American Journal of Epidemiology

Vol. 00, No. 00
https://doi.org/10.1093/aje/kwab176
Advance Access publication:

Original Contribution

Estimates of Childhood Overweight and Obesity at the Region, State, and County Levels: A Multilevel Small-Area Estimation Approach

Anja Zgodic, Jan M. Eberth, Charity B. Breneman, Marilyn E. Wende, Andrew T. Kaczynski, Angela D. Liese, and Alexander C. McLain*

* Correspondence to Dr. Alexander C. McLain, Department of Epidemiology and Biostatistics Arnold School of Public Health University of South Carolina 915 Greene Street Room 450 Columbia, SC 29208 (e-mail: mclaina@mailbox.sc.edu).

Initially submitted November 10, 2020; accepted for publication June 10, 2021.

Local-level childhood overweight and obesity data are often used to implement and evaluate community programs, as well as allocate resources to combat overweight and obesity. The most current substate estimates of US childhood obesity use data collected in 2007. Using a spatial multilevel model and the 2016 National Survey of Children's Health, we estimated childhood overweight and obesity prevalence rates at the Census regional division, state, and county levels using small-area estimation with poststratification. A sample of 24,162 children aged 10–17 years was used to estimate a national overweight and obesity rate of 30.7% (95% confidence interval: 27.0%, 34.9%). There was substantial county-to-county variability (range, 7.0% to 80.9%), with 31 out of 3,143 counties having an overweight and obesity rate significantly different from the national rate. Estimates from counties located in the Pacific region had higher uncertainty than other regions, driven by a higher proportion of underrepresented sociodemographic groups. Child-level overweight and obesity was related to race/ethnicity, sex, parental highest education ($P < 0.01$ for all), county-level walkability ($P = 0.03$), and urban/rural designation ($P = 0.02$). Overweight and obesity remains a vital issue for US youth, with substantial area-level variability. The additional uncertainty for underrepresented groups shows surveys need to better target diverse samples.

childhood obesity; childhood overweight; poststratification; small-area estimation; survey

Abbreviations: BMI, body mass index; CI, confidence interval; CV, cross-validated; NSCH, National Survey of Children's Health; RMSPE, root mean squared predictive error; SAE, small-area estimation; UIC, Urban Influence Codes.

Childhood overweight and obesity is a major public health problem with physical and mental health impacts, including reduced life expectancy (1–3), an association with adulthood obesity, and the onset of additional comorbidities and increased mortality (1, 4, 5). The United States has one of the highest rates of children experiencing overweight/obesity worldwide (6). Approximately 31.5% of US youth experience overweight (7), while 18.5% experience obesity, ranging from 14% among children aged 2–5 years to 21% among those aged 12–19 years (8). The most recent data available (9) indicates that obesity remains higher than the Healthy People 2020 target of 14.5% for children and adolescents aged 2–19 years (10). For the remainder of this article, we use "overweight" to indicate "at least overweight," which includes childhood overweight and obesity.

Disparities in childhood overweight exist across geographic regions and demographic subgroups (11, 12): female persons, members of historically marginalized groups, and persons with less educational experience higher overweight rates compared with their counterparts (13–20). Multiple studies have also identified the significant impact of children's environments on their weight status (21–24). Specifically, environmental factors such as climate, levels of green space, and safety have been associated with experiencing childhood overweight (18, 25). As a result, policy and community-level interventions to prevent childhood overweight have focused on improving environmental factors (26–28). Accessing estimates of childhood overweight across the United States is particularly important for surveillance and benchmarking efforts, policy and program development, and resource allocation toward communities with greatest needs (27, 29, 30). To understand spatial patterns of childhood overweight and to uncover the most influential determinants, current and accurate estimates of childhood overweight at the local level are necessary. However, the most recently available childhood overweight estimates are based on data from 2007 (9).

Our study aimed to bridge this gap by estimating childhood overweight rates at the county, state, and Census regional-division levels using the 2016 National Survey of Children's Health (NSCH) and the 2016 American Community Survey. Our objectives were to: 1) ensure that the current scope of overweight is well understood in a geographic context; 2) produce updated local-level predictive estimates that can help planning and evaluation of targeted community-level interventions against overweight; and 3) accurately characterize the uncertainty in overweight prediction estimates to identify where more data are needed. We implemented model-based small-area estimation (SAE) techniques, which are established methods to obtain prevalence predictions (31, 32) by extrapolating information from more data rich locations (9, 33–35). SAE has been successfully used to study obesity (15, 24, 36–43) and many other health behaviors and disease outcomes (43–47). Also, SAE models typically use population-based survey data linked to area-level covariates (e.g., county rurality, poverty, etc.) to model individual-level outcomes (9, 30, 34, 48) and poststratification to aggregate and weight individual-level estimates to an area level (46, 49). We demonstrate how multilevel modeling with poststratification can be used to accurately estimate uncertainty and identify areas with prevalence estimates different than the regional or national average while controlling the false discovery rate.

METHODS

Data sources

The 2016 NSCH, sponsored by the Health Resources and Services Administration, is a national survey examining the physical and emotional health of US children aged 0–17 years. This survey is conducted by the US Census Bureau in the 50 states and the District of Columbia via telephone (cellphone and landline), where information is gathered from parents/guardians of selected children. Within each state, an equal number of residences are randomly selected, and within each residence, a single child is randomly selected (50). For this study, only NSCH data for children aged 10–17 years were included because the body size of younger children changes more rapidly (51), and parent-reported height and weight measurements are more reliable for older children (52). We used the NSCH data for child-level variables.

For area-level variables (county, state, and division levels), we collected publicly available data from policy institutes and governmental organizations, including the Department of Health and Human Services, Department of Agriculture, Environmental Protection Agency, US Census Bureau, and American Community Survey among others (complete list in Web Table 1, available at https://doi.org/10.1093/aje/kwab176). Potential child- and area-level variables were chosen based upon previously established (9, 53) and hypothesized factors associated with childhood overweight and with input from obesity researchers at the University of South Carolina.

