

A Risk Score for Childhood Obesity in an Urban Latino Cohort

Jacob O. Robson, MD, MAS¹, Sofia G. Verstraete, MD, MAS¹, Stephen Shiboski, PhD², Melvin B. Heyman, MD, MPH¹, and Janet M. Wojcicki, PhD, MPH¹

Objective To assess whether individual obesity risk factors, present during gestation, and the first 6 months of life, can be combined into a simple prognostic model that has the ability to accurately predict childhood obesity at age 5 years in a high-risk cohort.

Study design A total of 201 Latina women were recruited during pregnancy, and their infants followed longitudinally. Ten risk factors for childhood obesity were included in an initial logistic model; a second reduced model was created via stepwise deletion (confirmed with nonparametric conditional random forest classifier), after which 5 risk factors remained. From each model, an obesity risk equation was derived, and an obesity risk score was generated for each patient. Derived algorithms were assessed using discrimination, calibration, and via predictive statistics.

Results Of the 166 children followed through age 5 years, 56 (32%) met criteria for childhood obesity. Discrimination accuracy for both derivation models was excellent, and after optimism-corrected bootstrapping, both models showed meaningful clinical performance. Both models were adequately calibrated, showed strong sensitivity and negative predictive value at conservatively set obesity risk thresholds, and displayed excellent specificity among those classified as highest risk. Birth weight z-score and change in weight-for-age z-score between birth and 6 months were the risk factors with the strongest contribution to the obesity risk score.

Conclusions Obesity risk algorithms are reliable in their prediction of childhood obesity and have the potential to be integrated into the electronic medical record. These models could provide a filter for directing early prevention resources to children with high obesity risk but should be evaluated in a larger external dataset. (*J Pediatr* 2016; ■: ■-■).

Despite the medical and financial severity of childhood obesity, it has proven difficult to treat. Longitudinal data show that once childhood obesity is present, it is likely to persist into adolescence and adulthood.¹ Pediatric health care practitioners (HCPs) have, thus, turned their focus to obesity prevention. Many recent studies have focused on single risk factors that are highly associated childhood obesity and are present during gestation or early infancy, such as maternal smoking and rapid early infant weight gain, as potential prevention targets.² However, as the development of childhood obesity is influenced by genetic, environmental, and socioeconomic factors, in a complex interaction, targeting single obesity risk factors for intervention may be ineffective.

Prognostic modeling, whereby influence weights from multiple risk factors are combined to estimate an individual's risk of a medical outcome,³ may be useful. An accurate childhood obesity risk score, derived from the presence of known prenatal and early postnatal obesity risk factors, could provide a simple means of identifying infants at low risk of obesity and directing them to standard weight monitoring,⁴ while reserving intensive obesity prevention resources for those at high risk. This would be particularly useful in medical centers serving the urban poor, where the prevalence of obesity is often high,⁵ yet resources are low. The purpose of this study was to examine whether a prognostic model for childhood obesity could be derived from data gathered among an urban, Latino cohort, using only objective measures available from the medical record in a low resource setting.

Methods

Latina women living in the San Francisco area were recruited during their pregnancy for a prospective cohort study. The full recruitment protocol has been described previously.⁶ Demographic and general health characteristics of maternal participants were collected and maternal body mass index (BMI) was

BMI	Body mass index
EMR	Electronic medical record
HCP	Health care practitioner
NPV	Negative predictive value
PPV	Positive predictive value
WFA	Weight-for-age

From the ¹Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, and ²Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA

Data collection was funded to J.W. by the National Institutes of Health-National Institute of Diabetes and Digestive and Kidney Diseases (K01 DK080825), North American Society for Pediatric Gastroenterology, Hepatology and Nutrition Foundation and Hellman Family Foundation. The research was also supported in part by National Institutes of Health-National Center for Research Resources (UL1 RR024131). J.R. and S.V.'s analysis and writing were funded by National Institutes of Health (T32 DK007762). No study funding sources had any role in study design, analysis, writing or decision to submit for publication. The authors declare no conflicts of interest.

0022-3476/\$ - see front matter. Copyright © 2016 Elsevier Inc. All rights reserved.

