American Journal of Epidemiology

Vol. 191, No. 8
https://doi.org/10.1093/aje/kwac062
Advance Access publication:
March 30, 2022

Original Contribution

Use of a Doubly Robust Machine-Learning–Based Approach to Evaluate Body Mass Index as a Modifier of the Association Between Fruit and Vegetable Intake and Preeclampsia

Lisa M. Bodnar*, Abigail R. Cartus, Edward H. Kennedy, Sharon I. Kirkpatrick, Sara M. Parisi, Katherine P. Himes, Corette B. Parker, William A. Grobman, Hyagriv N. Simhan, Robert M. Silver, Deborah A. Wing, Samuel Perry, and Ashley I. Naimi

* Correspondence to Dr. Lisa M. Bodnar, 5129 Public Health, Department of Epidemiology, School of Public Health, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261 (e-mail: lbodnar@pitt.edu).

Initially submitted March 26, 2021; accepted for publication March 25, 2022.

The Dietary Guidelines for Americans rely on summaries of the effect of dietary pattern on disease risk, independent of other population characteristics. We explored the modifying effect of prepregnancy body mass index (BMI; weight (kg)/height (m)^2) on the relationship between fruit and vegetable density (cup-equivalents/1,000 kcal) and preeclampsia using data from a pregnancy cohort study conducted at 8 US medical centers ($n = 9{,}412$; 2010–2013). Usual daily periconceptional intake of total fruits and total vegetables was estimated from a food frequency questionnaire. We quantified the effects of diets with a high density of fruits ($\geq$1.2 cups/1,000 kcal/day vs. $<$1.2 cups/1,000 kcal/day) and vegetables ($\geq$1.3 cups/1,000 kcal/day vs. $<$1.3 cups/1,000 kcal/day) on preeclampsia risk, conditional on BMI, using a doubly robust estimator implemented in 2 stages. We found that the protective association of higher fruit density declined approximately linearly from a BMI of 20 to a BMI of 32, by 0.25 cases per 100 women per each BMI unit, and then flattened. The protective association of higher vegetable density strengthened in a linear fashion, by 0.3 cases per 100 women for every unit increase in BMI, up to a BMI of 30, where it plateaued. Dietary patterns with a high periconceptional density of fruits and vegetables appear more protective against preeclampsia for women with higher BMI than for leaner women.

birth; body mass index; dietary patterns; machine learning; obesity; preeclampsia; pregnancy; pregnant women

Abbreviations: BMI, body mass index; EIF, efficient influence function; FFQ, food frequency questionnaire; LASSO, least absolute shrinkage and selection operator; nuMoM2b, Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be; TMLE, targeted maximum likelihood estimation.

In 2016, poor diet quality was associated with more deaths in the United States than smoking, obesity, or other modifiable risk factors (1). Interventions and education programs implemented at the national, regional, state, and local levels to improve dietary patterns are often based on the Dietary Guidelines for Americans (2), which provide dietary advice for health promotion and disease prevention. The Dietary Guidelines for Americans rely on a summary of the effect of dietary pattern on disease risk, independent of other features of the diet and other characteristics of the US population (3). These summary associations are of central importance in a population health setting, as they establish whether changes in a dietary exposure could have meaningful impacts overall in the target population of interest (4). Nevertheless, such summary associations are not always relevant for specific subgroups in the population (5). For instance, smokers may have greater requirements for antioxidant-rich foods than nonsmokers (6, 7). Vitamin D needs are higher among persons affected by obesity than among those not affected (8, 9). These differential needs should translate into subgroup-specific recommendations for the most healthful dietary patterns.

Evaluation of heterogeneity (statistical interaction or effect-measure modification) helps identify subgroups who may benefit most from dietary interventions (10), thereby enabling targeting of limited resources. Here, we use the

term *interaction* in its statistical sense. Conventional (parametric) analytical approaches for exploring heterogeneity require researchers to correctly specify all interactions among dietary variables and between dietary variables and covariates a priori when evaluating the risk of the adverse outcome (11–13). Incorrect specification of these models can result in bias. Machine learning algorithms can potentially overcome this limitation by using data-adaptive strategies that account for key interactions between variables (12, 14, 15). Further, they flexibly model interactions, which enables better handling of a number of other specifications and assumptions in comparison with traditional parametric regression (16).

We previously estimated associations between periconceptional fruit and vegetable intake and the risk of preeclampsia using doubly robust targeted maximum likelihood estimation (TMLE) paired with an ensemble machine learning algorithm (17). We compared these associations with results generated from multivariable logistic regression and found that the TMLE yielded larger reductions in preeclampsia risk with higher intake of fruits and vegetables (17). In the present study, we further explored the extent to which the relationship between periconceptional fruit and vegetable intake and preeclampsia is modified by prepregnancy body mass index (BMI). Nutrient deficiencies and high levels of oxidative stress are more common among persons affected by overweight or obesity than among leaner people (18–20). As a result, dietary patterns rich in fruits and vegetables and their accompanying nutrients may be more important for reducing the risk of preeclampsia among women with obesity than among leaner women, as has been shown with other nutrition exposures (21–24).

We evaluated this effect modification without categorizing BMI, and we used flexible machine learning methods to account for deviations from linearity, as well as any potential exposure-confounder or confounder-confounder interactions of importance. We hypothesized that the reductions in preeclampsia risk associated with high (versus low) periconceptional fruit and vegetable densities would strengthen as BMI increased.

METHODS

We used data from the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b), a large pregnancy cohort study, the details of which have been published elsewhere (25). The nuMoM2b investigators enrolled 10,038 women from 8 US medical centers (2010–2013). Women were eligible if they had a viable singleton pregnancy, were at 6–13 completed weeks of gestation, and had no previous pregnancy of ≥20 weeks' gestation. At all sites, the nuMoM2b researchers used trained and credentialed study personnel and a common protocol and manual of operations. At enrollment (6–13 completed weeks' gestation), usual periconceptional dietary intake, demographic characteristics, medical history, psychosocial assessments, and behaviors were ascertained. Data on birth outcomes, medical history, and delivery diagnoses and complications were collected from the medical record at least 30 days after delivery. Each site's local institutional review board approved the study protocol, and all women gave written, informed consent. Our analytical sample included 9,412 women who delivered at ≥20 weeks' gestation and had information on preeclampsia diagnosis (see Web Figure 1, available at https://doi.org/10.1093/aje/kwac062). We retained records with missing data on other relevant variables through an imputation process described below.

