19	67.00	-0.27	-0.46	-0.19	0.546	0.396	0.723
		6M	67.15	67.12	67.17	-0.03	0.02	0.05	0.933	0.928	0.881
	12M	PN	74.92	74.81	74.69	-0.11	-0.23	-0.12	0.291	0.099	0.407
		DL	75.12	74.84	74.66	-0.28	-0.47	-0.19	0.346	0.099	0.519
		6M	74.57	74.64	74.78	0.07	0.22	0.14	0.756	0.336	0.556
		12M	74.97	74.80	74.62	-0.18	-0.35	-0.18	0.299	0.015	0.056
	24M	PN	86.49	86.42	86.27	-0.08	-0.22	-0.14	0.704	0.298	0.425
		DL	86.75	86.39	86.06	-0.36	-0.69	-0.33	0.302	0.035	0.476
		6M	86.23	86.37	86.58	0.14	0.35	0.21	0.462	0.237	0.572
		12M	86.65	86.32	86.11	-0.33	-0.54	-0.21	0.158	0.041	0.341
		24M	87.07	86.32	85.76	-0.74	-1.31	-0.56	0.000 ‡	0.000 ‡	0.040

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.8: TMLE of mean length by tertiles of Total DAPs exposure at the time indicated by row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

	Exposure Time	TMLE: Q^1			TMLE Differences			Wald Test p-values		
		δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Total DAPs on Child Length (cm) @ 42M	PN	99.84	99.66	99.56	-0.18	-0.28	-0.10	0.613	0.599	0.807
	DL	100.16	99.65	99.34	-0.52	-0.82	-0.31	0.575	0.397	0.664
	6M	99.71	99.66	99.84	-0.05	0.13	0.19	0.947	0.814	0.800
	12M	100.02	99.60	99.27	-0.42	-0.75	-0.33	0.561	0.460	0.745
	24M	100.17	99.74	99.34	-0.43	-0.83	-0.39	0.436	0.375	0.685
	42M	99.53	99.80	99.50	0.26	-0.03	-0.30	0.664	0.968	0.778
Total DAPs on Child Length (cm) @ 60M	PN	110.83	110.28	109.90	-0.55	-0.92	-0.37	0.018	0.002	0.273
	DL	111.07	110.37	109.69	-0.70	-1.38	-0.68	0.145	0.237	0.545
	6M	110.20	110.28	110.51	0.09	0.31	0.22	0.901	0.601	0.694
	12M	110.51	110.24	109.95	-0.28	-0.56	-0.29	0.668	0.500	0.644
	24M	110.61	110.52	110.20	-0.09	-0.41	-0.32	0.835	0.471	0.570
	42M	110.37	110.33	110.21	-0.04	-0.16	-0.12	0.939	0.787	0.886
	60M	110.35	110.43	110.19	0.08	-0.16	-0.24	0.892	0.805	0.705

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.9: TMLE of mean length by tertiles of DEs exposure at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DEs on Child Length (cm) @	**6M**	PN	67.19	67.15	67.14	-0.04	-0.05	-0.01	0.852	0.813	0.941
		DL	67.08	67.21	67.31	0.12	0.23	0.11	0.397	0.129	0.192
		6M	67.40	67.16	66.98	-0.24	-0.42	-0.18	0.545	0.285	0.041
	12M	PN	74.82	74.80	74.83	-0.02	0.01	0.03	0.942	0.960	0.883
		DL	74.88	74.83	74.84	-0.05	-0.04	0.01	0.690	0.695	0.948
		6M	74.97	74.70	74.32	-0.27	-0.65	-0.38	0.320	0.319	0.592
		12M	74.93	74.73	74.71	-0.20	-0.21	-0.01	0.541	0.423	0.972
	24M	PN	86.51	86.38	86.32	-0.13	-0.20	-0.06	0.748	0.710	0.856
		DL	86.44	86.47	86.34	0.03	-0.10	-0.13	0.907	0.483	0.533
		6M	86.51	86.37	86.19	-0.14	-0.33	-0.19	0.661	0.210	0.316
		12M	86.72	86.42	86.02	-0.31	-0.71	-0.40	0.633	0.252	0.238
		24M	86.93	86.47	85.69	-0.46	-1.24	-0.78	0.422	0.000 ‡	0.187

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.10: TMLE of mean length by tertiles of DEs exposure at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**				**TMLE Differences**				**Wald Test p-values**		
			δ_1	δ_2	δ_3		$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$		$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DEs on Child Length (cm) @	**42M**	PN	99.72	99.74	99.62		0.02	-0.10	-0.12		0.969	0.892	0.871
		DL	99.76	99.71	99.56		-0.05	-0.20	-0.15		0.930	0.727	0.817
		6M	100.45	99.76	98.99		-0.69	-1.46	-0.77		0.498	0.164	0.080
		12M	100.03	99.52	99.22		-0.51	-0.81	-0.30		0.501	0.247	0.687
		24M	100.16	99.67	99.35		-0.50	-0.81	-0.31		0.622	0.004	0.732
		42M	99.68	99.67	99.60		-0.02	-0.08	-0.06		0.971	0.837	0.867
	60M	PN	110.60	110.21	110.06		-0.38	-0.54	-0.16		0.635	0.476	0.783
		DL	110.78	110.42	109.94		-0.36	-0.84	-0.48		0.294	0.260	0.489
		6M	110.94	110.21	109.65		-0.73	-1.29	-0.56		0.330	0.077	0.262
		12M	110.72	110.22	109.80		-0.50	-0.92	-0.42		0.362	0.173	0.316
		24M	110.74	110.46	110.22		-0.28	-0.52	-0.24		0.591	0.288	0.736
		42M	110.52	110.24	110.28		-0.29	-0.24	0.05		0.676	0.680	0.945
		60M	110.20	110.33	110.46		0.13	0.25	0.12		0.908	0.713	0.910