<http://dx.doi.org/10.1016/j.jpeds.2016.01.055>

calculated from self-reported prepregnancy weight and height on the intake questionnaire. Mothers with preexisting diabetes, polycystic ovary syndrome, insulin dependent gestational diabetes, and those with health issues or beliefs that would prevent breastfeeding were excluded from the cohort. Infants were excluded at delivery if they had any special care needs or an Apgar score ≤ 7 at 5 minutes of life. A total of 201 mothers were enrolled in the study, 196 infants met criteria for participation, and 166 mother-child pairs (83% and 85%, respectively) remained in follow-up at 5 years.

At birth, 6 months, 1 year, and subsequent annual visits, anthropometric measures were obtained on child participants, using standard digital scales for weight and tape measurements for length. Weight and height were calculated based on sex-specific Centers for Disease Control growth references. A small percentage of the participants (<5% at each time point) could not attend study visits, so weight and height measures were extracted from the medical record. Data were obtained on the child's nutritional intake via maternal interviews.

The Committee on Human Research at University of California, San Francisco approved all study procedures. Written informed consent was obtained from all participants at study entry and follow-up visits. Data collection ran from 2006-2012, and statistical analysis was conducted from 2014-2015.

For our predictive modeling, we included only risk factors that were clearly defined, reliably measurable, and available in standard clinical settings. We only considered predictors that have been previously described as affecting infant or childhood weight status.^{6,7} Whenever possible, we avoided dichotomization or categorization of linear predictors.⁸ All continuous predictors were checked for nonlinearity. The primary outcome was childhood obesity at age 5 years, defined as BMI ≥ 95 th percentile, using Centers for Disease Control growth references.⁹

In reviewing our longitudinal data, we identified 19 candidate predictors of childhood obesity, including 11 prenatal/maternal and 8 early postnatal risk factors (Table I; available at www.jpeds.com). We applied our guiding principles in creating the predictive model, removing several maternal variables, including maternal depression, employment status, and years in the US, because of concerns over infrequent inclusion in the standard medical record and/or time burden on the HCP to document. Maternal smoking was excluded because of low prevalence of smoking in the cohort (both during pregnancy and in the first year of follow-up, reported smoking prevalence was <3%). Gestational age was excluded because of collinearity with birth weight and little independent predictive value.

Statistical Analyses

The 10 remaining candidate predictors were placed into a logistic regression model, referred to as the "full model." An obesity risk score was generated for each participant, with the regression constant serving as the intercept and the

beta-coefficient indicating the adjusted contribution of each predictor to the obesity risk (Table II; available at www.jpeds.com). The predicted risk of obesity was calculated using the formula $1/(1 + e^{-\text{risk score}})$.¹⁰ We also developed an alternate predictive regression model using two variable selection strategies: stepwise backward deletion using the 20% significance level as a criterion for variable retention, and based on variable importance rankings from a nonparametric conditional random forest classifier (Figure 1; available at www.jpeds.com).¹¹ Both approaches resulted in the same group of 5 predictors, which were used to fit a "reduced model." Risk scores were also created for each patient using this model. Regression diagnostics for each model included identification of observations with high leverage, assessment for pairwise interactions and nonlinearity.

Both the full model and the reduced model were evaluated for discrimination and calibration performance. Discrimination was assessed by creating a receiver operating characteristic curve, with concordance index (area under the receiver operating characteristic) ≥ 0.8 considered to be excellent accuracy and ≥ 0.75 clinically meaningful.¹² Model calibration was analyzed using the Hosmer-Lemeshow goodness of fit test, with $P < .05$ defined as failure of adequate agreement between estimated and observed values.¹² To test predictive accuracy, an arbitrary risk threshold (eg, obesity risk score >50th percentile) was established, and those above the threshold were labeled as positive on the prognostic test for childhood obesity. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for the prognostic test at several different risk thresholds were calculated, as were likelihood ratios.

Because of the small sample size in this specific cohort, we elected not to split the dataset into derivation and validation sets. Rather, we used the entire sample to develop the prediction models and estimated internal measures of prediction performance using the bootstrap with 1000 samples.¹³ Statistics were performed with Stata 13 (StataCorp LP, College Station, Texas) and R v 3.2.2 (Vienna, Austria).

Because 17% and 12% of individuals had missing observations for at least 1 predictor in the full and reduced models, respectively, we used chained multiple imputation to assess the sensitivity of our estimated scores to missing information.¹⁴ In both cases, estimates from imputed data were quite similar to those obtained using observed data, so we present results from the latter only.