At enrollment, a self-administered modified Block 2005 food frequency questionnaire (FFQ) (available in English and Spanish) was used to assess usual dietary intake during the 3 months before conception. The FFQ's list of approximately 120 food and beverage items was based on analyses of 24-hour dietary recall data from the 1999–2002 administrations of the National Health and Nutrition Examination Survey. A series of "adjustment" questions in the FFQ are used to improve the estimation of fat and carbohydrate intake. Respondents were asked about portion size for each food, and pictures of portion sizes were given to participants to enhance accuracy. The FFQ has acceptable validity relative to other self-reported assessment tools in many samples of pregnant women (26–31). We modified the questionnaire to reflect a 3-month period. Study personnel checked all pages of the FFQ for completeness. Block Dietary Data Systems (Berkeley, California) performed scanning, nutrient and food group mapping, and summary analysis of the FFQ using software developed at the National Cancer Institute (32). The food items were disaggregated into their component parts and linked to a nutrient database that was developed from the US Department of Agriculture's Food and Nutrient Database for Dietary Studies (33) and the food-based MyPyramid Equivalents Database, version 2.0 (34). Finally, food group and nutrient summary variables were created.

Total fruit and total vegetable consumption, our 2 exposures of interest, were defined as densities (number of cups per 1,000 kcal) as per the construction of Healthy Eating Index 2010 (35). Potatoes and tomatoes were included with total vegetables, and avocadoes and fruit juices were included with total fruits, per the method of the MyPyramid Equivalents Database at the time of the FFQ processing. Legumes were first included in total protein intake, and after the protein standard was met, excess legumes were included as vegetables. We dichotomized density of fruit intake (1.2 cups/1,000 kcal of fruit per day) and density of vegetable intake (1.3 cups/1,000 kcal of vegetables per day) on the basis of the 80th percentile of the distributions, which reflects the upper quintile of the distribution. Analytical models for fruit density included vegetable density as a covariate and vice versa. To account for the multidimensional dietary patterns, we also included food group densities for all other Healthy Eating Index 2010 components, including whole grains, dairy products, total protein foods, seafood and plant proteins, fatty acids, refined grains, sodium, and "empty" calories in the models. We calculated the percentage of empty calories by summing the amounts of energy provided by added sugars, solid fats, and excess alcohol intake (>13 g/1,000 kcal) and dividing by total energy intake (35). All other dietary components were entered as continuous confounders (35).

Detailed definitions of pregnancy hypertensive disorders used in the nuMoM2b cohort are provided elsewhere (36). Briefly, women were classified as having preeclampsia if they had the following symptoms during the period from ≥20 weeks' gestation through 14 days postpartum: gestational hypertension (systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg on 2 occasions ≥6 hours apart or on 1 occasion with subsequent antihypertensive therapy, excluding blood pressures recorded during the second stage of labor), proteinuria (≥300 mg/24-hour collection or protein:creatinine ratio ≥0.3 or dipstick ≥2+), thrombocytopenia (platelet count <100,000/mm^3), or pulmonary edema. Preeclampsia included superimposed preeclampsia or eclampsia, regardless of the timing of onset. The nuMoM2b principal investigators reviewed clinical data and adjudicated final classification based on consensus judgment for cases that presented atypically and were difficult to classify according to study criteria.

At enrollment, women reported their prepregnancy weight and had their height measured using a stadiometer or measuring tape. We calculated prepregnancy body mass index (BMI) as self-reported weight (kg) divided by measured height (m) squared. We evaluated BMI as both a continuous variable and a categorical variable (underweight (<18.5), normal weight (18.5–24.9), overweight (25.0–29.9), or obese (≥30.0) (37)). There were too few underweight women in the stratified analyses, so we grouped them with normal-weight women.

At the first visit, women also self-reported their marital status, prepregnancy smoking, health insurance status, highest level of education, race/ethnicity, whether the pregnancy was planned, physical activity, and preexisting diabetes and chronic hypertension (25). They also reported whether they had been born in the United States and the language spoken at home. Women completed evaluations of depressive symptoms (Edinburgh Postnatal Depression Scale (38)) and anxiety (Spielberger State-Trait Anxiety Inventory (39)), stress (Perceived Stress Scale (40)), and insomnia symptoms (Women's Health Initiative Insomnia Rating Scale (41)).

Trained medical record abstractors collected data from ultrasound reports conducted by certified sonographers. Gestational age was determined by applying the algorithm defined by the nuMoM2b investigators (25).

We sought to quantify the effect of high fruit density and high vegetable density (denoted X) on preeclampsia risk (denoted Y), conditional on the effect-measure modifier of interest (BMI, denoted Z). If we let $Y^x = 1$ denote the potential outcome that would be observed if a woman consumed at least 1.2 cups/1,000 kcal of fruit per day (≥80th percentile) or 1.3 cups/1,000 kcal of vegetables per day (≥80th percentile) ($Y^x = 0$ otherwise), this effect can be defined as

$$\psi(Z) = E\left(Y^{x=1} - Y^{x=0} \mid Z\right),$$

which yields a conditional average treatment effect, or an average treatment effect for each unique value of BMI. We estimated $\psi(Z)$ using a doubly robust estimator, known as DR-Learner (42–44), implemented in 2 stages.

The first stage consisted of fitting confounder-adjusted treatment and outcome models to construct predictions of expected fruit or vegetable density and individual risk of preeclampsia under high and low intake. Confounders of each relationship (collectively denoted C) were the remaining Healthy Eating Index 2010 components conceptualized as part of a multidimensional dietary pattern, maternal race/ethnicity, age, smoking, education, marital status, health insurance, acculturation, planning of pregnancy, preexisting diabetes or chronic hypertension, sleep quality, physical activity level, and scores for depression, anxiety, and stress. Age, physical activity, the psychosocial scores, and the dietary measures were entered as continuous variables, and the rest were entered as discrete variables. We did not include energy intake in the models because all dietary variables were specified as densities, thereby adjusting for energy intake. This set of confounders was identified via theory-based causal graphs (45). We also added to this outcome model BMI, the effect-measure modifier of interest. The exposure model regressed fruit or vegetable intake against the previously listed confounders and BMI.