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.11: TMLE mean length by tertiles of DMs exposure at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

DMs on Child Length (cm) @		Exposure Time	TMLE: Q^1			TMLE Differences			Wald Test p-values		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
	6M	PN	67.22	67.15	66.95	-0.08	-0.27	-0.19	0.650	0.471	0.626
		DL	67.41	67.19	67.06	-0.22	-0.35	-0.13	0.167	0.401	0.777
		6M	67.04	67.18	67.27	0.15	0.24	0.09	0.379	0.213	0.658
	12M	PN	74.90	74.79	74.80	-0.11	-0.10	0.00	0.537	0.832	0.994
		DL	75.01	74.79	74.71	-0.22	-0.30	-0.08	0.264	0.040	0.741
		6M	74.33	74.66	74.91	0.33	0.59	0.26	0.251	0.043	0.101
		12M	74.99	74.82	74.61	-0.18	-0.39	-0.21	0.413	0.008	0.195
	24M	PN	86.58	86.37	86.22	-0.21	-0.36	-0.15	0.190	0.144	0.632
		DL	86.67	86.38	86.18	-0.29	-0.49	-0.20	0.280	0.189	0.674
		6M	86.20	86.32	86.61	0.12	0.41	0.29	0.695	0.253	0.309
		12M	86.74	86.35	86.00	-0.39	-0.74	-0.35	0.313	0.012	0.356
		24M	86.96	86.36	85.70	-0.60	-1.26	-0.66	0.023	0.000 ‡	0.007

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.12: TMLE mean length by tertiles of DMs exposure at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**				**TMLE Differences**				**Wald Test p-values**		
			δ_1	δ_2	δ_3		$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$		$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DMs on Child Length (cm) @	**42M**	PN	100.02	99.66	99.46		-0.36	-0.56	-0.20		0.180	0.081	0.508
		DL	99.96	99.61	99.40		-0.35	-0.56	-0.21		0.679	0.333	0.780
		6M	99.17	99.78	100.22		0.61	1.05	0.45		0.484	0.282	0.555
		12M	100.19	99.70	99.11		-0.48	-1.07	-0.59		0.450	0.531	0.748
		24M	100.29	99.67	99.36		-0.62	-0.93	-0.31		0.315	0.146	0.364
		42M	99.54	99.62	99.64		0.08	0.10	0.02		0.878	0.869	0.977
	60M	PN	110.69	110.27	109.92		-0.42	-0.76	-0.34		0.101	0.141	0.488
		DL	111.09	110.35	109.76		-0.74	-1.34	-0.60		0.220	0.112	0.483
		6M	109.83	110.14	110.82		0.31	0.99	0.68		0.635	0.088	0.158
		12M	110.62	110.35	109.90		-0.26	-0.72	-0.45		0.640	0.435	0.648
		24M	110.79	110.45	110.00		-0.34	-0.79	-0.45		0.324	0.060	0.042
		42M	110.33	110.34	110.32		0.01	-0.01	-0.02		0.990	0.988	0.984
		60M	110.59	110.25	110.02		-0.34	-0.56	-0.22		0.308	0.421	0.686

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.13: TMLE of mean BMI by tertiles of Total DAPs exposure at the time indicated by row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Total DAPs on Child BMI (kg/m^2) @	**24M**	PN	17.43	17.34	17.31	-0.09	-0.12	-0.03	0.525	0.457	0.808
		DL	17.50	17.35	17.19	-0.15	-0.31	-0.16	0.610	0.229	0.662
		6M	17.57	17.32	17.12	-0.24	-0.45	-0.20	0.049	0.027	0.389
		12M	17.56	17.34	17.17	-0.22	-0.39	-0.17	0.107	0.027	0.329
		24M	17.71	17.34	17.00	-0.37	-0.71	-0.34	0.003	0.000 ‡	0.056
	42M	PN	17.69	17.60	17.48	-0.09	-0.20	-0.11	0.626	0.463	0.632
		DL	17.54	17.61	17.60	0.07	0.06	-0.02	0.882	0.908	0.969
		6M	17.90	17.54	17.27	-0.35	-0.63	-0.27	0.386	0.122	0.421
		12M	17.59	17.56	17.53	-0.03	-0.06	-0.04	0.952	0.897	0.944
		24M	17.71	17.54	17.58	-0.17	-0.13	0.04	0.584	0.791	0.915
		42M	17.85	17.50	17.38	-0.34	-0.47	-0.13	0.246	0.281	0.807
	60M	PN	18.19	17.88	17.66	-0.31	-0.53	-0.22	0.049	0.007	0.343
		DL	18.01	17.89	17.72	-0.12	-0.29	-0.17	0.745	0.652	0.793
		6M	18.17	17.91	17.56	-0.25	-0.60	-0.35	0.579	0.151	0.409
		12M	17.93	17.98	17.92	0.04	-0.02	-0.06	0.920	0.973	0.884
		24M	18.00	17.92	17.82	-0.08	-0.18	-0.10	0.779	0.579	0.772
		42M	18.34	17.94	17.65	-0.40	-0.69	-0.29	0.313	0.100	0.567
		60M	17.99	17.97	17.78	-0.03	-0.22	-0.19	0.937	0.628	0.645