Results

Our cohort had relatively low prevalence of higher education, employment, and English proficiency, comparable with the underserved, recent immigrant populations typically seen at urban safety net hospitals (Table III). Maternal prepregnancy BMI was reflective of national obesity trends among Latina women,¹⁵ with 33% of mothers in the overweight range and an additional 18% obese.

Table III. Maternal and early infant candidate predictors of childhood obesity by child's obesity status at age 5 years

	Total	Obese, N = 53	Not obese, N = 113
Maternal predictors	N = 201		
Mexican ethnicity (%)	123 (61)	32 (60)	67 (59)
English language proficiency (%)	46 (23)	14 (26)	26 (23)
Maternal age, y (SD)	26.3 (5.2)	25.6 (5.9)	26.8 (4.9)
Maternal prepregnancy BMI, kg/m ² (SD)	25.8 (5.5)	27.4 (6.9)	25.2 (4.6)*
Depression at intake or peripartum (%)	67 (33)	15 (28)	41 (36)
First child (%)	95 (47)	26 (49)	52 (46)
>High school education (%)	44 (22)	9 (18)	23 (21)
In US <5 y (%)	99 (49)	27 (51)	52 (46)
Cohabitation with partner (%)	166 (83)	44 (85)	94 (83)
Maternal smoking (%)	4 (2)	1 (2)	3 (3)
Employed (%)	57 (28)	14 (26)	32 (28)
Infant predictors, measured at birth	N = 196		
Male sex (%)	98 (50)	25 (47)	55 (49)
Gestational age, wk (SD)	39.3 (1.5)	39.6 (1.4)	39.1 (1.6)†
Vaginal delivery (%)	166 (85)	44 (83)	99 (88)
Birth weight z-score, CDC (SD)	-0.14 (0.9)	0.16 (0.9)	-0.24 (0.9)*
Infant predictors, measured over first 6 mo			
WFA z-score change between birth and 6 mo, CDC (N = 173, SD)	0.62 (1.1)	0.91 (1.2)	0.46 (1.1)†
Exclusively breastfed at 4-6 wk (N = 190, %)	71 (37)	16 (30)	48 (42)
Any breastfeeding at 6 mo (N = 180, %)	120 (66)	30 (57)	75 (66)
Introduction of solids (N = 169, %)			
<6 mo	98 (58)	31 (66)	61 (58)
>6 mo	71 (42)	16 (34)	45 (42)

CDC, Centers for Disease Control.

Boldface indicates statistical significance by *t* test.

**P* < .01.

†*P* < .05.

Only 7% of included children were born preterm (<37 weeks gestational age). The mean birth weight (3368 g, SD 476) was similar to 2005 data for term babies born to Hispanic mothers (mean 3369 g, SD 455) in the US and slightly below the overall US birth weight average (mean 3389 g, SD 466).¹⁶ Ninety-one percent of infants were breastfed at 4-6 weeks after birth, with 37% receiving breast milk as their only source of nutrition. Sixty-six percent of infants continued to breastfeed through the first 6 months of life, though over one-half of breastfeeding babies (57%) were supplemented with some formula. Children who remained in follow-up through the 3-, 4- and 5-year-old visits had obesity point prevalence of 28%, 25%, and 32%, respectively.

Candidate Predictor Analysis

Maternal prepregnancy BMI, infant birth weight z-score, and weight-for-age (WFA) z-score change between birth and 6 months were independent risk factors for childhood obesity in unadjusted analysis and remained significant predictors after adjustment in both regression models (Tables III and IV). Maternal age was a significant protector against childhood obesity in the reduced model but not the full model.

Model Performance

The regression output for the full model and the reduced model are displayed in Table IV. Both models contained predictors that conferred increased childhood obesity risk as they increased in value, marked by a positive beta-coefficient (eg, maternal prepregnancy BMI). They also had variables with negative beta-coefficients, such that increase (eg, maternal age) was associated with a decrease in obesity risk.

Discrimination accuracy for both models was excellent in the derivation cohort (full model concordance index = 0.84, 95% CI 0.77-0.91; reduced model concordance index = 0.82, 95% CI 0.74-0.89; Figure 2) and clinically meaningful in the optimism-corrected bootstrap-based estimates (full model concordance index = 0.78; reduced model concordance index = 0.76). Both models displayed adequate calibration (*P* values for Hosmer-Lemeshow test >.5).