Predictions for the exposure and outcome model were obtained using the cross-validated targeted maximum likelihood estimator (the "tmle3shift" package (46)), incorporating the cross-validated ensemble Super Learner (or stacking) (47, 48), which included a prespecified library of algorithms, each fitted over a grid of hyperparameters. This library included 1) random forests (ranger) with minimum node sizes of 10, 500, and 2,500 trees, 2, 3, and 4 predictor variables selected at random for each split, and sampling with and without replacement; 2) extreme gradient boosting (xgboost) with 200, 500, and 1,000 trees, maximum tree depth of 4, 5, or 6, and shrinkage parameters of 0.01, 0.001, or 0.0001; 3) least absolute shrinkage and selection operator (LASSO) and elastic-net regularized generalized linear models (glmnet) with elastic net mixing parameter $\alpha = 0.0$ (ridge penalty), 0.2, 0.4, 0.6, 0.8, or 1.0 (LASSO penalty); 4) generalized linear models (glm); and 5) simple mean. Each ensemble learner was fitted using 10-fold cross validation. Super Learner was optimized via the binomial loss function. An additional layer of 10-fold cross-validation (also referred to as sample-splitting or cross-fitting) for the TMLE algorithm was used to avoid the need for empirical process conditions (49).

The second stage of the DR-Learner approach involved obtaining efficient influence function (EIF) values for each woman in the sample from the original TMLE fit and regressing these on the effect-measure modifier of interest. These EIF values were defined as

$$\mathrm{EIF}_i = \frac{(2X_i - 1)\left[Y - \hat{E}(Y \mid X_i, C_i, Z_i)\right]}{\hat{P}(X_i C_i, Z_i)} + \hat{E}(YX_i = 1, C_i, Z_i) - \hat{E}(Y \mid X_i = 0, C_i, Z_i),$$

where $\hat{E}(Y \mid X_i, C_i, Z_i)$ denotes outcome model predictions and $\hat{P}(X_i C_i, Z_i)$ denotes observed exposure predictions obtained from the Super Learner described above. Conceptually, these EIF_i values can be thought of as nearly unbiased

Table 1. Characteristics of 9,412 Deliveries According to Prepregnancy Body Mass Index Category in the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be, 2010–2013

Characteristic	Prepregnancy Body Mass Index					
	Underweight or Normal Weight (*n* = 5,433)		Overweight (*n* = 2,038)		Obese (*n* = 1,941)	
	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)
Maternal age, years						
<25	33		36		41	
25–34	58		54		49	
≥35	9		10		10	
Maternal race/ethnicity						
Non-Hispanic White	65		58		49	
Non-Hispanic Black	10		15		24	
Hispanic	16		19		19	
Other	10		7		8	
Maternal education						
High school or less	17		21		26	
Some college	25		32		38	
College graduate	31		26		22	
Graduate degree	27		21		14	
Smoking status						
Smoker	15		19		23	
Nonsmoker	85		81		77	
Marital status						
Not married	33		40		48	
Married	67		60		52	
Health insurance at delivery						
Public	24		29		36	
Private	76		71		64	
Acculturation						
Mother and her parents born in US	70		70		74	
Mother born in US with ≥1 immigrant parent	13		15		16	
Mother born outside US						
Mother immigrated at age <6 years	5		6		4	
Mother immigrated at age ≥6 years	12		10		6	
Pregnancy was planned						
Yes	64		57		48	
No	36		43		52	
Preexisting diabetes						
Yes	1		2		4	
No	99		98		96	
Preexisting chronic hypertension						
Yes	1		3		9	
No	99		97		91	

Table continues

Table 1. Continued

Characteristic	Prepregnancy Body Mass Index					
	Underweight or Normal Weight (*n* = 5,433)		Overweight (*n* = 2,038)		Obese (*n* = 1,941)	
	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)
Edinburgh Postnatal Depression Scale[a]		5 (6)		5 (5)		5 (6)
Spielberger State-Trait Anxiety Inventory—Trait Subscale[b]		32 (11)		33 (11)		34 (13)
Perceived Stress Scale[c]		12 (9)		12 (9)		13 (9)
Sleep quality[d]						
Restless	17		21		23	
Average	40		42		44	
Restful/very restful	43		37		33	
Total physical activity, MET-hours/week		420 (1,050)		315 (841)		210 (675)
Dietary components						
Total fruit, cups/1,000 kcal		0.8 (0.7)		0.7 (0.6)		0.8 (0.6)
Total vegetables, cups/1,000 kcal		0.9 (0.6)		0.8 (0.6)		0.8 (0.5)
Dairy foods, cups/1,000 kcal		0.8 (0.5)		0.8 (0.5)		0.8 (0.5)
Total protein foods, ounces[e]/1,000 kcal		2.4 (0.9)		2.4 (0.8)		2.4 (1.0)
Seafood and plant proteins, ounces/1,000 kcal		0.7 (0.7)		0.6 (0.6)		0.6 (0.5)
Fatty acid ratio[f]		1.9 (0.5)		1.9 (0.5)		1.8 (0.5)
Refined grains, ounces/1,000 kcal		2.2 (1.0)		2.2 (0.9)		2.2 (0.9)
Whole grains, ounces/1,000 kcal		0.6 (0.6)		0.5 (0.5)		0.5 (0.5)
Sodium, g/1,000 kcal		1.6 (0.3)		1.6 (0.3)		1.6 (0.3)
Empty calories, % of energy		30 (10)		31 (11)		33 (11)

Abbreviations: IQR, interquartile range; MET, metabolic equivalent of task; US, United States.
[a] Edinburgh Postnatal Depression Scale scores range from 0 to 30 (38).
[b] Spielberger State-Trait Anxiety Inventory—Trait Subscale scores range from 20 to 80 (39).
[c] Perceived Stress Scale scores range from 0 to 40 (40).
[d] Item from the Women's Health Initiative Insomnia Rating Scale (41).
[e] 1 ounce = 28.4 g.
[f] Ratio of poly- and monounsaturated fatty acid intake to saturated fatty acid intake.

pseudo-outcomes, centered at individual risk differences for each woman in the sample, indexed by i. These EIF_i values are then regressed against the effect-measure modifier Z to estimate a function that captures the relationship between the risk difference for the effect of fruit and vegetable consumption on preeclampsia and the effect-measure modifier of interest. This latter regression is implemented with the same (cross-validated) stacked generalization described above. Predictions from the fit of this model are then obtained across the range of BMI to obtain the risk difference for each value of the (continuous) effect-measure modifier. We bootstrapped this second-stage fit to obtain the uncertainty bands depicted in the figures. Importantly, these uncertainty bands are not guaranteed to yield 95% confidence intervals for the underlying parameter of interest, but they do provide valid confidence intervals for a smoothed sample version of the parameter (50). We also estimated risk differences using a standard cross-validated TMLE approach (51) within categorized versions of BMI, as a check against DR-Learner results, which treated BMI as a continuous variable.