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.14: TMLE of mean BMI by tertiles of DEs exposure at the time indicated by row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**				**TMLE Differences**				**Wald Test p-values**		
			δ_1	δ_2	δ_3		$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$		$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DEs on Child BMI (kg/m^2) @	**24M**	PN	17.34	17.34	17.31		0.01	-0.02	-0.03		0.977	0.942	0.872
		DL	17.40	17.36	17.32		-0.04	-0.08	-0.04		0.815	0.402	0.788
		6M	17.36	17.37	17.27		0.01	-0.09	-0.10		0.947	0.579	0.422
		12M	17.34	17.32	17.35		-0.02	0.01	0.04		0.944	0.969	0.862
		24M	17.77	17.38	16.98		-0.39	-0.79	-0.39		0.090	0.000 ‡	0.181
	42M	PN	17.46	17.62	17.66		0.16	0.20	0.04		0.545	0.663	0.939
		DL	17.42	17.61	17.82		0.19	0.40	0.21		0.548	0.184	0.550
		6M	17.79	17.59	17.42		-0.20	-0.37	-0.16		0.817	0.676	0.377
		12M	17.65	17.57	17.51		-0.07	-0.14	-0.07		0.882	0.777	0.897
		24M	17.84	17.55	17.38		-0.29	-0.47	-0.18		0.517	0.007	0.687
		42M	17.65	17.56	17.51		-0.09	-0.14	-0.05		0.740	0.555	0.836
	60M	PN	17.81	17.81	17.97		0.01	0.16	0.16		0.987	0.681	0.696
		DL	17.81	17.95	18.11		0.14	0.29	0.16		0.594	0.569	0.740
		6M	17.98	17.86	17.77		-0.12	-0.21	-0.10		0.770	0.658	0.827
		12M	18.18	17.92	17.72		-0.26	-0.46	-0.19		0.574	0.347	0.512
		24M	18.29	17.80	17.66		-0.49	-0.63	-0.15		0.151	0.072	0.767
		42M	18.03	17.90	17.85		-0.13	-0.19	-0.06		0.778	0.657	0.897
		60M	18.22	17.86	17.55		-0.36	-0.67	-0.31		0.668	0.195	0.703

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.15: TMLE of mean BMI by tertiles of DMs exposure at the time indicated by row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
DMs on Child BMI (kg/m^2) @	**24M**	PN	17.36	17.37	17.33	0.01	-0.03	-0.04	0.921	0.843	0.835
		DL	17.52	17.32	17.16	-0.20	-0.36	-0.16	0.232	0.213	0.630
		6M	17.45	17.32	17.24	-0.12	-0.20	-0.08	0.563	0.429	0.620
		12M	17.63	17.29	17.10	-0.34	-0.53	-0.19	0.243	0.009	0.560
		24M	17.69	17.37	17.06	-0.31	-0.62	-0.31	0.064	0.001 ‡	0.059
	42M	PN	17.55	17.60	17.64	0.04	0.09	0.05	0.775	0.631	0.804
		DL	17.50	17.60	17.67	0.11	0.17	0.07	0.821	0.607	0.888
		6M	17.78	17.56	17.42	-0.22	-0.37	-0.14	0.793	0.710	0.777
		12M	17.69	17.58	17.41	-0.12	-0.29	-0.17	0.784	0.423	0.754
		24M	17.77	17.57	17.55	-0.20	-0.22	-0.02	0.566	0.558	0.935
		42M	17.79	17.53	17.41	-0.27	-0.38	-0.12	0.364	0.450	0.839
	60M	PN	17.94	17.93	17.89	-0.01	-0.05	-0.04	0.939	0.837	0.867
		DL	18.00	17.91	17.85	-0.09	-0.15	-0.06	0.852	0.770	0.918
		6M	18.15	17.85	17.63	-0.30	-0.52	-0.22	0.556	0.329	0.519
		12M	17.94	18.02	17.90	0.08	-0.04	-0.12	0.852	0.936	0.834
		24M	18.04	17.92	17.82	-0.11	-0.22	-0.10	0.561	0.355	0.440
		42M	18.34	17.94	17.71	-0.40	-0.63	-0.23	0.511	0.253	0.768
		60M	17.98	17.94	17.78	-0.04	-0.19	-0.15	0.828	0.663	0.676

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.16: TMLE of mean waist circumference at 5 years by tertiles of exposure from Total DAPs, DEs, and DMs at the time indicated by row.

	Exposure		**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
		Time	δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
5Y Waist Circumference (cm)	**Total DAPs**	PN	59.00	58.45	58.01	-0.55	-0.99	-0.44	0.054	0.084	0.451
		DL	58.96	58.38	57.76	-0.58	-1.20	-0.62	0.563	0.483	0.704
		6M	58.95	58.42	57.82	-0.53	-1.13	-0.60	0.687	0.369	0.583
		12M	58.99	58.61	57.80	-0.37	-1.19	-0.82	0.775	0.452	0.463
		24M	58.61	58.73	58.40	0.12	-0.20	-0.32	0.899	0.813	0.724
		42M	59.44	58.49	57.75	-0.95	-1.69	-0.74	0.297	0.030	0.509
		60M	58.92	58.50	57.78	-0.43	-1.14	-0.72	0.594	0.312	0.492
	DEs	PN	58.28	58.34	58.41	0.06	0.13	0.07	0.966	0.908	0.950
		DL	58.58	58.48	58.58	-0.10	0.00	0.10	0.862	1.000	0.940
		6M	58.80	58.44	57.82	-0.36	-0.98	-0.62	0.698	0.362	0.469
		12M	59.33	58.53	57.62	-0.80	-1.71	-0.91	0.373	0.151	0.298
		24M	59.51	58.20	57.70	-1.31	-1.81	-0.50	0.129	0.015	0.673
		42M	58.81	58.46	58.10	-0.35	-0.72	-0.36	0.747	0.374	0.733
		60M	59.15	58.39	57.76	-0.77	-1.39	-0.63	0.725	0.318	0.776
	DMs	PN	58.21	58.53	58.56	0.32	0.35	0.03	0.452	0.618	0.958
		DL	58.92	58.34	58.02	-0.59	-0.90	-0.31	0.593	0.506	0.814
		6M	58.53	58.12	58.43	-0.42	-0.10	0.31	0.757	0.941	0.687
		12M	59.01	58.83	57.85	-0.18	-1.17	-0.98	0.854	0.413	0.518
		24M	58.80	58.58	58.29	-0.23	-0.51	-0.28	0.714	0.470	0.409
		42M	59.06	58.56	57.99	-0.50	-1.07	-0.57	0.724	0.201	0.679
		60M	58.84	58.48	57.96	-0.37	-0.88	-0.51	0.461	0.387	0.535