Child- and area-level model covariates

NSCH child-level variables tested included race/ethnicity (Hispanic, non-Hispanic White, Non-Hispanic Black, multiracial/other), sex (male, female), age, and educational attainment of the child's parent with the most education (less than high school, completed high school, some college, college or higher).

Area-level variables considered included Census-defined divisions (contiguous groups of states broken into New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific), state bullying laws, state Medicaid expansion status, state-level school wellness policy, federal/state/county school funding, county incarceration rates, county child insurance rates, county food insecurity rates, county rate of children living in households with a single parent, number of primary care/pediatric providers per 100,000 residents in the county, county adult obesity rate, county population-weighted walkability index (54), county school access and proximity status, county rate of access to exercise locations, county food desert status, and county urban-rural designation based upon 2013 Urban Influence Codes (UIC) (Web Table 1 provides a complete list). UIC codes were aggregated together to classify counties as either metropolitan (UIC codes 1 and 2; $n = 1{,}166$), micropolitan (UIC codes 3, 5, and 8; $n = 641$), or rural (UIC codes 4, 6, 7, 9–12; $n = 1{,}335$).

Primary outcome: childhood overweight/obesity

Children were classified as overweight/obese based on their body mass index (BMI, calculated as weight (kg)/height (m)^2) values. Using the 2000 Centers for Disease Control and Prevention sex- and age-specific growth charts (55), children with a BMI value of ≥95th percentile were classified as obese, and children with a BMI value between the 85th and 95th percentiles were considered overweight. Thus, a child with a BMI value of ≥85th percentile was considered to be overweight/obese. To calculate BMI for each child, NSCH relied on parental report of a child's height (feet/inches or meters/centimeters) and weight (pounds/ounces or kilograms/grams).

Statistical model

To predict the county-level prevalence of overweight, we used a multilevel mixed-effects logistic regression model of child-level overweight (for child j in county k, y_{jk} = yes or no) with a county-level intrinsic conditional autoregressive random intercept term (56, 57). Since $y_{jk} \sim \text{Bernoulli}(p_{jk})$, our model can then be written as:

$$\text{logit}\left(p_{jk}\right) = \mathbf{X}_{jk}\beta^{[1]} + \mathbf{C}_k\beta^{[2]} + b_k,$$

where $\mathbf{X}_{jk}$ represents a child-level vector of fixed-effects covariates (race/ethnicity, age, sex, parents' highest educational attainment) and $\mathbf{C}_k$ represents an area-level vector of fixed-effects covariates. The b_k intrinsic conditional autoregressive random effect incorporated the spatial dependence between a county and its neighboring counties, defined as all counties that share a border (contiguity of edges), and provided some spatial smoothing. Specifically, an intrinsic conditional autoregressive component implies that $b_k \mid b_{-k} \sim N(\bar{b}_k, \sigma^2/n_k)$, where b_{-k} represents the random intercepts of all counties excluding county k, $\bar{b}_k$ is the average random intercept for counties that neighbor county k, and n_k the number of neighbors of county k. Consistent with the predictive goal of the analysis, variable selection used 5-fold cross-validated root mean squared predictive error (CV RMSPE) to maximize the model's predictive ability while avoiding overfitting (58–60). Specifically, starting from a null model, we entered the variable that had the lowest CV RMSPE, checked if any variable already in the model could be dropped (i.e., dropping resulted in a lower CV RMSPE), and iterated until no variables improved the CV RMSPE. The intrinsic conditional autoregressive random effect decreased the CV RMSPE (compared with a standard random intercept). Further details are provided in Web Appendix 1. We also accounted for the NSCH survey design by leveraging provided sample survey weights which were reweighted to have a unit mean. To perform model validation, we compared our model-based state-level prevalence estimates with direct state-level estimates from the 2016 NSCH public use file (61).

Poststratification

The fitted regression model resulted in estimates of prevalence for all 256 strata of child-level variables (4 race/ethnicity r, 2 sex g, 8 age a, and 4 parental education e strata) denoted as $\hat{p}_{rgaek}$ for each county k. To create a single county-level estimate $\hat{p}_k$ that reflected the underlying population demographics, we conducted poststratification where we computed a weighted average of the strata overweight rates:

$$\hat{p}_k = \frac{\sum_r \sum_g \sum_a \sum_e \hat{p}_{rgaek} \mathrm{POP}_{rgaek}}{\sum_r \sum_g \sum_a \sum_e \mathrm{POP}_{rgaek}}.$$

The weights used in this average, POP_{rgaek}, are traditionally population counts; however, Census population estimates for the various strata of race/ethnicity, age, sex, and parent's education were not publicly available. To estimate these quantities, we used a multistep approach utilizing a combination of Census data, the publicly available NSCH data, and simplifying assumptions (description provided in Web Appendix 1).

Uncertainty and significance testing

To estimate the variability and perform hypothesis testing for the estimated overweight prevalence rates, we used a Monte Carlo–based parametric bootstrapping approach (62). Here, 10,000 bootstrap samples of estimated county-level overweight rates were computed using regression coefficients and random effects randomly drawn from estimated distributions. The 95% confidence intervals were given by the 2.5th and 97.5th percentiles of the bootstrap sample and incorporated uncertainty from both the regression coefficients and random effects. For the random effects, we used empirical Bayes predictions and standard errors assuming a normal distribution. For counties with no sampled individuals (n = 839), random effects were sampled conditional on data from neighboring counties.

We applied 2 different significance testing procedures to identify counties with outlying overweight rates. In the first procedure, P values from the county-level predicted empirical Bayes random effects were used. This approach identified counties with overweight rates significantly different from what was expected given their child- and area-level predictors (the fixed effects). In the second procedure, differences between the county and national rates were calculated for each bootstrap iteration. The proportion of times the difference was lower/higher than zero (indicating that the county rate was lower/higher than the national rate) multiplied by 2 was used as the P value. This approach identified counties that were significantly different from the national average. Both procedures used the Benjamini-Hochberg correction (63) to control the false discovery rate at a significance level of $\alpha = 0.05$.