A total of 24% of women (n = 2,258) had missing FFQ or covariate data. We addressed missing data in several steps. For the 0.8% of women who were missing data on BMI (n = 79), mean imputation was performed separately depending on maternal race/ethnicity. For remaining covariates with missing data, missing values were median- or mode-imputed for continuous and categorical covariates, respectively. We also included missing indicators in the TMLE models (52).

We explored the impact of the competing risk of spontaneous pregnancy loss (fetal death at <20 weeks' gestation), preterm birth (live birth at <37 weeks' gestation), and

Table 2. Association[a] Between Periconceptional Fruit Density and Risk of Preeclampsia, Overall and by Prepregnancy Body Mass Index (Effect-Measure Modifier), in the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be, 2010–2013

BMI Category and Total Fruit Density, cups/1,000 kcal	No. of Women at Risk	Preeclampsia Cases		Adjusted[b] No. of Excess Cases per 100 Deliveries	95% CI
		No.	%		
Overall					
<1.2[c]	7,837	718	9.2	0	Referent
≥1.2	1,575	113	7.2	−2.6	−3.9, −1.2
Prepregnancy BMI category					
Underweight or normal weight (<25.0)					
<1.2	4,418	296	6.7	0	Referent
≥1.2	1,015	63	6.2	−0.5	−2.4, 1.3
Overweight (25.0–29.9)					
<1.2	1,736	187	10.8	0	Referent
≥1.2	302	15	5.0	−6.9	−9.3, −4.6
Obese (≥30.0)					
<1.2	1,683	235	14.0	0	Referent
≥1.2	258	35	13.6	−1.8	−5.0, 1.4

Abbreviations: BMI, body mass index; CI, confidence interval.

[a] Calculated using targeted maximum likelihood estimation with Super Learner.

[b] Results were adjusted for Healthy Eating Index 2010 components, maternal race/ethnicity, smoking, education, marital status, age, health insurance, acculturation, pregnancy planning, preexisting diabetes or chronic hypertension, sleep quality, physical activity level, and depression, anxiety, and stress assessment scores, as well as prepregnancy BMI as appropriate.

[c] 1.2 cups of fruit/1,000 kcal was the 80th percentile of the fruit density distribution.

stillbirth (fetal death at ≥20 weeks' gestation) on the associations of interest by approximating the subdistribution risk differences of the effects of interest (see Web Appendix).

To determine whether our results were sensitive to extreme energy intakes, we reran our analysis after replacing extreme energy intakes at the <5th or >95th percentiles of the distribution with the 5th and 95th percentile values (756 kcal and 3,327 kcal), as well as replacing self-reported prepregnancy weight with measured first-trimester weight for calculation of BMI.

All analyses were conducted in R 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Most women in the sample were normal-weight (58%), with 22% affected by overweight and 20% by obesity. Sixty percent of women were non-Hispanic White, 14% non-Hispanic Black, 17% Hispanic, and 9% another race/ethnicity. Half of women were college-educated (28%) or had a graduate degree (23%). Preterm birth occurred in 8.3% of pregnancies. For every 1,000 kcal of energy intake, women in the sample consumed an average of 0.83 (standard deviation, 0.53) cups of fruit per day and 0.96 (standard deviation, 0.54) cups of vegetables per day. Approximately 7% of women reported usual daily densities of ≥1.2 cups/1,000 kcal of fruits and ≥1.3 cups/1,000 kcal of vegetables, while 13% reported ≥1.2 cups/1,000 kcal intakes of fruit only and 13% reported ≥1.3 cups/1,000 kcal intakes of vegetables only.

Women with obesity were more likely than leaner women to be under age 25 years, non-Hispanic Black, non–college-educated, smokers, and unmarried and to have public health insurance, preexisting diabetes, or hypertension (Table 1). They also more often reported less restful sleep and less physical activity per week and that their pregnancy was unplanned. Daily densities of fruit, vegetables, seafood and plant proteins, and whole grains were lower, and daily density of empty calories was higher, among women with obesity compared with their counterparts.

Preeclampsia occurred in 8.8% of pregnancies (n = 831). Of these cases, 26% delivered preterm. The incidence rose with increasing prepregnancy BMI category (underweight and normal weight, 6.6%; overweight, 9.9%; obese, 14.0%). In the overall cohort, the unadjusted incidence of preeclampsia was lower among women who reported usual daily fruit density greater than or equal to 1.2 cups/1,000 kcal (7.2%) compared with fruit density less than 1.2 cups/1,000 kcal (9.2%) (Table 2). After adjustment for confounders via TMLE, women who consumed at least 1.2 cups of fruit per 1,000 kcal had 2.6 fewer preeclampsia cases than women who consumed less than 1.2 cups of fruit per 1,000 kcal, for every 100 women in the sample (adjusted risk difference = −0.026, 95% confidence interval: −0.039, −0.012).

The association between fruit intake and preeclampsia was modified by prepregnancy BMI (Figure 1). Using the DR-Learner, we found no meaningful change in the adjusted

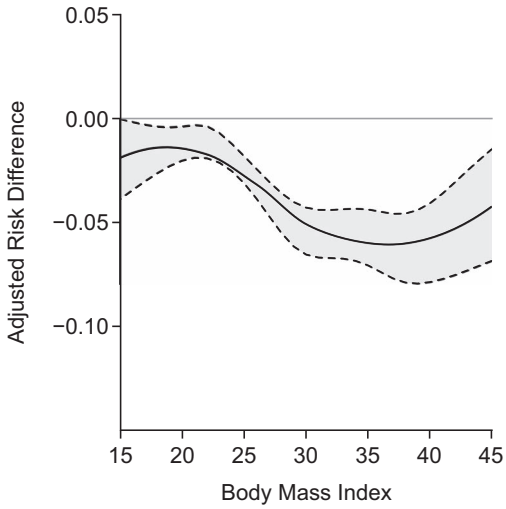

Figure 1. Adjusted difference in the risk of preeclampsia associated with a fruit density of ≥1.2 cups/1,000 kcal versus <1.2 cups/1,000 kcal, according to prepregnancy body mass index (weight (kg)height $(m)^2$), Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be, 2010–2013. Risk differences were adjusted for Healthy Eating Index 2010 components, maternal race/ethnicity, smoking, education, marital status, age, health insurance, acculturation, pregnancy planning, preexisting diabetes or chronic hypertension, sleep quality, physical activity level, and depression, anxiety, and stress assessment scores. The solid line represents the point estimate, and the dashed lines and shaded area represent uncertainty bands.

risk difference up to a BMI of approximately 20. However, as BMI increased past 20, the risk difference became larger at a rate of approximately 0.25 cases per 100 women in the sample for every unit increase in BMI, up to approximately 32. As BMI increased beyond 32, the risk difference did not change meaningfully. Although the point estimate increased at very high BMI values, the uncertainty intervals were wide owing to a small number of women with class 3 obesity. These findings obtained with BMI on a continuous scale were consistent with the results obtained from the BMI-stratified analysis shown in Table 2.