To illustrate how risk scores were calculated, consider the following patient from the cohort: female sex, birth weight z-score 0.33, WFA z-score change between birth and 6 months 2.44, solids introduced at <6 months, not exclusively breastfed at 4-6 weeks, and no breastfeeding at 6 months. This child was born to an 18 year-old mother, who had no previous children, no English language proficiency, and a prepregnancy BMI of 28.4. Using the prediction algorithms in Table II, we arrive at a risk score of 3.89 using the full model algorithm (3.22 with reduced model), with a predicted obesity risk of 98% (96% with reduced model). This patient was obese at age 5 years, with a BMI >99.9th percentile.

Birth weight z-score and change in WFA z-score between birth and 6 months were consistently the strongest contributors, across both models, to the obesity risk score. There were modest protective effects of breastfeeding and later introduction of solid foods in the full model, but only exclusive breastfeeding at 4-6 weeks met criteria for inclusion in the reduced model.

Predictive Statistics

The predictive metrics of the algorithms are displayed in Table V. The sensitivity and NPV were high (>90%) when

Table IV. Multivariable logistic regression models for prediction of childhood obesity

Predictor	Full model		Reduced model	
	Beta-coefficient (SE)	OR (95% CI)	Beta-coefficient (SE)	OR (95% CI)
WFA z-score change between birth and 6 mo, CDC	1.35 (0.32)	3.85 (2.04-7.28)*	1.14 (0.28)	3.15 (1.82-5.46)*
Birth weight z-score, CDC	1.7 (0.41)	5.47 (2.47-12.1)*	1.39 (0.35)	4.02 (2.01-8.03)*
Maternal prepregnancy BMI	0.11 (0.05)	1.12 (1.02-1.22)†	0.1 (0.04)	1.11 (1.02-1.2)†
Maternal age	-0.09 (0.05)	0.92 (0.83-1.01)	-0.1 (0.05)	0.9 (0.82-0.99)†
Exclusively breastfed at 4-6 wk	-0.72 (0.48)	0.48 (0.18-1.26)	-0.76 (0.46)	0.47 (0.19-1.15)
Introduction of solids after 6 mo	-0.7 (0.51)	0.5 (0.18-1.37)		
Male sex	0.57 (0.49)	1.76 (0.67-4.6)		
Any breastfeeding at 6 mo	-0.33 (0.51)	0.72 (0.27-1.94)		
First child	-0.5 (0.54)	0.61 (0.23-1.62)		
English language proficiency	0.55 (0.54)	1.73 (0.6-5)		
Intercept (SE)		-1.58 (1.7)		-1.14 (1.50)
Concordance index		0.84		0.82

CDC, Centers for Disease Control.

Boldface indicates statistical significance by *t* test.

**P* < .001.

†*P* < .05.

the obesity risk score threshold was set at the 25th percentile, yet remain strong when the threshold is raised to the 50th percentile, especially in the full model algorithm (86% sensitivity, 91% NPV). In setting the risk threshold higher (eg, at the 75th percentile and above), we found that patients above the threshold reliably went on to childhood obesity (high specificity and positive likelihood ratio), but this came at the cost of lowering sensitivity and NPV significantly.

Discussion

Previously published childhood obesity prediction algorithms¹⁷⁻²² were most often derived from European cohorts,

with childhood obesity rates as low as 3%.¹⁷ While adhering to predefined guiding principles in candidate predictor selection in a high-risk, United States-based cohort, we were able to derive 2 prognostic algorithms for childhood obesity with adequate discrimination and calibration and predictive capabilities with potential for meaningful clinical applications and cost savings.