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.17: TMLE of mean weight by tertiles of Total DAPs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted Total DAPs on Child Weight (kg) @	**6M**	PN	8.204	8.177	8.174	-0.028	-0.030	-0.003	0.662	0.836	0.984
		DL	8.246	8.186	8.128	-0.060	-0.118	-0.058	0.481	0.223	0.519
		6M	8.241	8.183	8.141	-0.058	-0.099	-0.042	0.665	0.524	0.588
	12M	PN	10.190	10.128	10.088	-0.062	-0.101	-0.040	0.473	0.331	0.648
		DL	10.258	10.128	10.027	-0.131	-0.231	-0.101	0.554	0.309	0.173
		6M	10.248	10.114	9.952	-0.134	-0.296	-0.162	0.010	0.000 ‡	0.025
		12M	10.246	10.165	10.021	-0.081	-0.224	-0.144	0.547	0.000 †	0.212
	24M	PN	13.018	12.978	12.991	-0.040	-0.027	0.013	0.868	0.854	0.939
		DL	13.176	12.995	12.813	-0.182	-0.363	-0.181	0.604	0.001 ⋆	0.582
		6M	13.212	12.971	12.736	-0.241	-0.476	-0.235	0.003	0.000 ‡	0.071
		12M	13.351	12.887	12.715	-0.464	-0.636	-0.172	0.008	0.001 ⋆	0.220
		24M	13.422	13.002	12.600	-0.420	-0.822	-0.402	0.040	0.000 ‡	0.011

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.18: TMLE of mean weight by tertiles of Total DAPs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**				**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3		$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted Total DAPs on Child Weight (kg) @	**42M**	PN	17.764	17.532	17.444		-0.232	-0.319	-0.088	0.531	0.386	0.789
		DL	17.638	17.634	17.488		-0.004	-0.150	-0.146	0.996	0.634	0.805
		6M	17.867	17.544	17.217		-0.323	-0.650	-0.327	0.432	0.039	0.305
		12M	17.879	17.520	17.197		-0.359	-0.682	-0.323	0.207	0.021	0.217
		24M	17.844	17.642	17.420		-0.203	-0.424	-0.221	0.408	0.152	0.366
		42M	17.709	17.402	17.140		-0.307	-0.569	-0.262	0.545	0.234	0.451
	60M	PN	22.302	21.956	21.711		-0.347	-0.591	-0.244	0.402	0.227	0.527
		DL	22.384	22.054	21.692		-0.330	-0.691	-0.362	0.596	0.246	0.436
		6M	22.293	21.856	21.575		-0.437	-0.717	-0.280	0.247	0.038	0.389
		12M	22.349	22.052	21.871		-0.297	-0.478	-0.181	0.612	0.282	0.636
		24M	22.300	22.105	21.788		-0.195	-0.512	-0.316	0.706	0.622	0.742
		42M	22.340	21.755	21.483		-0.585	-0.856	-0.272	0.398	0.250	0.626
		60M	22.569	21.927	21.450		-0.641	-1.119	-0.478	0.334	0.237	0.561

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.19: TMLE mean weight by tertiles of DEs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DEs on Child Weight (kg) @	**6M**	PN	8.258	8.172	8.151	-0.086	-0.106	-0.021	0.508	0.447	0.789
		DL	8.176	8.188	8.192	0.012	0.016	0.004	0.788	0.802	0.952
		6M	8.297	8.181	8.079	-0.115	-0.218	-0.103	0.288	0.014	0.056
	12M	PN	10.147	10.132	10.120	-0.015	-0.027	-0.012	0.951	0.910	0.925
		DL	10.149	10.121	10.109	-0.028	-0.040	-0.012	0.641	0.301	0.760
		6M	10.241	10.103	9.982	-0.137	-0.258	-0.121	0.337	0.010	0.177
		12M	10.137	10.159	10.118	0.022	-0.019	-0.041	0.883	0.784	0.740
	24M	PN	12.892	12.966	13.094	0.074	0.202	0.127	0.784	0.515	0.567
		DL	13.027	12.933	12.961	-0.093	-0.066	0.028	0.480	0.440	0.817
		6M	13.210	12.973	12.741	-0.237	-0.470	-0.233	0.231	0.001	0.022
		12M	13.100	12.991	12.884	-0.109	-0.217	-0.107	0.733	0.378	0.623
		24M	13.457	12.995	12.535	-0.462	-0.922	-0.460	0.001 ⋆	0.000 ‡	0.001 †

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.20: TMLE mean weight by tertiles of DEs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

	Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
		δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DEs on Child Weight (kg) @ 42M	PN	17.464	17.530	17.719	0.066	0.254	0.188	0.890	0.582	0.298
	DL	17.440	17.626	17.705	0.186	0.264	0.079	0.767	0.435	0.866
	6M	18.027	17.545	17.205	-0.482	-0.822	-0.340	0.368	0.359	0.670
	12M	17.673	17.591	17.462	-0.083	-0.211	-0.129	0.896	0.685	0.804
	24M	18.085	17.674	17.130	-0.411	-0.954	-0.544	0.333	0.000 ‡	0.151
	42M	17.748	17.358	17.248	-0.390	-0.500	-0.110	0.497	0.385	0.683
60M	PN	22.058	21.887	22.129	-0.170	0.071	0.241	0.821	0.942	0.653
	DL	21.836	22.070	22.243	0.234	0.407	0.173	0.492	0.540	0.762
	6M	22.571	21.821	21.469	-0.750	-1.102	-0.351	0.381	0.264	0.741
	12M	22.349	22.030	21.728	-0.318	-0.620	-0.302	0.753	0.298	0.776
	24M	22.650	21.798	21.414	-0.852	-1.235	-0.383	0.058	0.024	0.587
	42M	22.338	21.982	21.427	-0.356	-0.911	-0.556	0.604	0.098	0.190
	60M	22.600	22.130	21.343	-0.470	-1.256	-0.786	0.418	0.027	0.197