Higher vegetable density was also associated with fewer preeclampsia cases in the cohort than lower vegetable density, and the relationship was also modified by BMI (Table 3 and Figure 2). After adjustment for confounders via TMLE, women who consumed at least 1.3 cups of vegetables per 1,000 kcal had 3.0 fewer preeclampsia cases than women who consumed less than 1.3 cups of vegetables per 1,000 kcal, for every 100 women in the sample (adjusted risk difference = −0.030, 95% confidence interval: −0.046, −0.015). The protective association of higher vegetable density strengthened in a linear fashion, by approximately 0.3 cases per 100 women for every unit increase in BMI, up to a BMI of 30, where it plateaued. As with the fruit density model, there was some evidence of an increasing risk difference at high BMI values, but results were imprecise.

Our results were not sensitive to extreme energy intakes (Web Figure 2) or the use of self-reported BMI (Web Figure 3). Additionally, no meaningful changes were observed when accounting for the impact of competing risks (Web Figures 4 and 5).

DISCUSSION

In this US cohort of nulliparous women, we found that the protective association between diets with a high density of fruits or vegetables (compared with lower densities) and risk of preeclampsia strengthened as prepregnancy BMI increased through a BMI of approximately 30, and plateaued thereafter. These associations were consistent after controlling for all components of Healthy Eating Index 2010, race/ethnicity, education, and other confounders. Our use of TMLE for nonparametric effect modification allowed us to observe strong and meaningful differences in the risk of preeclampsia with each unit increase in BMI, while avoiding the pitfalls of arbitrarily categorizing continuous BMI (53, 54).

Our finding that dietary patterns rich in fruits and vegetables are related to a reduced risk of preeclampsia is consistent with existing evidence (55–58), but previous studies have not evaluated effect modification by BMI. Some data suggest that BMI modifies the association between dietary supplements and risk of preeclampsia. Conde-Agudelo et al. (59), in a meta-analysis of 2 large double-blinded randomized controlled trials (60, 61), found that supplementation with vitamins C and E reduced the risk of preeclampsia (as compared with placebo) more strongly in women with BMI ≥30 than in those with BMI <30. Some observational studies have found that regular use of folic acid supplements (62) or multivitamins (63, 64) was associated with a reduced risk of preeclampsia, but only among lean women. One previous study modeled BMI continuously, assuming that the effect modification was linear (62); all others categorized BMI (63, 64). Investigators in these studies reported odds or risk ratios estimated via parametric regression models that relied on the assumption that there were no additional effect-measure modifiers present and that there was no interaction between confounders. In principle, the use of an ensemble machine learning algorithm to model both the propensity score and the outcome model can overcome these limitations. Machine learning methods are better at adapting to features of the data, such as the presence of important confounder-confounder interactions, curvilinear relationships, and other features that would have to be specified a priori when parametric regression is used.

Several previous studies have shown a stronger effect of some nutritional exposures on disease risk among individuals with a higher BMI than among leaner people (21–24). There are several possible explanations for why high-fruit- or high-vegetable-density dietary patterns may protect against preeclampsia to a greater extent in women with a high BMI than in leaner women. Multiple micronutrient deficiencies, including deficiencies in vitamin D, antioxidants, folate, calcium, and iron, are common among overweight or obese individuals (18, 19). These deficiencies may be due to metabolic changes that alter nutrient availability (18). Additionally, obesity is a state of increased systemic and adipose-tissue–specific oxidative stress, which results in

Table 3. Association[a] Between Periconceptional Vegetable Density and Risk of Preeclampsia, Overall and by Prepregnancy Body Mass Index (Effect-Measure Modifier), in the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be, 2010–2013

BMI Category and Total Vegetable Density, cups/1,000 kcal	No. of Women at Risk	Preeclampsia Cases		Adjusted[b] No. of Excess Cases per 100 Deliveries	95% CI
		No.	%		
Overall					
<1.3[c]	7,841	726	9.3	0	Referent
≥1.3	1,571	105	6.7	−3.0	−4.6, −1.5
Prepregnancy BMI category					
Underweight or normal weight (<25.0)					
<1.3	4,429	302	6.8	0	Referent
≥1.3	1,004	57	5.7	−2.3	−4.0, −0.7
Overweight (25.0–29.9)					
<1.3	1,691	183	10.8	0	Referent
≥1.3	347	19	5.5	−6.7	−9.4, −3.9
Obese (≥30.0)					
<1.3	1,721	241	14.0	0	Referent
≥1.3	220	29	13.2	−2.1	−7.2, 3.0

Abbreviations: BMI, body mass index; CI, confidence interval.

[a] Calculated using targeted maximum likelihood estimation with Super Learner.

[b] Results were adjusted for Healthy Eating Index 2010 components, maternal race/ethnicity, smoking, education, marital status, age, health insurance, acculturation, pregnancy planning, preexisting diabetes or chronic hypertension, sleep quality, physical activity level, and depression, anxiety, and stress assessment scores, as well as prepregnancy BMI as appropriate.

[c] 1.3 cups of vegetables/1,000 kcal was the 80th percentile of the vegetable density distribution.

a greater need for antioxidants to scavenge the increased production of reactive oxygen species (20). As a result, dietary patterns rich in fruits and vegetables may be more important for reducing the risk of preeclampsia among women with obesity.

A downside to our approach was the reliance on a dichotomous primary exposure (58). We chose to dichotomize total fruit and total vegetable consumption at the 80th percentiles of their respective distributions because this corresponds to the cutpoint for the upper quintile, which is often of interest in nutritional epidemiology, and it approximates the recommended intake defined in the Dietary Guidelines for Americans (2). While future work should explore the effects of continuous densities, our primary interest in this work was to explore whether there were meaningful differences in preeclampsia risk if women consumed more than a prespecified threshold amount of fruits and vegetables (which roughly corresponds to a "representative intervention" related to stochastic dynamic treatment strategies (65, 66)) and whether these effects differed across the spectrum of BMI.