Further statistical details and results are provided in Web Appendix 1. Our analyses/maps used SAS, version 9.4 (SAS Institute Inc., Cary, North Carolina) (64); R (R Foundation for Statistical Computing, Vienna, Austria) (65); and ArcGIS Pro 2.0 (ESRI, Redlands, California) (66). Study approval was given from the University of South Carolina Institutional Review Board (Protocol #00073928) and output approved through the US Census Bureau Disclosure Review Board.

RESULTS

Multilevel SAE model selection and validation

Table 1 shows the weighted and unweighted descriptive statistics of our sample. Out of 50,112 children in the 2016 NSCH data, 26,094 were aged 10–17 and of those, 24,162 (92.6%) had complete data; those excluded were due to missing BMI (6.5%) and missing other predictor data (<1%). In this sample, 53.9% were Non-Hispanic White (weighted proportion), 51.0% were male, and the average age was 13.5 years. Most children lived in the South Atlantic division (19.3%), Pacific division (16.2%), and East North Central division (14.7%). Our final model included the predictors that best explained the current scope of overweight based on CV RMSPE. These predictors were Census division, state wellness policy, 5 county-level variables (percentage of children in a single parent household in 2015, number of primary care providers per 100,000 residents, rurality, 2015 adult obesity rate, 2016 population-weighted walkability index), the child's age, main effects, and all 2-way interactions for child race/ethnicity, sex, and parent's education, along with a 3-way interaction of race/ethnicity by sex by parent's education (see Web Tables 2 and 3). Our model-based state-level overweight estimates performed well compared with state-level direct estimates from the 2016 NSCH public use file. The average difference

Table 1. Descriptive Statistics of Children Aged 10–17 Years, 2016 National Survey of Children's Health Data Set, United States

Predictor	Weighted No.	Weighted %[a]	Unweighted No.	Unweighted %[a]
Child's race/ethnicity				
Hispanic	7,047,399	23.8	2,521	10.4
Non-Hispanic White	15,974,168	53.9	17,408	72.0
Non-Hispanic Black	3,739,768	12.6	1,334	5.5
Multiple/other races	2,871,037	9.7	2,899	12.0
Child's sex				
Male	15,125,398	51.0	12,243	50.7
Female	14,506,974	49.0	11,919	49.3
Parents' highest educational attainment				
Less than high school	2,813,883	9.5	543	2.2
High school or GED	6,084,955	20.5	3,186	13.2
Some college	6,930,247	23.4	5,735	23.7
College degree or higher	13,803,286	46.6	14,698	60.8
Census division				
New England	1,296,098	4.4	3,203	13.3
Middle Atlantic	3,615,314	12.2	1,564	6.5
East North Central	4,346,146	14.7	2,666	11.0
West North Central	2,007,404	6.8	3,649	15.1
South Atlantic	5,724,112	19.3	3,970	16.4
East South Central	1,658,640	5.6	1,563	6.5
West South Central	3,897,042	13.2	1,476	6.1
Mountain	2,280,613	7.7	3,666	15.2
Pacific	4,807,002	16.2	2,405	10.0
Child age, years[b]	13.5 (2.3)		13.9 (2.3)	

Abbreviation: GED, General Educational Development.
[a] Values do not all sum to 1 due to rounding.
[b] Values are expressed as weighted and unweighted mean (standard deviation).

between the estimates was 1.55 (standard deviation, 1.25) percentage points (Pearson correlation $\rho = 0.89$; see Web Figure 1).

Predicted childhood overweight/obesity prevalence from multilevel SAE model

Table 2 shows the predicted national and regional division overweight rates, while Figure 1 shows a map of all predicted county-level overweight rates. The national overweight rate was 30.7% (95% confidence interval (CI): 27.0%, 34.9%, Table 2). The East South Central division had the highest rate of 36.6% (95% CI: 32.3%, 41.2%), while New England had the lowest with 27.8% (95% CI: 21.6%, 34.9%; Web Figure 2). More counties in the West South Central, South Atlantic, and East South Central divisions had higher estimated rates of childhood overweight (57.5% had rates ≥33%) compared with counties in the West North Central, Mountain and Pacific divisions (21.8% had rates ≥33%). After the false discovery rate correction, 31 counties had overweight rates significantly different from the national rate (23 lower, 8 higher). The Pacific division had the most counties that were significantly lower. We also found 14 counties with overweight rates significantly higher than expected given their child- and area-level fixed effects. Most of these counties were in California (4), Florida (4), and Texas (5). Web Table 4 gives a list of outlying counties, and Web Table 5 additional state-level results.

Our SAE model was also able to provide insights for targeted community-level interventions against overweight by giving predictive overweight estimates for groups of interest. Table 3 shows these predicted overweight rates by various demographic groups for each regional division. Nationally and by division, female children, non-Hispanic White children, children with a college-educated parent, and children aged 15 to 17 years had lower overweight rates. Children in metropolitan counties had lower overweight rates compared with children in micropolitan and rural counties. These results are displayed in Web Figures 3–7.