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.21: TMLE of mean weight by tertiles of DMs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DMs on Child Weight (kg) @	**6M**	PN	8.187	8.183	8.193	-0.004	0.006	0.010	0.932	0.976	0.958
		DL	8.278	8.196	8.110	-0.082	-0.167	-0.085	0.475	0.164	0.309
		6M	8.209	8.184	8.172	-0.025	-0.037	-0.012	0.854	0.594	0.939
	12M	PN	10.165	10.140	10.081	-0.025	-0.084	-0.060	0.730	0.520	0.482
		DL	10.260	10.118	10.021	-0.142	-0.239	-0.097	0.431	0.049	0.476
		6M	10.149	10.118	10.068	-0.030	-0.081	-0.051	0.852	0.456	0.674
		12M	10.184	10.146	10.081	-0.039	-0.104	-0.065	0.672	0.038	0.440
	24M	PN	13.061	12.995	12.919	-0.066	-0.142	-0.076	0.580	0.175	0.242
		DL	13.158	12.989	12.809	-0.170	-0.350	-0.180	0.712	0.004	0.681
		6M	13.153	12.937	12.854	-0.216	-0.299	-0.083	0.317	0.138	0.570
		12M	13.251	12.973	12.750	-0.278	-0.501	-0.223	0.053	0.000 ‡	0.036
		24M	13.233	12.999	12.754	-0.233	-0.479	-0.245	0.221	0.013	0.140

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.22: TMLE of mean weight by tertiles of DMs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DMs on Child Weight (kg) @	**42M**	PN	17.784	17.540	17.403	-0.244	-0.381	-0.137	0.461	0.198	0.715
		DL	17.659	17.693	17.424	0.033	-0.235	-0.268	0.939	0.417	0.518
		6M	17.778	17.519	17.417	-0.258	-0.360	-0.102	0.680	0.571	0.727
		12M	17.944	17.250	17.488	-0.694	-0.457	0.238	0.000 ‡	0.010	0.166
		24M	17.743	17.605	17.496	-0.138	-0.247	-0.109	0.573	0.462	0.738
		42M	17.743	17.443	17.193	-0.300	-0.551	-0.251	0.370	0.309	0.519
	60M	PN	22.316	21.932	21.634	-0.384	-0.682	-0.298	0.308	0.079	0.321
		DL	22.440	21.956	21.709	-0.484	-0.731	-0.247	0.502	0.391	0.626
		6M	22.149	21.815	21.758	-0.333	-0.391	-0.058	0.731	0.688	0.869
		12M	22.189	22.061	21.911	-0.128	-0.278	-0.149	0.654	0.310	0.557
		24M	22.016	22.040	21.955	0.024	-0.060	-0.085	0.947	0.876	0.854
		42M	22.386	21.806	21.477	-0.580	-0.909	-0.329	0.395	0.313	0.620
		60M	22.439	21.945	21.445	-0.494	-0.994	-0.500	0.323	0.176	0.353

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.23: TMLE of mean length by tertiles of Total DAPs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted Total DAPs on Child Length (cm) @	**6M**	PN	67.17	67.14	67.09	-0.03	-0.08	-0.05	0.821	0.831	0.892
		DL	67.47	67.23	66.90	-0.25	-0.58	-0.33	0.359	0.030	0.197
		6M	67.31	67.18	67.04	-0.14	-0.27	-0.13	0.601	0.352	0.453
	12M	PN	74.95	74.79	74.70	-0.17	-0.26	-0.09	0.375	0.237	0.620
		DL	75.13	74.86	74.66	-0.27	-0.46	-0.20	0.579	0.363	0.138
		6M	74.93	74.67	74.42	-0.27	-0.51	-0.24	0.009	0.000 ‡	0.127
		12M	75.03	74.87	74.54	-0.16	-0.49	-0.33	0.574	0.000 †	0.188
	24M	PN	86.54	86.34	86.28	-0.21	-0.26	-0.06	0.687	0.329	0.890
		DL	86.64	86.35	86.18	-0.29	-0.46	-0.17	0.574	0.008	0.727
		6M	86.54	86.38	86.24	-0.16	-0.31	-0.14	0.302	0.147	0.446
		12M	86.83	86.37	86.00	-0.46	-0.83	-0.37	0.209	0.031	0.150
		24M	87.18	86.31	85.65	-0.87	-1.54	-0.67	0.002	0.000 ‡	0.003

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.24: TMLE of mean length by tertiles of Total DAPs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**				**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3		$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted Total DAPs on Child Length (cm) @	**42M**	PN	100.18	99.61	99.31		-0.58	-0.87	-0.29	0.193	0.023	0.350
		DL	99.90	99.76	99.43		-0.14	-0.47	-0.33	0.901	0.303	0.743
		6M	99.98	99.59	99.53		-0.39	-0.44	-0.06	0.538	0.224	0.912
		12M	100.33	99.60	99.09		-0.74	-1.25	-0.51	0.030	0.000 †	0.073
		24M	100.42	99.75	99.20		-0.67	-1.22	-0.55	0.077	0.041	0.279
		42M	99.65	99.44	99.32		-0.21	-0.33	-0.12	0.744	0.544	0.841
	60M	PN	110.82	110.30	109.85		-0.53	-0.98	-0.45	0.190	0.020	0.136
		DL	110.98	110.34	109.88		-0.64	-1.09	-0.45	0.178	0.019	0.158
		6M	110.49	110.26	110.14		-0.23	-0.35	-0.13	0.518	0.288	0.685
		12M	110.86	110.31	109.89		-0.55	-0.97	-0.41	0.284	0.024	0.219
		24M	110.96	110.39	110.04		-0.57	-0.92	-0.35	0.104	0.171	0.583
		42M	110.34	110.27	110.16		-0.06	-0.18	-0.11	0.932	0.805	0.807
		60M	110.77	110.26	109.88		-0.51	-0.89	-0.39	0.376	0.331	0.655