Current recommendations for the universal assessment of childhood obesity risk includes monitoring BMI at well child visits and considering medical, behavioral, and attitude risks,⁴ but there is no method by which the clinician can combine individual risk factors in a quantifiable risk assessment. Prognostic modeling has the potential to aid childhood obesity prediction tremendously; the weighted

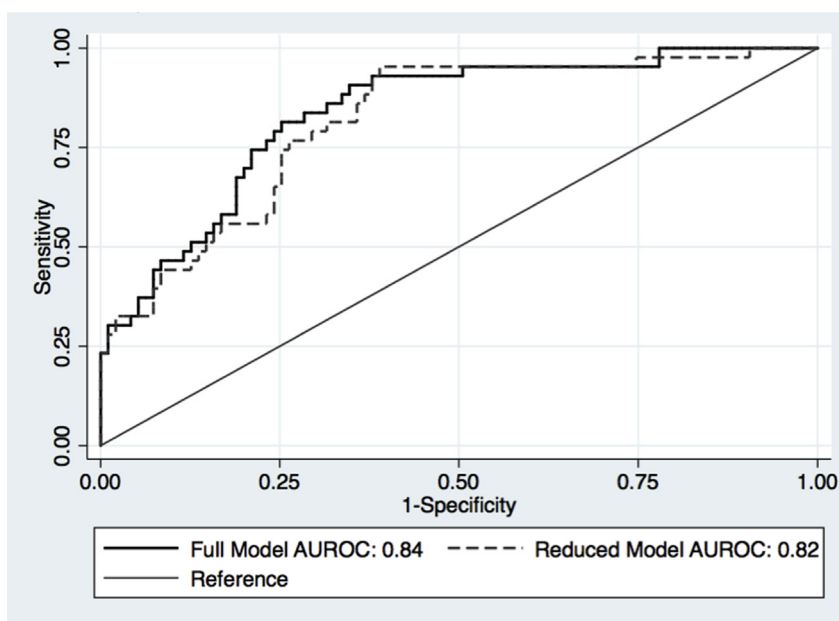


Figure 2. Receiver operating characteristic curve for obesity prediction models. Prediction algorithms applied to derivation cohort. AUROC, area under the receiver operating characteristic.

Table V. Predictive accuracy of childhood obesity algorithms based on placement of risk threshold

Risk threshold	Population above threshold (%)	Sensitivity (%)		Specificity (%)		NPV (%)		PPV (%)		Positive likelihood ratio	
		Full model	Reduced model	Full model	Reduced model	Full model	Reduced model	Full model	Reduced model	Full model	Reduced model
25th Percentile	75	95	96	35	37	94	95	40	41	1.46	1.52
50th Percentile	50	86	80	66	64	91	88	54	51	2.55	2.23
75th Percentile	25	51	46	85	84	79	77	61	57	3.47	2.85
90th Percentile	10	30	24	99	97	76	73	93	79	28.72	7.97

algorithm allows infant HCPs to quickly obtain a risk score that puts their patient's obesity risk in contrast against population risk. Further, as all measures in the described models are either taken in the prenatal period or in the infant's first 6 months, there is potential for early intervention. Applying the obesity risk algorithm at the 6-month visit would allow infant HCPs to present tangible concern about obesity risk, starting the discussion around healthy weight and nutrition, while gauging the family's readiness for change.²³ Risk score data may be particularly useful in counseling Latino families, where cultural practices lead to perceptions of healthy weight that are skewed toward higher weight ranges.²⁴ Although consensus continues to build that early obesity prevention efforts are effective,²⁵ best timing, setting, and methodology of intervention continues to be a point of debate. Studies in obesity prevention during infancy are few and although some have shown promise,²⁶ additional study is necessary to show which interventions lead to sustained effects.²⁷

Applying risk thresholds across the spectrum of obesity risk could lead to an innovative, multitiered, prevention strategy. Using our full model prediction equation and risk score threshold at the 25th percentile, 94% of patients below this risk threshold were in the normal weight range at age 5 years (this is the NPV of the model using this risk threshold, true negatives divided by all testing negative). In contrast, 61% of those above the 75th percentile risk score (and 93% of those above the 90th percentile) were obese at age 5 years (this corresponds to the PPV, true positives divided by all testing positive). HCPs can use this type of data to allocate prevention resources within their own patient populations, safely directing those at low obesity risk to standard monitoring, while reserving high-value prevention resources, such as gym memberships, dietician visits, and in-home behavioral counseling, for patients at high risk of actually developing the disease outcome. Even though there is a vast literature on effectiveness of childhood obesity prevention efforts, there is an opportunity to study how intervention efficacy may differ across obesity risk strata. Identifying those at low risk and minimizing resources applied to this group could confer cost savings over efforts that prescribe obesity prevention resources to the entire population.

As electronic medical records (EMRs) become a standard of practice across the US,²⁸ childhood obesity risk scoring has the potential for immediate and impactful applications.