To characterize the uncertainty in our overweight predicted estimates, Figure 2 displays a bivariate map of overweight rate (low, medium, high) and the estimated 95% confidence

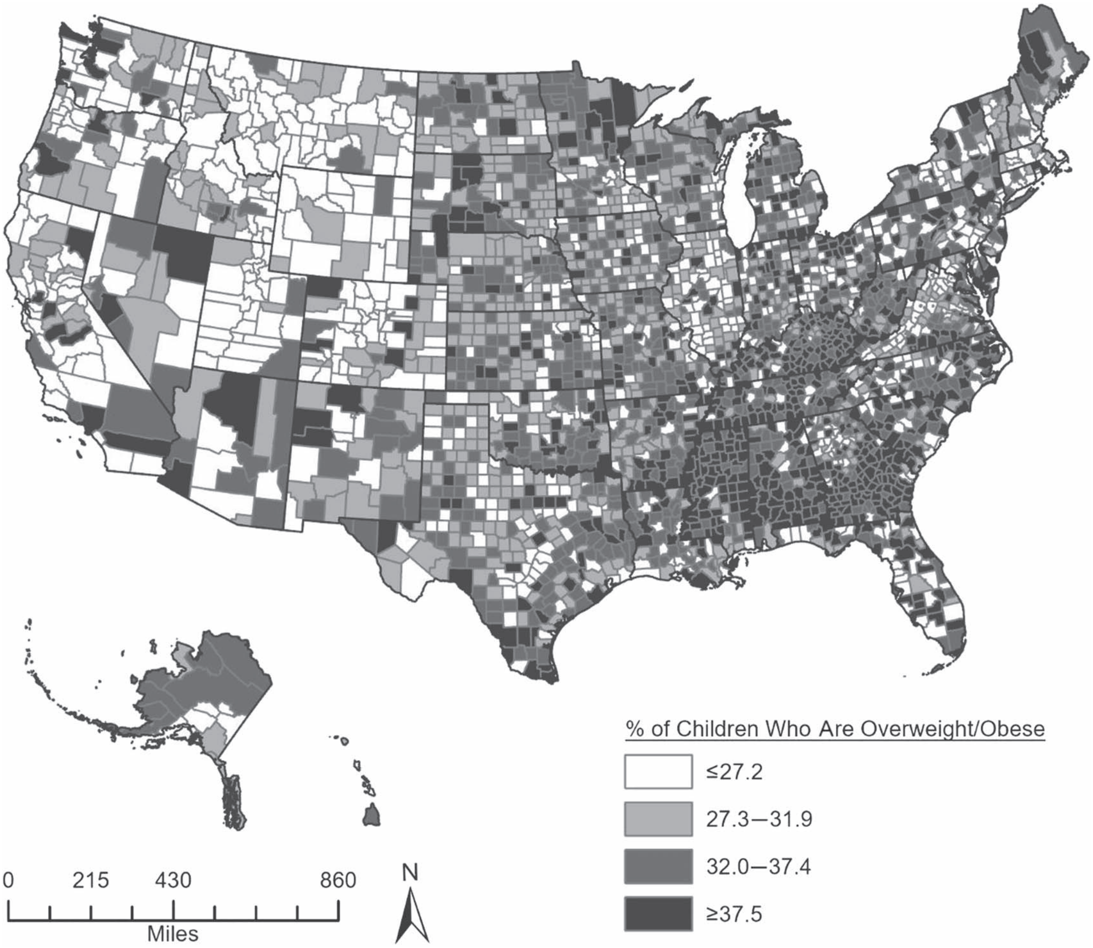

Figure 1. Estimated proportion of children aged 10–17 years experiencing overweight or obesity, county level, United States, 2016. Overweight/obesity estimates are indicated by color gradient: lightest color, 27% or lower; the second color, 28%–32%; the third color, 33%–37%; the darkest color, 38% or higher. The lighter the color of the county, the lower the overweight/obesity prevalence is. The darker the color of the county, the higher the overweight/obesity prevalence is. For example, counties in the Southeast United States have higher overweight/obesity prevalence than counties in New England. States included in the 9 Census divisions are New England: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont; Middle Atlantic: New Jersey, New York, and Pennsylvania; East North Central: Illinois, Indiana, Michigan, Ohio, and Wisconsin; West North Central: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota; South Atlantic: Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, and District of Columbia; East South Central: Alabama, Kentucky, Mississippi, and Tennessee; West South Central: Arkansas, Louisiana, Oklahoma, and Texas; Mountain: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming; and Pacific: Alaska, California, Hawaii, Oregon, and Washington. Full black lines on the map delineate state boundaries.

intervals (low, medium, high). "Low" or narrow confidence intervals indicated a more certain estimate. The higher child overweight rates in the South Atlantic division also had narrower confidence intervals, indicating a higher level of confidence in these estimates compared with those of the East South Central division, for instance. In the East South Central division, child overweight rates were high, but so were the confidence intervals of the estimated rates, indicating less certainty. An unexpected finding from our study is the large amount of uncertainty for Los Angeles County, California (overweight rate = 42.7%, 95% CI: 11.9%, 78.8%), given that it is a large metropolitan county. As an alternative example, Dallas County, Texas (overweight rate = 44.4%, 95% CI: 31.3%, 57.9%), had a similar sample size and overweight rate as Los Angeles but a much narrower confidence interval.

DISCUSSION

Childhood overweight/obesity prevalence from SAE- and inference-based models

Our study used SAE techniques to illustrate the widespread issue of childhood overweight and its geographic variation in the United States. To do so, we produced updated local-level predictive overweight estimates and characterized their