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.25: TMLE of mean length by tertiles of DEs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DEs on Child Length (cm) @	**6M**	PN	67.03	67.39	67.05	0.36	0.03	-0.34	0.182	0.931	0.071
		DL	67.16	67.20	67.24	0.04	0.08	0.05	0.770	0.623	0.781
		6M	67.38	67.17	66.98	-0.20	-0.39	-0.19	0.406	0.043	0.127
	12M	PN	74.90	74.74	74.92	-0.15	0.02	0.17	0.789	0.969	0.520
		DL	74.87	74.84	74.82	-0.03	-0.05	-0.03	0.856	0.561	0.781
		6M	75.10	74.70	74.27	-0.40	-0.84	-0.43	0.148	0.000 ‡	0.015
		12M	74.98	74.91	74.62	-0.06	-0.35	-0.29	0.851	0.027	0.317
	24M	PN	86.31	86.43	86.46	0.12	0.15	0.03	0.829	0.813	0.946
		DL	86.62	85.92	86.68	-0.70	0.05	0.76	0.007	0.713	0.004
		6M	86.56	86.39	86.22	-0.17	-0.34	-0.17	0.549	0.100	0.343
		12M	86.67	86.48	86.11	-0.18	-0.55	-0.37	0.730	0.121	0.387
		24M	86.96	86.41	85.71	-0.55	-1.25	-0.70	0.166	0.000 ‡	0.080

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.26: TMLE of mean length by tertiles of DEs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

	Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
		δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DEs on Child Length (cm) @ 42M	PN	99.80	99.63	99.79	-0.17	-0.01	0.17	0.817	0.993	0.516
	DL	99.69	99.69	99.73	0.00	0.04	0.04	0.996	0.914	0.935
	6M	100.37	99.74	99.07	-0.63	-1.30	-0.67	0.272	0.154	0.408
	12M	99.99	99.80	99.34	-0.19	-0.65	-0.46	0.808	0.225	0.461
	24M	100.21	99.72	99.31	-0.48	-0.90	-0.42	0.279	0.000 ‡	0.330
	42M	99.78	99.45	99.36	-0.32	-0.42	-0.10	0.677	0.595	0.790
Creatinine adjusted DEs on Child Length (cm) @ 60M	PN	110.43	110.23	110.21	-0.19	-0.22	-0.03	0.707	0.724	0.948
	DL	110.50	110.42	110.23	-0.08	-0.27	-0.19	0.842	0.637	0.639
	6M	110.92	110.17	109.65	-0.75	-1.27	-0.52	0.366	0.172	0.510
	12M	110.54	110.27	109.99	-0.26	-0.55	-0.28	0.799	0.250	0.770
	24M	110.85	110.41	110.00	-0.44	-0.86	-0.42	0.367	0.065	0.539
	42M	110.50	110.31	110.06	-0.18	-0.43	-0.25	0.801	0.461	0.507
	60M	110.21	110.53	110.33	0.32	0.12	-0.20	0.596	0.827	0.708

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.27: TMLE of mean length by tertiles of DM exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DMs on Child Length (cm) @	**6M**	PN	67.21	67.16	67.16	-0.05	-0.05	0.00	0.691	0.923	0.993
		DL	67.50	67.22	66.89	-0.27	-0.60	-0.33	0.263	0.044	0.121
		6M	67.25	67.18	67.10	-0.08	-0.15	-0.07	0.792	0.265	0.814
	12M	PN	75.04	74.81	74.58	-0.23	-0.46	-0.23	0.132	0.064	0.157
		DL	75.05	74.85	74.73	-0.20	-0.32	-0.13	0.507	0.215	0.537
		6M	74.75	74.61	74.63	-0.14	-0.12	0.03	0.633	0.583	0.902
		12M	74.94	74.89	74.65	-0.05	-0.29	-0.24	0.814	0.004	0.243
	24M	PN	86.66	86.38	86.11	-0.28	-0.54	-0.26	0.242	0.013	0.022
		DL	86.60	86.34	86.20	-0.26	-0.40	-0.14	0.698	0.022	0.819
		6M	86.55	86.30	86.25	-0.25	-0.30	-0.05	0.506	0.274	0.873
		12M	86.60	86.40	86.16	-0.20	-0.44	-0.24	0.424	0.035	0.225
		24M	86.98	86.38	85.78	-0.59	-1.20	-0.60	0.035	0.000 ‡	0.043

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.28: TMLE of mean length by tertiles of DM exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

	Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
		δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DMs on Child Length (cm) @ 42M	PN	100.23	99.58	99.21	-0.65	-1.02	-0.37	0.133	0.005	0.470
	DL	99.88	99.74	99.60	-0.14	-0.28	-0.14	0.776	0.471	0.772
	6M	99.59	99.67	99.80	0.08	0.21	0.14	0.907	0.733	0.722
	12M	100.14	99.62	99.20	-0.52	-0.94	-0.43	0.023	0.012	0.290
	24M	100.19	99.72	99.38	-0.47	-0.81	-0.33	0.165	0.105	0.543
	42M	99.71	99.52	99.30	-0.19	-0.42	-0.23	0.588	0.458	0.582
Creatinine adjusted DMs on Child Length (cm) @ 60M	PN	110.94	110.22	109.66	-0.71	-1.28	-0.57	0.087	0.006	0.027
	DL	110.97	110.37	109.83	-0.60	-1.14	-0.54	0.296	0.103	0.224
	6M	110.18	110.18	110.45	0.00	0.28	0.27	1.000	0.781	0.360
	12M	110.54	110.23	110.04	-0.31	-0.50	-0.19	0.115	0.055	0.408
	24M	110.80	110.36	109.99	-0.44	-0.81	-0.37	0.148	0.031	0.410
	42M	110.36	110.30	110.29	-0.06	-0.08	-0.01	0.928	0.923	0.978
	60M	110.85	110.22	109.78	-0.63	-1.07	-0.44	0.168	0.107	0.364