The models presented here lend themselves to the development of an EMR-based risk calculator. Most EMRs have the capability to generate a risk score automatically from risk factor data entered into the medical record and/or a clinic visit note. A pop-up message could alert providers to their patient's obesity risk score and whether they are above a set risk threshold where intervention is indicated. Risk score alerts could serve as an important tool for busy HCPs,^{21,29} who frequently fail to classify patients in the appropriate weight category³⁰ and may not appreciate rapid infant growth if contained within the normal weight range (eg, 15th% at birth to 75th% at 6 months).

Compared with other recently published predictive risk equations for childhood obesity, the models described in this study have several relative methodological advantages. First, the candidate predictors based on anthropometric measures are continuous, allowing for relative contribution of values across the spectrum of the variable. The beta-coefficients for birth weight z-score and WFA z-score change are stronger than those presented previously for dichotomized or categorized versions of these same variables. Second, other than self-reported maternal prepregnancy BMI, we only included candidate predictors with reliable, objective measures, minimizing measurement error. Lastly, we measured the obesity outcome at age 5 years, a time where children often display issues with portion control³¹ and are prone to choosing heavily advertised sugary beverages and snacks.³² Further, it has been shown that those who are obese entering school have higher risk of obesity into adulthood than obese toddlers.¹

One important limitation to this analysis is that the models derived have yet to be validated and evaluated for predictive accuracy in a larger, external dataset with similar demographics. Validation in an external dataset is critical in confirming the discrimination, calibration, and predictive accuracy of the models and will allow for further efficacy comparisons between the full and reduced models, as well as the establishment of the most clinically useful obesity risk thresholds in considering resource allocation. Further, because of the relatively small size of this dataset and some missing data across variables, several predictors that have been shown to be protective against obesity, such as breastfeeding in the first 6 months and timing of solids introduction, may have been underpowered to reach statistical significance and meet inclusion criteria in the reduced model.

Lastly, this is a “locally” derived model.¹⁷ Although it is likely to display high predictive accuracy among Latinos in urban areas, it may not generalize as well to mixed ethnicity and socioeconomic status cohorts. However, based on obesity risk data for other underserved minorities (African Americans, Native Americans) living in metropolitan areas, there is reason to believe that this model may still be useful for obesity prediction in these groups.^{5,33}

Further study is needed to show which prevention measures are most clinically effective (and cost-effective) in preventing obesity when directed towards infants and toddlers with a high obesity risk. ■

We thank Joel Simon, MD, MPH, Michael Kohn, MD, MPP, David Glidden, PhD, Saunak Sen, PhD, MS, and Emily Perito, MD, MAS, for mentorship in this work as part of the UCSF Master's Program in Advanced Science and Clinical Research.

Submitted for publication Aug 14, 2015; last revision received Dec 16, 2015; accepted Jan 21, 2016.

Reprint requests: Jacob O. Robson, MD, MAS, 81 N Mario Capecchi Dr, Salt Lake City, UT 84113. E-mail: jacob.robson@hsc.utah.edu