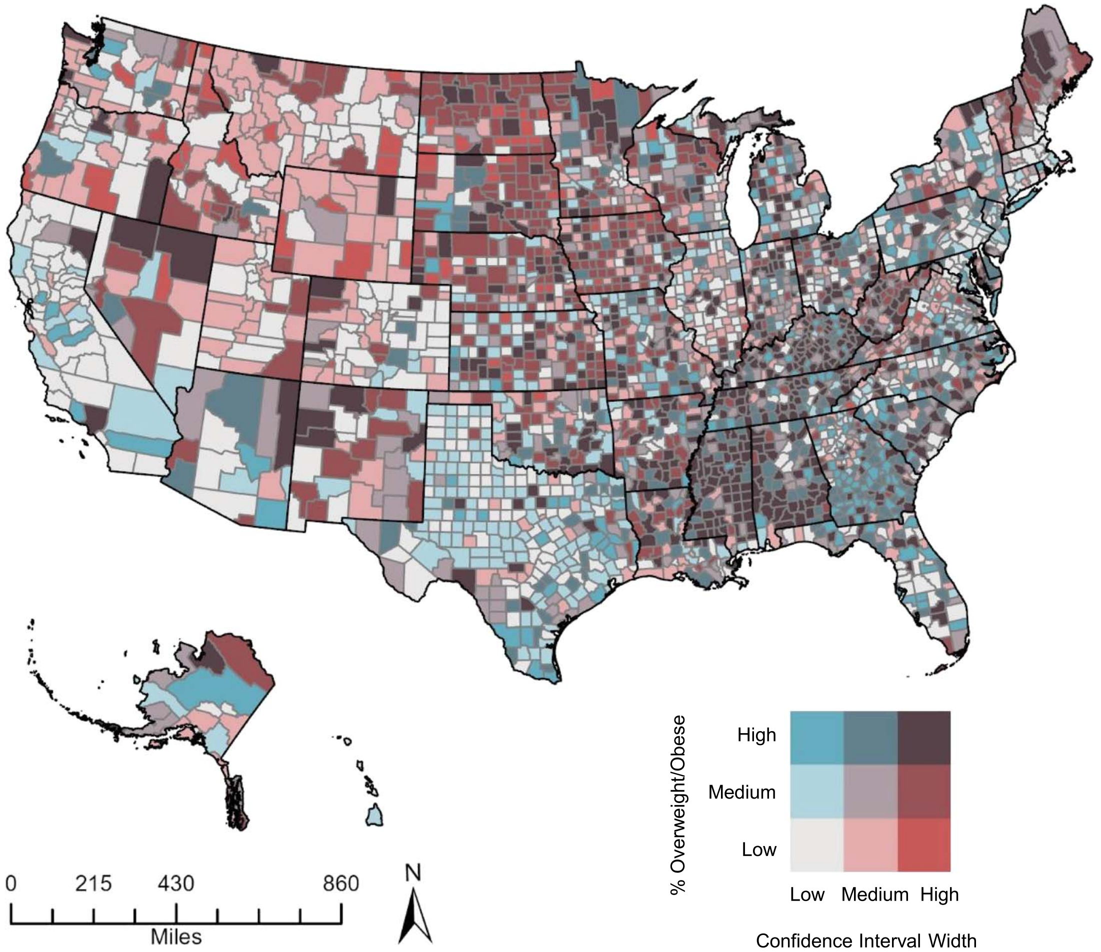

Figure 2. Bivariate map of estimated childhood overweight/obesity and confidence interval width, county level, United States, 2016. The lightest color represents counties with a low overweight/obesity estimate and narrow confidence intervals. The darkest color represents counties with a high overweight/obesity estimate and wide confidence intervals. Counties in shades of blue have a medium to high overweight/obesity estimate with narrow- to medium-width confidence intervals. Counties in shades of red have a low to medium overweight/obesity estimate with medium to wide confidence intervals. States included in the 9 Census divisions are New England: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont; Middle Atlantic: New Jersey, New York, and Pennsylvania; East North Central: Illinois, Indiana, Michigan, Ohio, and Wisconsin; West North Central: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota; South Atlantic: Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, and District of Columbia; East South Central: Alabama, Kentucky, Mississippi, and Tennessee; West South Central: Arkansas, Louisiana, Oklahoma, and Texas; Mountain: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming; and Pacific: Alaska, California, Hawaii, Oregon, and Washington. Full black lines on the map delineate state boundaries.

uncertainty per study objectives. Our national overweight rate of 30.7% is higher than the 16.8% obesity-only rate found by Zhang et al. using 2007 data (9). Our analysis predicted elevated rates of childhood overweight in the deep South, Four Corners, and Appalachian regions, in small pockets of the Northern Plains, and along the US-Mexico border. In prevalence studies using earlier NSCH data, patterns of elevated childhood obesity had been observed in the deep South, Appalachian region, San Joaquin Valley (California), and Four Corners region, as well as small pockets in Oklahoma, Northern Plains, and along the US-Mexico border (9, 67). Although those studies applied varied approaches to producing SAEs, there were some similarities between our maps displaying childhood overweight prevalence and those of previous studies examining only childhood obesity (9, 67). Trends in childhood overweight mirror trends in adult overweight (36).

Our study's predicted overweight prevalence by sociodemographic groups indicated that male sex, minority race/ethnicity, younger age, lower parental educational achievement, and micropolitan/rural residence were associated with higher childhood overweight rates. While our study was not designed to look at the inferential effects of sociodemographic characteristics, other studies have observed similar trends (9, 14, 67–70). For example, other studies confirmed higher prevalence rates of childhood overweight among male persons (67, 70), younger youth (9, 67), and those of minority race/ethnicity (9, 14, 67, 69, 70). Additionally, studies using National Health and Nutrition Examination Survey data also demonstrated an inverse association between the prevalence of childhood obesity and the educational level of the head of household (14, 69). Current literature also suggests that rural children have a greater risk of obesity compared with their urban counterparts (71), as demonstrated by a 2015 meta-analysis examining urban-rural differences in childhood obesity in a pooled sample of 74,168 participants aged 2–19 years old (72).

The increased uncertainty found by our study for specific county overweight estimates, such as Los Angeles, was driven by the model's fixed effects, in particular race/ethnicity and parental education (see Web Tables 6 and 7). The main effects for Hispanic and non-Hispanic Black respondents had over 50% more variability than the main effect for non-Hispanic White respondents, and the main effects for lower parental education attainment had over 40% more variance than for higher educational levels. The demographics of Dallas County had a higher concentration of sociodemographic groups with low uncertainty compared with Los Angeles County, thus resulting in the marked difference in uncertainty.