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.29: TMLE of mean BMI by tertiles of Total DAPs exposure after adjusting for creatinine at the time indicated by row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	TMLE: Q^1			TMLE Differences			Wald Test p-values		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted Total DAPs on Child BMI (kg/m^2) @	24M	PN	17.31	17.29	17.46	-0.02	0.15	0.17	0.946	0.410	0.444
		DL	17.50	17.33	17.19	-0.18	-0.31	-0.13	0.654	0.010	0.729
		6M	17.61	17.32	17.06	-0.29	-0.55	-0.26	0.004	0.000 ‡	0.067
		12M	17.54	17.34	17.16	-0.21	-0.38	-0.18	0.260	0.065	0.292
		24M	17.58	17.37	17.17	-0.21	-0.41	-0.20	0.386	0.021	0.274
	42M	PN	17.57	17.58	17.63	0.01	0.07	0.05	0.951	0.743	0.786
		DL	17.49	17.63	17.62	0.14	0.13	-0.01	0.805	0.661	0.982
		6M	17.81	17.51	17.36	-0.30	-0.45	-0.15	0.455	0.134	0.574
		12M	17.68	17.57	17.54	-0.11	-0.15	-0.04	0.616	0.550	0.856
		24M	17.53	17.63	17.67	0.10	0.14	0.04	0.668	0.564	0.873
		42M	17.73	17.49	17.32	-0.24	-0.41	-0.17	0.495	0.108	0.482
	60M	PN	17.91	17.91	17.91	0.00	0.01	0.01	0.996	0.982	0.982
		DL	18.02	17.88	17.92	-0.14	-0.11	0.04	0.693	0.764	0.878
		6M	18.13	17.83	17.63	-0.30	-0.50	-0.20	0.212	0.025	0.328
		12M	17.94	17.96	18.04	0.02	0.10	0.08	0.962	0.742	0.719
		24M	17.92	17.93	17.92	0.01	-0.01	-0.01	0.981	0.993	0.979
		42M	18.16	17.80	17.55	-0.36	-0.61	-0.25	0.295	0.130	0.438
		60M	18.17	17.88	17.66	-0.29	-0.52	-0.23	0.451	0.396	0.675

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.30: TMLE of mean BMI by tertiles of DEs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DEs on Child BMI (kg/m^2) @	**24M**	PN	17.26	17.28	17.42	0.02	0.16	0.14	0.950	0.648	0.465
		DL	17.40	17.25	17.31	-0.14	-0.09	0.06	0.341	0.426	0.689
		6M	17.53	17.34	17.13	-0.19	-0.40	-0.20	0.410	0.020	0.071
		12M	17.26	17.34	17.40	0.08	0.13	0.05	0.829	0.574	0.859
		24M	17.73	17.37	17.03	-0.36	-0.70	-0.34	0.039	0.000 ‡	0.033
	42M	PN	17.39	17.60	17.72	0.21	0.33	0.12	0.544	0.305	0.379
		DL	17.38	17.64	17.73	0.26	0.35	0.09	0.578	0.146	0.804
		6M	17.74	17.54	17.45	-0.19	-0.29	-0.10	0.644	0.672	0.872
		12M	17.59	17.57	17.64	-0.02	0.05	0.08	0.959	0.898	0.848
		24M	17.88	17.62	17.33	-0.26	-0.55	-0.29	0.248	0.000 ‡	0.156
		42M	17.65	17.46	17.41	-0.19	-0.24	-0.05	0.533	0.441	0.799
	60M	PN	17.90	17.88	18.04	-0.02	0.14	0.16	0.966	0.782	0.604
		DL	17.74	17.95	18.07	0.20	0.33	0.12	0.290	0.409	0.727
		6M	18.02	17.84	17.73	-0.18	-0.29	-0.11	0.758	0.578	0.869
		12M	18.11	17.94	17.81	-0.17	-0.30	-0.13	0.770	0.359	0.839
		24M	18.28	17.83	17.52	-0.45	-0.76	-0.31	0.096	0.017	0.454
		42M	18.12	17.87	17.53	-0.24	-0.58	-0.34	0.571	0.098	0.179
		60M	18.43	17.95	17.43	-0.49	-1.00	-0.51	0.173	0.003	0.150

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.31: TMLE of mean BMI by tertiles of DMs exposure after adjusting for creatinine at the time indicated on the row. The time of the outcome is indicated on the far left column; allowing only for exposures prior or concurrent to the time of the outcome.