References

- Guo SS, Wu W, Chumlea WC, Roche AF. Predicting overweight and obesity in adulthood from body mass index values in childhood and adolescence. *Am J Clin Nutr* 2002;76:653-8.
- Taveras EM, Rifas-Shiman SL, Belfort MB, Kleinman KP, Oken E, Gillman MW. Weight status in the first 6 months of life and obesity at 3 years of age. *Pediatrics* 2009;123:1177-83.
- Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
- Barlow SE, Expert Committee. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics* 2007;120: S164-92.
- Goodell LS, Wakefield DB, Ferris AM. Rapid weight gain during the first year of life predicts obesity in 2- to 3-year-olds from a low-income, minority population. *J Community Health* 2009;34:370-5.
- Wojcicki JM, Holbrook K, Lustig RH, Epel E, Caughey AB, Muñoz RF, et al. Chronic maternal depression is associated with reduced weight gain in Latino infants from birth to 2 years of age. *PLoS One* 2011;6:e16737.
- Weng SF, Redsell SA, Swift JA, Yang M, Glazebrook CP. Systematic review and meta-analyses of risk factors for childhood overweight identifiable during infancy. *Arch Dis Child* 2012;97:1019-26.
- Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2005;25:127-41.
- Ogden CL, Flegal KM. Changes in terminology for childhood overweight and obesity. *Natl Health Stat Report* 2010;22:1-6.
- Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604.
- Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinformatics* 2008;9:307.
- Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York, NY: John Wiley and Sons; 2000.
- Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
- van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007;16:219-42.
- Hedley AA, Ogden CL, Johnson CL, Carroll MD, Curtin LR, Flegal KM. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. *JAMA* 2004;291:2847-50.
- Donahue SMA, Kleinman KP, Gillman MW, Oken E. Trends in birth weight and gestational length among singleton term births in the United States. *Obstet Gynecol* 2010;115:357-64.
- Morandi A, Meyre D, Lobbens S, Kleinman K, Kaakinen M, Rifas-Shiman SL, et al. Estimation of newborn risk for child or adolescent obesity: lessons from longitudinal birth cohorts. *PLoS One* 2012;7: e49919.
- Weng SF, Redsell SA, Nathan D, Swift JA, Yang M, Glazebrook C. Estimating overweight risk in childhood from predictors during infancy. *Pediatrics* 2013;132:e414-21.
- Druet C, Stettler N, Sharp S, Simmons RK, Cooper C, Davey Smith G, et al. Prediction of childhood obesity by infancy weight gain: an individual-level meta-analysis. *Paediatr Perinat Epidemiol* 2011;26: 19-26.
- Levine RS, Dahly DL, Rudolf MCJ. Identifying infants at risk of becoming obese: can we and should we? *Public Health* 2012;126:123-8.
- Santorelli G, Petherick ES, Wright J, Wilson B, Samiei H, Cameron N, et al. Developing prediction equations and a mobile phone application to identify infants at risk of obesity. *PLoS One* 2013;8:e71183.
- Graversen L, Sørensen TIA, Gerds TA, Petersen L, Sovio U, Kaakinen M, et al. Prediction of adolescent and adult adiposity outcomes from early life anthropometrics. *Obesity (Silver Spring)* 2014;23:162-9.
- Rhee KE, DeLago CW, Arscott-Mills T, Mehta SD, Krysko Davis R. Factors associated with parental readiness to make changes for overweight children. *Pediatrics* 2005;116:e94-101.
- Rich SS, DiMarco NM, Huettig C, Essery EV, Andersson E, Sanborn CF. Perceptions of health status and play activities in parents of overweight Hispanic toddlers and preschoolers. *Fam Community Health* 2005;28: 130-41.
- Waters E, de Silva-Sanigorski A, Hall BJ, Brown T, Campbell KJ, Gao Y, et al. Interventions for preventing obesity in children. *Cochrane Database Syst Rev* 2011;12:CD001871.
- Paul IM, Savage JS, Anzman SL, Beiler JS, Marini ME, Stokes JL, et al. Preventing obesity during infancy: a pilot study. *Obesity* 2010;19:353-61.
- Wen LM, Baur LA, Simpson JM, Rissel C, Wardle K, Flood VM. Effectiveness of home based early intervention on children's BMI at age 2: randomised controlled trial. *BMJ* 2012;344:e3732.
- Adler-Milstein J, DesRoches CM, Furukawa MF, Worzala C, Charles D, Kralovec P, et al. More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most. *Health Aff (Millwood)* 2014;33:1664-71.
- Rattay KT, Ramakrishnan M, Atkinson A, Gilson M, Drayton V. Use of an electronic medical record system to support primary care recommendations to prevent, identify, and manage childhood obesity. *Pediatrics* 2008;123:S100-7.
- Riley MR, Bass NM, Rosenthal P, Merriman RB. Underdiagnosis of pediatric obesity and underscreening for fatty liver disease and metabolic syndrome by pediatricians and pediatric subspecialists. *J Pediatr* 2005; 147:839-42.
- Rolls BJ, Engell D, Birch LL. Serving portion size influences 5-year-old but not 3-year-old children's food intakes. *J Am Diet Assoc* 2000;100: 232-4.
- Chamberlain LJ, Wang Y, Robinson TN. Does children's screen time predict requests for advertised products? Cross-sectional and prospective analyses. *Arch Pediatr Adolesc Med* 2006;160:363-8.
- Kimbro RT, Brooks-Gunn J, McLanahan S. Racial and ethnic differentials in overweight and obesity among 3-year-old children. *Am J Public Health* 2007;97:298-305.