FFQ data have been shown to be subject to a greater degree of systematic measurement error than other self-reporting methods, at least for absolute intakes (67, 68). This error can have unexpected effects, but attenuation of associations and reduced precision are common (69). Validation studies using the few recovery biomarkers identified have shown that potassium intake is better estimated from self-reported data than is energy or sodium intake (67, 68). Because fruits and vegetables are important food sources of potassium, but not energy or sodium, they may be reported with less error than other dietary components. Furthermore, densities are better estimated in FFQ data than absolute intakes of nutrients, indicating that our adjustment for energy intake via the density approach helped to reduce the bias (70). Further, we have no reason to believe that women who later develop preeclampsia have greater error in their measurement of the exposures.

Our sample was nulliparous and had a lower prevalence of overweight and obesity than the general US population. We sacrificed some generalizability by using a cohort of nulliparous women to take advantage of the large, geographically and racially/ethnically diverse nuMoM2b sample. Finally, naive use of machine learning methods can yield results that perform as poorly as misspecified parametric models (71, 72). To avoid these problems, we used a diverse set of machine learning algorithms in our ensemble learner, as well as an additional layer of sample splitting, which has been shown to yield optimally performing algorithms in simple and complex settings (43). The measures of prepregnancy BMI and periconceptional dietary intake were ascertained at visit 1. Thus, it is possible that measures of diet caused changes in measures of BMI.

Analyses targeting effect-measure modification are essential in tailoring public health recommendations to specific subpopulations, but several challenges arise when data

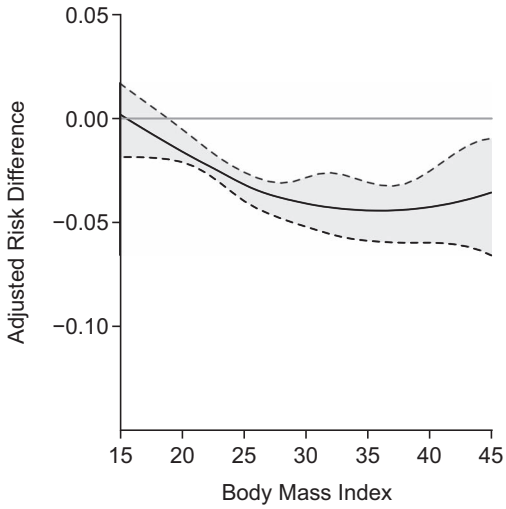

Figure 2. Adjusted difference in the risk of preeclampsia associated with a vegetable density of ≥1.3 cups/1,000 kcal versus <1.3 cups/1,000 kcal, according to prepregnancy body mass index, Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be, 2010–2013. Risk differences were adjusted for Healthy Eating Index 2010 components, maternal race/ethnicity, smoking, education, marital status, age, health insurance, acculturation, pregnancy planning, preexisting diabetes or chronic hypertension, sleep quality, physical activity level, and depression, anxiety, and stress assessment scores. The solid line represents the point estimate, and the dashed lines and shaded area represent uncertainty bands.

sets contain numerous continuous, categorical, and binary confounders and continuous effect-measure modifiers. Such challenges are commonly encountered in nutritional epidemiology. Machine learning methods can be used to deal with many of these challenges, and they provide informative results that can be used to establish empirically grounded guidelines on optimal dietary patterns for pregnant women. Our results suggest that dietary patterns with a high density of fruits and vegetables may be more beneficial for reducing preeclampsia among women affected by overweight or obesity than among lean women. Additional research is needed to determine whether dietary guidelines for pregnant women may require tailoring by maternal prepregnancy BMI to optimize pregnancy and birth outcomes.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Lisa M. Bodnar, Abigail R. Cartus, Sara M. Parisi); Department of Obstetrics, Gynecology, and Reproductive Sciences, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Lisa M. Bodnar, Katherine P. Himes, Hyagriv N. Simhan); Magee-Womens Research Institute, Pittsburgh, Pennsylvania, United States (Lisa M. Bodnar, Katherine P. Himes, Hyagriv N. Simhan); Department of Statistics and Data Science, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States (Edward H. Kennedy); School of Public Health and Health Systems, University of Waterloo, Waterloo, Ontario, Canada (Sharon I. Kirkpatrick); RTI International, Research Triangle Park, North Carolina, United States (Corette B. Parker); Department of Obstetrics and Gynecology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States (William A. Grobman); Department of Obstetrics and Gynecology, School of Medicine, University of Utah, Salt Lake City, Utah, United States (Robert M. Silver); Department of Obstetrics and Gynecology, School of Medicine, University of California, Irvine, Irvine, California, United States (Deborah A. Wing); Miller Children's and Women's Hospital, Long Beach, California, United States (Deborah A. Wing); Long Beach Memorial Medical Center, Long Beach, California, United States (Deborah A. Wing); Department of Obstetrics and Gynecology, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States (Samuel Perry); and Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States (Ashley I. Naimi).

This work was funded by National Institutes of Health grant R01 HD102313 (to L.M.B. and A.I.N.), as well as the following National Institutes of Health grants awarded to various companies and universities: grant U10 HD063036 (RTI International, Research Triangle Park, North Carolina); grant U10 HD063072 (Case Western Reserve University); grant U10 HD063047 (Columbia University); grant U10 HD063037 (Indiana University); grant U10 HD063041 (University of Pittsburgh); grant U10 HD063020 (Northwestern University); grant U10 HD063046 (University of California, Irvine); grant U10 HD063048 (University of Pennsylvania); and grant U10 HD063053 (University of Utah). Support was also provided by respective Clinical and Translational Science Institutes to Indiana University (grant UL1TR001108) and the University of California, Irvine (grant UL1TR000153).

The data described in this article and accompanying software code will not be made available until they are released to the public by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

All software code with which to reproduce this analysis can be found on GitHub (73).

The views expressed in this article are those of the authors and do not reflect those of the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

1. Mokdad AH, Ballestros K, Echko M, et al. The state of US health, 1990–2016: burden of diseases, injuries, and risk factors among US states. *JAMA*. 2018;319(14):1444–1472.