		Exposure Time	**TMLE: Q^1**			**TMLE Differences**			**Wald Test p-values**		
			δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
Creatinine adjusted DMs on Child BMI (kg/m^2) @	**24M**	PN	17.33	17.35	17.39	0.02	0.06	0.04	0.911	0.698	0.630
		DL	17.52	17.33	17.16	-0.20	-0.37	-0.17	0.537	0.002 ⋆	0.561
		6M	17.50	17.27	17.18	-0.22	-0.32	-0.10	0.364	0.155	0.602
		12M	17.58	17.33	17.16	-0.24	-0.42	-0.17	0.118	0.002 ⋆	0.141
		24M	17.46	17.37	17.27	-0.09	-0.19	-0.10	0.664	0.368	0.572
	42M	PN	17.60	17.60	17.62	0.00	0.02	0.03	0.988	0.921	0.926
		DL	17.58	17.65	17.51	0.07	-0.07	-0.14	0.853	0.781	0.680
		6M	17.77	17.46	17.46	-0.31	-0.31	0.00	0.532	0.485	0.995
		12M	17.86	17.20	17.74	-0.66	-0.12	0.54	0.000 ‡	0.422	0.000 ‡
		24M	17.50	17.61	17.68	0.11	0.18	0.07	0.589	0.506	0.811
		42M	17.73	17.51	17.33	-0.23	-0.40	-0.18	0.393	0.221	0.414
	60M	PN	17.94	17.91	17.89	-0.03	-0.06	-0.03	0.902	0.816	0.883
		DL	18.01	17.88	17.90	-0.12	-0.11	0.02	0.783	0.841	0.949
		6M	18.09	17.84	17.63	-0.25	-0.45	-0.20	0.705	0.473	0.348
		12M	17.89	17.99	18.07	0.10	0.18	0.08	0.515	0.371	0.605
		24M	17.77	17.95	18.00	0.18	0.23	0.05	0.453	0.292	0.860
		42M	18.18	17.79	17.55	-0.38	-0.62	-0.24	0.299	0.197	0.515
		60M	18.10	17.90	17.73	-0.21	-0.37	-0.17	0.541	0.459	0.641

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
⋆: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Table B.32: TMLE of mean waist circumference at 5 years by tertiles of exposure from Total DAPs, DEs, and DMs after adjusting for creatinine at the time indicated by row.

	Exposure		TMLE: Q^1			TMLE Differences			Wald Test p-values		
		Time	δ_1	δ_2	δ_3	$\psi_{2,1}$	$\psi_{3,1}$	$\psi_{3,2}$	$p_{2,1}$	$p_{3,1}$	$p_{3,2}$
5Y Waist Circumference (cm)	Creatinine adjusted	Total DAPs	PN	58.49	58.52	58.36	0.03	-0.13	-0.17	0.956	0.837	0.764
			DL	58.92	58.48	58.06	-0.44	-0.86	-0.42	0.617	0.326	0.552
			6M	59.04	58.26	57.68	-0.77	-1.35	-0.58	0.189	0.009	0.240
			12M	58.87	58.57	58.24	-0.29	-0.63	-0.34	0.773	0.385	0.596
			24M	58.25	58.66	58.80	0.40	0.55	0.14	0.578	0.666	0.892
			42M	59.03	58.32	57.62	-0.72	-1.41	-0.69	0.431	0.151	0.361
			60M	59.22	58.37	57.52	-0.85	-1.71	-0.85	0.404	0.305	0.559
		DEs	PN	58.61	58.44	58.67	-0.17	0.05	0.23	0.868	0.968	0.804
			DL	58.28	58.48	58.50	0.20	0.22	0.02	0.715	0.819	0.977
			6M	59.54	58.15	57.40	-1.39	-2.14	-0.74	0.336	0.132	0.636
			12M	58.91	58.39	58.42	-0.51	-0.49	0.03	0.763	0.647	0.988
			24M	59.51	58.40	57.41	-1.10	-2.10	-0.99	0.198	0.017	0.411
			42M	59.17	58.44	57.67	-0.73	-1.50	-0.77	0.533	0.138	0.176
			60M	59.75	58.44	57.45	-1.31	-2.30	-0.99	0.363	0.010	0.478
		DMs	PN	58.56	58.54	58.23	-0.02	-0.32	-0.30	0.966	0.648	0.491
			DL	58.90	58.47	57.95	-0.42	-0.95	-0.53	0.666	0.369	0.386
			6M	58.99	58.18	58.06	-0.81	-0.93	-0.12	0.640	0.587	0.836
			12M	58.75	58.60	58.41	-0.15	-0.34	-0.18	0.707	0.352	0.599
			24M	58.13	58.61	58.85	0.48	0.72	0.24	0.479	0.224	0.629
			42M	58.91	58.23	57.91	-0.68	-0.99	-0.31	0.499	0.439	0.725
			60M	59.25	58.25	57.55	-1.00	-1.70	-0.70	0.181	0.150	0.420

PN: prenatal; DL: delivery; 6M: 6 months; 12M: 12 months; 24M: 24 months; 42M: 3.5 years; 60M: 5 years.
$\star$: adjusted p-value < 0.1, †: adjusted p-value < 0.05, ‡: adjusted p-value < 0.01

Appendix C

Longitudinal effect estimation algorithm

1. Estimate the intermediate and final outcome models using SuperLearner with the desired library of candidate learners. In our case: glm, gam, polymars, and D/S/A.

2. Sample with replacement 10,000 observations from the original data set to begin the estimation of our parameter of interest.

   (a) Simulate the first intermediate model fixing the level of exposure at the one of interest and using the actual values of the covariates. This returns a matrix of probabilities for the weight levels at the first intermediate time point.

   (b) Using the probabilities obtained in the previous step, sample the weight level for each subject at the first intermediate time point.

   (c) Simulate the second intermediate model fixing the level of exposure at the one of interest which is also consistent with the first simulation, the actual covariate values, and the simulated weight levels calculated in the previous step. This returns a second matrix of probabilities for the weight levels at the second intermediate time point.

   (d) Using the probabilities obtained in the last step, sample the weight level for each of the individuals at the second intermediate time point.

   (e) Simulate the final model fixing the level of exposure at the one of interest which is also consistent with the first two simulations, the actual covariate values, and the simulated weight categories for the intermediate time points. This returns a vector of weights for your population at the desired level of exposure.

3. Target the level of exposure used in the simulations and update your estimates. This returns the targeted estimate under the exposure of interest.

4. To obtain the inference on the estimate, and build confidence intervals around it, bootstrap $n \geq 1,000$ times. Sample with replacement from the original population a new population of the same size. Then perform steps 2 and 3 on the new population.