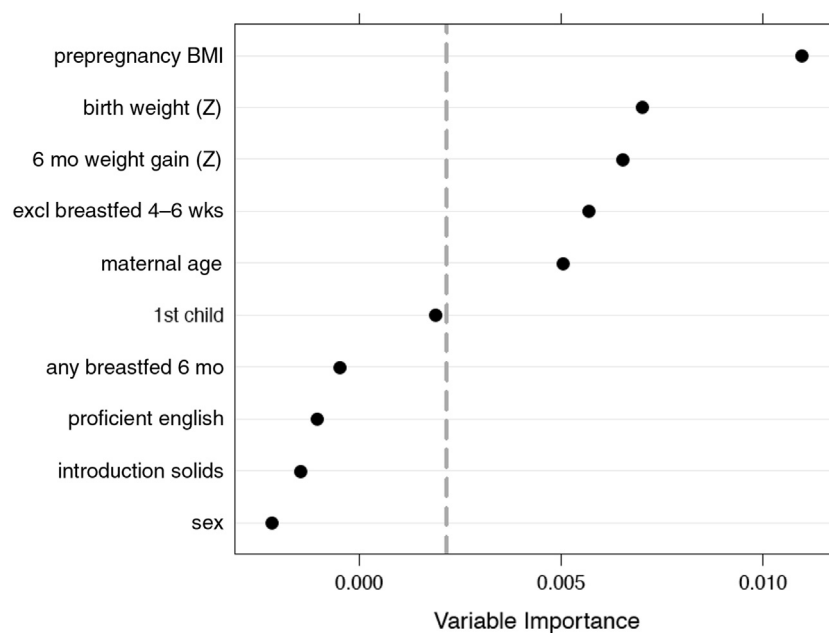


Figure 1. Variable importance rankings from a nonparametric conditional random forest classifier.

Table I. Candidate predictors of childhood obesity within the dataset

Infant	
Delivery type (dichotomous, cesarean or vaginal)	
Gestational age (continuous, in y)	
Sex (dichotomous, male or female)	
Birth weight (continuous, in z-scores, plotted on both CDC and WHO references)	
Infant weight gain (continuous, in z-score change between birth and 6 mo, plotted on both CDC and WHO references)	
Exclusively breastfed at 4-6 wk, no formula or solids (dichotomous, yes or no)	
Any breastfeeding at 6 mo (dichotomous, yes or no)	
Timing of introduction of solids (dichotomous, <6 mo or >6 mo)	
Maternal	
Prepregnancy BMI (continuous, in kg/m ²)	
First child (dichotomous, yes or no)	
Age (continuous, in y)	
Employment (dichotomous, prior to or during pregnancy, yes or no)	
Education (dichotomous, beyond high school, yes or no)	
Depression (dichotomous, clinical depression at intake or immediately postpartum, yes or no)	
Cohabitation status (dichotomous, living with partner, yes or no)	
Y living in the US (dichotomous, <5 y, yes or no)	
English language proficiency (dichotomous, proficient in English, yes or no)	
Maternal smoking (dichotomous, smoking at intake or postpartum period, yes or no)	
Mexican ethnicity (dichotomous, yes or no)	

CDC, Centers for Disease Control; WHO, World Health Organization.

Table II. Childhood obesity risk score algorithms

Full model	
$-1.58 + (1.35 \times \text{CDC WFA z-score change between birth and 6 mo}) + (1.7 \times \text{CDC birth weight z-score}) + (0.11 \times \text{maternal prepregnancy BMI}) - (0.72 \times \text{exclusive breastfeeding at 4-6 wk}) - (0.33 \times \text{any breastfeeding at 6 mo}) + (0.57 \times \text{sex}) - (0.09 \times \text{maternal age}) - (0.7 \times \text{introduction of solids after 6 mo}) - (0.5 \times \text{first child}) + (0.55 \times \text{English language proficiency})$	
Reduced model	
$-1.14 + (1.14 \times \text{CDC WFA z-score change between birth and 6 mo}) + (1.39 \times \text{CDC birth weight z-score}) + (0.1 \times \text{maternal prepregnancy BMI}) - (0.1 \times \text{maternal age}) - (0.76 \times \text{exclusive breastfeeding at 4-6 wk})$	

CDC, Centers for Disease Control.