Table 2. Overweight/Obesity Rates for Children Aged 10–17 Years, Overall and in Census Divisions, Using Small-Area Estimation Model Predictions, Based on the 2016 National Survey of Children's Health Data Set, United States

Regional Division	Overweight Rate	Overweight Rate 95% CI	No. of Counties	Interquartile Range	No. of Counties Below National Rate	No. of Counties Above National Rate	No. of Counties Below Given Predictors	No. of Counties Above Given Predictors
United States	30.7	27.0, 34.9	3,143	2.9	23	8	0	14
New England	27.8	21.6, 34.9	67	4.5	0	0	0	0
Mid Atlantic	31.9	27.3, 36.9	150	3.3	1	0	0	1
East North Central	30.7	26.0, 36.2	437	3.7	4	0	0	0
West North Central	29.7	26.1, 33.6	618	2.5	0	0	0	0
South Atlantic	31.8	28.3, 35.6	589	2.5	2	2	0	4
East South Central	36.6	32.3, 41.2	364	3.1	0	0	0	0
West South Central	33.2	28.4, 38.3	470	3.4	5	5	0	5
Mountain	25.2	20.2, 31.2	281	3.9	1	0	0	0
Pacific	28.6	19.6, 39.7	167	7.6	10	1	0	4

Abbreviation: CI, confidence interval.

Implications for research and practice

Our study has important implications for future research. First, this research drew upon publicly available data and utilized an SAE and poststratification approach reproducible for other important health outcomes (details in Web Appendix 1). Second, future research may use our predictions for inferential work highlighting the relationships between food environments (i.e., presence of healthy and unhealthy food outlets), sociodemographic composition, and childhood overweight at the local level. Finally, additional validation work that compares model-based estimates with prevalence values obtained through Census-based approaches where nearly 100% of the population/area has data available (e.g., aggregation of school records) is an important next step.

This research also has implications for future public heath practice. Our small area estimates of childhood overweight and associated visualizations allow for more precise identification of populations or areas in need of resources and/or programs to combat childhood overweight. Our study showed that some areas of the United States have markedly higher childhood overweight rates and/or wide confidence intervals. This information can be used to understand which geographic areas or sociodemographic groups need additional programming, resources, or policy interventions (i.e., male adolescents, rural residents, etc.) and which geographic areas need better data collection (e.g., oversampling minority groups) to improve the certainty of estimates. Having publicly available local-level estimates of overweight is particularly important for organizations creating strategic plans, grants, toolkits, and other resources to highlight the areas/groups likely to benefit from these additional resources. Knowing which geographic areas or populations have less prediction uncertainty is also important, so that survey planners can modify their sampling

Table 3. Childhood Overweight/Obesity Rates (%) for Children Aged 10–17 Years, According to Demographic Groups and Census Divisions, Using Small-Area Estimation Model Predictions, Based on the 2016 National Survey of Children's Health Data Set, United States

Group	United States		New England		Middle Atlantic		East North Central		West North Central	
	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI
Overall	30.7	27.0, 34.9	27.8	21.6, 34.9	31.9	27.3, 36.9	30.7	26.0, 36.2	29.7	26.1, 33.6
Child's sex										
Male	33.0	29.0, 37.4	30.5	24.1, 38.0	34.3	29.5, 39.5	33.1	28.1, 38.7	32.5	28.5, 36.6
Female	28.3	24.7, 32.5	24.9	19.3, 31.6	29.1	24.8, 34.0	28.3	23.8, 33.6	26.9	23.4, 30.5
Child's race/ethnicity										
Hispanic	39.5	33.0, 46.1	38.6	30.3, 47.4	41.5	35.4, 47.8	40.8	31.7, 50.6	40.9	36.1, 45.6
Non-Hispanic White	26.5	23.9, 29.7	24.5	19.1, 31.2	28.1	24.0, 32.5	27.6	24.1, 31.8	27.8	24.3, 31.5
Non-Hispanic Black	39.2	34.7, 43.7	38.6	30.1, 48.0	37.6	31.4, 44.2	39.7	31.5, 48.5	36.3	30.3, 42.6
Multiple or other races	28.6	23.5, 34.9	28.6	21.5, 36.4	30.4	25.0, 36.4	30.4	24.1, 38.0	30.1	25.8, 34.6
Parents' highest educational attainment										
Less than high school	42.1	36.1, 48.0	42.2	33.7, 51.0	41.5	35.3, 47.7	41.3	34.2, 48.5	42.6	37.7, 47.8
High school or GED	38.8	34.6, 43.3	36.7	29.5, 44.7	40.1	34.8, 45.8	38.2	32.9, 43.8	37.4	33.2, 41.7
Some college	36.1	32.0, 40.3	34.5	27.3, 42.5	38.5	33.2, 44.0	36.7	31.6, 42.2	36.1	31.7, 40.4
College degree or higher	27.2	23.7, 31.3	25.2	19.4, 32.0	28.3	23.9, 33.0	27.8	23.4, 33.2	27.1	23.5, 30.7
Age, years										
10–14	32.9	29.0, 37.2	30.1	23.8, 37.5	34.1	29.3, 39.3	33.0	28.1, 38.5	32.0	28.1, 36.0
15–17	27.1	23.6, 31.3	24.3	18.6, 30.7	28.1	23.8, 32.8	27.1	22.6, 32.3	26.0	22.6, 29.7
Metropolitan Counties	30.0	25.9, 34.8	27.6	21.2, 34.9	31.7	27.0, 36.9	29.9	24.6, 36.2	27.9	23.9, 32.3
Micropolitan Counties	35.6	32.4, 38.9	28.8	20.6, 38.4	34.1	27.9, 40.9	35.5	30.7, 40.4	34.1	29.3, 39.5
Rural Counties	34.5	30.9, 38.4	30.8	21.8, 40.9	32.9	25.7, 40.9	32.6	27.5, 38.2	33.3	28.7, 38.2