2. Dietary Guidelines Advisory Committee. *Scientific Report of the 2020 Dietary Guidelines Advisory Committee: Advisory Report to the Secretary of Agriculture and the Secretary of Health and Human Services*. Washington, DC: Agricultural Research Service, US Department of Agriculture; 2020. https://doi.org/10.52570/DGAC2020. Accessed February 24, 2022.
3. Kennedy EH. Nonparametric causal effects based on incremental propensity score interventions. *J Am Stat Assoc.* 2018;114(526):1–12.
4. Robins JM, Greenland S. Comment on "causal inference without counterfactuals" by A.P. Dawid. *J Am Stat Assoc.* 2000;95(450):477–482.
5. Lee Y, Neider JA. Conditional and marginal models: another view. *Stat Sci.* 2004;19(2):219–228.
6. Orhon FS, Ulukol B, Kahya D, et al. The influence of maternal smoking on maternal and newborn oxidant and antioxidant status. *Eur J Pediatr.* 2009;168(8): 975–981.
7. Institute of Medicine (US) Panel on Dietary Antioxidants and Related Compounds. *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids*. Washington, DC: National Academies Press; 2000.
8. Vanlint S. Vitamin D and obesity. *Nutrients.* 2013;5(3): 949–956.
9. Holick MF. Vitamin D deficiency. *N Engl J Med.* 2007; 357(3):266–281.
10. Foster JC, Taylor JM, Ruberg SJ. Subgroup identification from randomized clinical trial data. *Stat Med.* 2011;30(24): 2867–2880.
11. Garcia-Magarinos M, Lopez-de-Ullibarri I, Cao R, et al. Evaluating the ability of tree-based methods and logistic regression for the detection of SNP-SNP interaction. *Ann Hum Genet.* 2009;73(3):360–369.
12. Gromping U. Variable importance assessment in regression: linear regression versus random forest. *Am Stat.* 2009;63(4): 308–319.
13. Yang PY, Yang YH, Zhou BB, et al. A review of ensemble methods in bioinformatics. *Curr Bioinform.* 2010;5(4): 296–308.
14. Hastie T, Tibshirani R, Friedman JJH. *The Elements of Statistical Learning*. New York, NY: Springer Publishing Company; 2009.
15. Denisko D, Hoffman MM. Classification and interaction in random forests. *Proc Natl Acad Sci U S A.* 2018;115(8): 1690–1692.
16. Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. *Eur J Epidemiol.* 2018;33(5):459–464.
17. Bodnar LM, Cartus AR, Kirkpatrick SI, et al. Machine learning as a strategy to account for dietary synergy: an illustration based on dietary intake and adverse pregnancy outcomes. *Am J Clin Nutr.* 2020;111(6):1235–1243.
18. Bodnar LM, Parrott MS. Intervention strategies to improve outcome in obese pregnancies: micronutrients and dietary supplements. In: Poston L, Gillman MW, eds. *Maternal Obesity*. Cambridge, United Kingdom: Cambridge University Press; 2012:199–208.
19. Scholing JM, Olthof MR, Jonker FA, et al. Association between pre-pregnancy weight status and maternal micronutrient status in early pregnancy. *Public Health Nutr.* 2018;21(11):2046–2055.
20. Vincent HK, Innes KE, Vincent KR. Oxidative stress and potential interventions to reduce oxidative stress in overweight and obesity. *Diabetes Obes Metab.* 2007;9(6): 813–839.
21. Bodnar LM, Tang G, Ness RB, et al. Periconceptional multivitamin use reduces the risk of preeclampsia. *Am J Epidemiol.* 2006;164(5):470–477.
22. Catov JM, Bodnar LM, Ness RB, et al. Association of periconceptional multivitamin use and risk of preterm or small-for-gestational-age births. *Am J Epidemiol.* 2007; 166(3):296–303.
23. Villar J, Abdel-Aleem H, Merialdi M, et al. World Health Organization randomized trial of calcium supplementation among low calcium intake pregnant women. *Am J Obstet Gynecol.* 2006;194(3):639–649.
24. Fall CH, Fisher DJ, Osmond C, et al. Multiple micronutrient supplementation during pregnancy in low-income countries: a meta-analysis of effects on birth size and length of gestation. *Food Nutr Bull.* 2009;30(4 suppl):S533–S546.
25. Haas DM, Parker CB, Wing DA, et al. A description of the methods of the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b). *Am J Obstet Gynecol.* 2015;212(4):539.e1–539.e24.
26. Block G, Hartman AM, Dresser CM, et al. A data-based approach to diet questionnaire design and testing. *Am J Epidemiol.* 1986;124(3):453–469.
27. Block G, Woods M, Potosky A, et al. Validation of a self-administered diet history questionnaire using multiple diet records. *J Clin Epidemiol.* 1990;43(12):1327–1335.
28. Johnson BA, Herring AH, Ibrahim JG, et al. Structured measurement error in nutritional epidemiology: applications in the Pregnancy, Infection, and Nutrition (PIN) Study. *J Am Stat Assoc.* 2007;102(479):856–866.
29. Mares-Perlman JA, Klein BE, Klein R, et al. A diet history questionnaire ranks nutrient intakes in middle-aged and older men and women similarly to multiple food records. *J Nutr.* 1993;123(3):489–501.
30. Boucher B, Cotterchio M, Kreiger N, et al. Validity and reliability of the Block98 food-frequency questionnaire in a sample of Canadian women. *Public Health Nutr.* 2006;9(1): 84–93.
31. Block G, Coyle LM, Hartman AM, et al. Revision of dietary analysis software for the Health Habits and History Questionnaire. *Am J Epidemiol.* 1994;139(12):1190–1196.
32. Epidemiology and Genomics Research Program, National Cancer Institute. *Diet∗Calc Analysis Program, Version 1.5.0*. Bethesda, MD: National Cancer Institute; 2012.
33. Food Surveys Research Group, Agricultural Research Service, US Department of Agriculture. *USDA Food and Nutrient Database for Dietary Studies, Version 1.0.* Beltsville, MD: US Department of Agriculture; 2004.
34. Bowman SA, Friday JE, Moshfegh A. MPED databases for downloading. MPED 2.0. (MyPyramid Equivalents Food Database, version 2.0). http://www.ars.usda.gov/Services/docs.htm?docid=17565. Published 2008. Modified May 13, 2021. Accessed September 8, 2016.
35. Guenther PM, Casavale KO, Reedy J, et al. Update of the Healthy Eating Index: HEI-2010. *J Acad Nutr Diet.* 2013; 113(4):569–580.
36. Facco FL, Parker CB, Reddy UM, et al. Association between sleep-disordered breathing and hypertensive disorders of pregnancy and gestational diabetes mellitus. *Obstet Gynecol.* 2017;129(1):31–41.
37. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic*. Geneva, Switzerland: World Health Organization; 2000.
38. Cox J. Origins and development of the 10-item Edinburgh Depression Scale. In: Cox J, Holden J, Royal College of Psychiatrists, eds. *Perinatal Psychiatry: Use and Misuse of*