Table continues

Table 3. Continued

Group	South Atlantic		East South Central		West South Central		Mountain		Pacific	
	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI
Overall	31.8	28.3, 35.6	36.6	32.3, 41.2	33.2	28.4, 38.3	25.2	20.2, 31.2	28.6	19.6, 39.7
Child's sex										
Male	33.5	29.8, 37.5	38.6	34.0, 43.4	35.5	30.6, 40.8	27.6	22.3, 33.7	31.2	21.4, 42.5
Female	30.1	26.7, 33.8	34.5	30.3, 39.1	30.7	26.4, 35.7	22.6	18.0, 28.1	25.9	17.4, 37.0
Child's race/ethnicity										
Hispanic	40.2	34.4, 46.6	46.9	41.4, 52.4	42.4	36.4, 48.8	33.7	27.3, 41.2	37.8	25.3, 50.6
Non-Hispanic White	27.0	23.8, 30.3	33.2	29.0, 37.7	27.5	23.5, 32.0	21.0	16.8, 26.1	22.8	16.2, 32.0
Non-Hispanic Black	39.3	35.3, 43.6	44.5	39.0, 50.4	39.2	33.5, 45.0	32.7	24.9, 41.7	35.1	22.5, 49.2
Multiple or other races	29.7	26.0, 33.8	35.7	30.5, 41.2	30.6	25.5, 36.4	25.2	19.6, 31.6	26.5	17.3, 38.6
Parents' highest educational attainment										
Less than high school	41.2	36.6, 46.2	45.0	39.1, 50.9	45.9	39.7, 51.9	38.6	31.4, 46.8	41.6	28.6, 53.8
High school or GED	40.1	36.0, 44.4	43.1	38.3, 48.4	40.2	35.2, 45.6	33.1	26.7, 39.8	36.9	26.5, 47.6
Some college	38.2	34.4, 42.2	41.7	37.0, 47.0	37.0	32.1, 42.5	29.5	23.7, 36.0	33.0	23.4, 43.6
College degree or higher	28.6	25.3, 32.3	32.9	28.7, 37.5	28.9	24.5, 34.0	21.8	17.4, 27.3	24.6	16.3, 35.8
Age, years										
10–14	34.0	30.4, 37.9	39.0	34.6, 43.9	35.4	30.6, 40.7	27.1	21.9, 33.2	30.7	21.1, 41.9
15–17	28.2	24.8, 31.9	32.5	28.4, 37.1	29.3	24.9, 34.3	21.7	17.2, 27.1	25.3	16.7, 36.6
Metropolitan Counties	30.8	27.3, 34.8	33.9	29.3, 38.9	32.9	27.8, 38.5	24.4	19.0, 30.9	28.6	19.2, 40.1
Micropolitan Counties	41.5	36.1, 46.9	42.3	36.3, 48.5	35.8	30.5, 41.3	30.1	24.2, 36.4	28.7	22.8, 35.6
Rural Counties	37.7	32.3, 43.4	42.4	36.2, 48.8	33.1	28.0, 38.9	27.5	21.8, 34.0	27.6	21.1, 35.1

Abbreviations: CI, confidence interval; GED, General Educational Development.

plans accordingly (e.g., reducing target sample size for some groups/areas) and potentially reduce cost.

Strengths and limitations

This study had many strengths that emphasize its contribution to the field. Robust statistical techniques were used to provide estimates of childhood overweight at the county and state levels that were previously unavailable. For counties with little data, SAE allowed for reliable estimates to be produced by conditioning on data from neighboring counties. Additionally, our poststratification method accounted for each area's underlying population when producing final estimates. Moreover, estimates were presented using choropleth and bivariate maps, offering rich visualization of the scope of the childhood overweight epidemic in the United States, as well as understanding geographic uncertainty in the data. Among 10–17-year-olds in the NSCH, the major reason for exclusion was missing BMI (6.5%). Those with and without BMI data did not differ in terms of our other covariate data (data not shown). Our study findings add to a growing body of knowledge on the prevalence and determinants of childhood overweight across the United States.

An important limitation was that parent-reported height and weight were used to calculate BMI for classification of overweight and were not independently verified. BMI derived from parent-reported height and weight was found in a meta-analysis to overestimate (prevalence ratio = 1.12) the prevalence of overweight compared with using BMI derived from direct height and weight measurements (73), yet parent-reported height and weight were more accurate for estimating overweight and obesity combined than for obesity alone (73). Additionally, there is evidence demonstrating that BMI derived from parent-reported height and weight increases in accuracy with the child's age (73, 74).

Another limitation was that different regional divisions showed overweight rates with wide confidence intervals when the division included counties with a higher prevalence of underrepresented demographic groups (e.g., overweight rates in the Pacific division have wider confidence intervals than other divisions). This had a large impact for counties where the prevalence of rare population subgroups was relatively high but had little impact for most counties, where the prevalence of these subgroups was very low. This is indicative of a need to understand and survey demographic subgroups more effectively, to obtain more precise estimates for all demographic groups. Despite of the availability of survey weights and good sample sizes, these results point to the importance of improving the survey sampling and administration process.

Conclusion

Model-based SAE was performed to update prevalence estimates of overweight among children aged 10–17 across the United States. Overweight predictions differed by geographic area as well as child sociodemographic groups, and areas with low precision were observed (i.e., greater uncertainty in regional division/counties with a high proportion of underrepresented population subgroups). This study highlights the importance of broadening survey sampling considerations to include representation of diverse population subgroups and geographic regions.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, United States (Anja Zgodic, Jan M. Eberth, Angela D. Liese, Alexander C. McLain); Rural and Minority Health Research Center, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, United States (Jan M. Eberth, Charity B. Breneman); Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, United States (Marilyn E. Wende, Andrew T. Kaczynski); and Prevention Research Center, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, United States (Andrew T. Kaczynski).

This study was supported by the Federal Office of Rural Health Policy (FORHP), Health Resources and Services Administration (HRSA), US Department of Health and Human Services (HHS) (cooperative agreement U1CRH30539), and by the National Institutes of Health (NIH), National Institute of General Medical Sciences (NIGMS), through the Research Center for Child Well-Being (NIGMS, grant P20GM130420).