the Edinburgh Postnatal Depression Scale. London, United Kingdom: Edward Gaskell Publishers; 1994:25–36.
39. Spielberger CD, Gorsuch RL, Lushene RE. *STAI Manual for State-Trait Anxiety Inventory ("Self-Evaluation Questionnaire")*. Palo Alto, CA: Consulting Psychologists Press; 1970.
40. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav.* 1983;24(4):385–396.
41. Levine DW, Kaplan RM, Kripke DF, et al. Factor structure and measurement invariance of the Women's Health Initiative Insomnia Rating Scale. *Psychol Assess.* 2003;15(2):123–136.
42. van der Laan MJ, Luedtke AR. *Targeted Learning of an Optimal Dynamic Treatment, and Statistical Inference Inference for Its Mean Outcome*. (UC Berkeley Division of Biostatistics Working Paper 329). Berkeley, CA: University of California, Berkeley; 2014. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.684.7117&rep=rep1&type=pdf. Accessed February 24, 2022.
43. Kennedy EH. Optimal doubly robust estimation of heterogeneous causal effects [preprint]. *arXiv.* 2020. https://arxiv.org/pdf/2004.14497.pdf. Accessed February 24, 2022.
44. Foster DJ, Syrgkanis V. Orthogonal statistical learning [preprint]. *arXiv.* 2020. (https://arxiv.org/abs/1901.09036). Accessed February 24, 2022.
45. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol.* 2008;8(1):70.
46. Hejazi NS, Coyle JR, van der Laan MJ. tmle3shift: Targeted Learning of the Causal Effects of Stochiastic Interventions. (R package, version 0.2.0). https://github.com/tlverse/tmle3shift. Published August 29, 2021. Accessed February 24, 2022.
47. Wolpert DH. Stacked generalization. *Neural Netw.* 1992;5(2):241–259.
48. Breiman L. Stacked regressions. *Machine Learning.* 1996;24(1):49–64.
49. Benkeser D, Cai W, van der Laan MJ. A nonparametric super-efficient estimator of the average treatment effect. *Statist Sci.* 2020;35(3):484–495.
50. Wasserman L. *All of Nonparametric Statistics*. New York, NY: Springer Publishing Company; 2006.
51. Kennedy EH, Ma Z, McHugh MD, et al. Nonparametric methods for doubly robust estimation of continuous treatment effects. *J R Stat Soc Series B Stat Methodol.* 2017;79(4):1229–1245.
52. Ghazaleh D, Lee KJ, Simpson JA, et al. Handling missing data for causal effect estimation in cohort studies using targeted maximum likelihood estimation. *Int J Epidemiol.* 2021;50(suppl 1):150.
53. Altman DG, Royston P. The cost of dichotomizing continuous variables. *BMJ.* 2006;332(7549):1080.
54. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* 2006;25(1):127–141.
55. Zhang C, Williams MA, King IB, et al. Vitamin C and the risk of preeclampsia—results from dietary questionnaire and plasma assay. *Epidemiology.* 2002;13(4):409–416.
56. Raghavan R, Dreibelbis C, Kingshipp BL, et al. Dietary patterns before and during pregnancy and maternal outcomes: a systematic review. *Am J Clin Nutr.* 2019;109(suppl 1):705S–728S.
57. Vieira MC, Poston L, Fyfe E, et al. Clinical and biochemical factors associated with preeclampsia in women with obesity. *Obesity (Silver Spring).* 2017;25(2):460–467.
58. Hamad R, Collin DF, Baer RJ, et al. Association of revised WIC food package with perinatal and birth outcomes: a quasi-experimental study. *JAMA Pediatr.* 2019;173(9):845–852.
59. Conde-Agudelo A, Romero R, Kusanovic JP, et al. Supplementation with vitamins C and E during pregnancy for the prevention of preeclampsia and other adverse maternal and perinatal outcomes: a systematic review and metaanalysis. *Am J Obstet Gynecol.* 2011;204(6):503e1–503e12.
60. Villar J, Purwar M, Merialdi M, et al. World Health Organisation multicentre randomised trial of supplementation with vitamins C and E among pregnant women at high risk for pre-eclampsia in populations of low nutritional status from developing countries. *BJOG.* 2009;116(6):780–788.
61. Poston L, Briley AL, Seed PT, et al. Vitamin C and vitamin E in pregnant women at risk for pre-eclampsia (VIP trial): randomised placebo-controlled trial. *Lancet.* 2006;367(9517):1145–1154.
62. Martinussen MP, Bracken MB, Triche EW, et al. Folic acid supplementation in early pregnancy and the risk of preeclampsia, small for gestational age offspring and preterm delivery. *Eur J Obstet Gynecol Reprod Biol.* 2015;195:94–99.
63. Bodnar LM, Ness RB, Markovic N, et al. The risk of preeclampsia rises with increasing prepregnancy body mass index. *Ann Epidemiol.* 2005;15(7):475–482.
64. Catov JM, Nohr EA, Bodnar LM, et al. Association of periconceptional multivitamin use with reduced risk of preeclampsia among normal-weight women in the Danish National Birth Cohort. *Am J Epidemiol.* 2009;169(11):1304–1311.
65. VanderWeele TJ, Hernan MA. Causal inference under multiple versions of treatment. *J Causal Inference.* 2013;1(1):1–20.
66. Young JG, Logan RW, Robins JM, et al. Inverse probability weighted estimation of risk under representative interventions in observational studies. *J Am Stat Assoc.* 2019;114(526):938–947.
67. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. *Am J Epidemiol.* 2015;181(7):473–487.
68. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol.* 2014;180(2):172–188.
69. Freedman LS, Schatzkin A, Midthune D, et al. Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst.* 2011;103(14):1086–1092.
70. Subar AF, Freedman LS, Tooze JA, et al. Addressing current criticism regarding the value of self-report dietary data. *J Nutr.* 2015;145(12):2639–2645.
71. Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. *J Econom.* 2018;21(1):C1–C68.
72. Naimi AI, Mishler AE, Kennedy EH. Challenges in obtaining valid causal effect estimates with machine learning algorithms [published online ahead of print July 15, 2021]. *Am J Epidemiol*. https://doi.org/10.1093/aje/kwab201.
73. Parisi SM. DietML_bmi-modifier-tmle3. https://github.com/smparisi/DietML_bmi-modifier-tmle3. Published March 7, 2022. Accessed March 7, 2022.