This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2086. All results have been reviewed to ensure that no confidential information is disclosed.

The information, conclusions, and opinions expressed in this brief are those of the authors, and no endorsement by the US Census Bureau, FORHP, HRSA, HHS, NIH, or NIGMS is intended or should be inferred.

Conflict of interest: none declared.

REFERENCES

1. Ebbeling CB, Pawlak DB, Ludwig DS. Childhood obesity: public-health crisis, common sense cure. *Lancet*. 2002; 360(9331):473–482.
2. US Preventive Services Task Force, Grossman DC, Bibbins-Domingo K, et al. Screening for obesity in children and adolescents us preventive services task force recommendation statement. *JAMA*. 2017;317(23):2417–2426.
3. Halfon N, Larson K, Slusser W. Associations between obesity and comorbid mental health, developmental, and physical health conditions in a nationally representative sample of us children aged 10 to 17. *Acad Pediatr*. 2013;13(1):6–13.
4. Serdula MK, Ivery D, Coates RJ, et al. Do obese children become obese adults? A review of the literature. *Prev Med (Baltim)*. 1993;22(2):167–177.

5. Barlow SE. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics*. 2007;120(suppl 4):S164–S192.
6. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384(9945): 766–781.
7. Hedley A, Ogden C, Johnson C, et al. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *JAMA*. 2004;291(23):2847–2850.
8. Hales CM, Carroll MD, Fryar CD, et al. Prevalence of obesity among adults and youth: United States, 2015–2016. *NCHS Data Brief*. 2017. https://www.cdc.gov/nchs/products/databriefs/db288.htm. Accessed August 25, 2021.
9. Zhang X, Onufrak S, Holt JB, et al. A multilevel approach to estimating small area childhood obesity prevalence at the Census block-group level. *Prev Chronic Dis*. 2013;10(5):1–10.
10. Office of Disease Prevention and Health Promotion. Nutrition and weight status. 2020. https://www.healthypeople.gov/2020/topics-objectives/topic/nutrition-and-weight-status/. Accessed February 9, 2020.
11. Wang YC, McPherson K, Marsh T, et al. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet*. 2011;378(9793):815–825.
12. Wang Y, Beydoun MA, Liang L, et al. Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. *Obesity*. 2008;16(10): 2323–2330.
13. Gartner DR, Taber DR, Hirsch JA, et al. The spatial distribution of gender differences in obesity prevalence differs from overall obesity prevalence among US adults. *Ann Epidemiol*. 2016;26(4):293–298.
14. Ogden CL, Carroll MD, Fakhouri TH, et al. Prevalence of obesity among youths by household income and education level of head of household—United States 2011–2014. *Morb Mortal Wkly Rep*. 2018;67(6):186–189.
15. Tabano DC, Bol K, Newcomer SR, et al. The spatial distribution of adult obesity prevalence in Denver County, Colorado: an empirical Bayes approach to adjust EHR-derived small area estimates. *EGEMS (Wash DC)*. 2017; 5(1):24.
16. Jun HJ, Namgung M. Gender difference and spatial heterogeneity in local obesity. *Int J Environ Res Public Health*. 2018;15(2):311.
17. Penney TL, Rainham DGC, Dummer TJB, et al. A spatial analysis of community level overweight and obesity. *J Hum Nutr Diet*. 2014;27(suppl 2):65–74.
18. Hales CM, Fryar CD, Carroll MD, et al. Differences in obesity prevalence by demographic characteristics and urbanization level among adults in the United States, 2013–2016. *JAMA*. 2018;319(23):2419–2429.
19. Arroyo-Johnson C, Mincey KD. Obesity epidemiology trends by race/ethnicity, gender, and education: National Health Interview Survey, 1997–2012. *Gastroenterol Clin North Am*. 2016;45(4):571–579.
20. Jia P, Cheng X, Xue H, et al. Applications of geographic information systems (GIS) data and methods in obesity-related research. *Obes Rev*. 2017;18(4):400–411.
21. Huang R, Moudon AV, Cook AJ, et al. The spatial clustering of obesity: does the built environment matter? *J Hum Nutr Diet*. 2015;28(6):604–612.
22. Loprinzi PD, Addoh O, Ortega Hinojosa AM, et al. Proximity effects in obesity rates in the US: a spatial Markov chains approach. *Spat Spatiotemporal Epidemiol*. 2017;11(1):1–12.
23. van der Zwaard B, Schalkwijk AAH, Elders PJM, et al. Does environment influence childhood BMI? A longitudinal analysis of children aged 3–11. *J Epidemiol Community Health*. 2018;72(12):1110–1116.
24. Fang D, Thomsen MR, Nayga RM Jr, et al. Association of Neighborhood Geographic Spatial Factors with rates of childhood obesity. *JAMA Netw Open*. 2018;1(4):e180954.
25. Von Hippel P, Benson R. Obesity and the natural environment across US counties. *Am J Public Health*. 2014; 104(7):1287–1293.
26. Vine M, Hargreaves MB, Briefel RR, et al. Expanding the role of primary care in the prevention and treatment of childhood obesity: a review of clinic- and community-based recommendations and interventions. *J Obes*. 2013; 2013:172035.
27. O'Connor EA, Evans CV, Burda BU, et al. Screening for obesity and intervention for weight management in children and adolescents evidence report and systematic review for the us preventive services task force. *JAMA*. 2017;317(23): 2427–2444.
28. Messiah SE, Jiang S, Kardys J, et al. Reducing childhood obesity through coordinated care: development of a park prescription program. *World J Clin Pediatr*. 2016;5(3): 234–243.
29. Stafford M, Duke-Williams O, Shelton N. Small area inequalities in health: are we underestimating them? *Soc Sci Med*. 2008;67(6):891–899.
30. D'Agostino-McGowan L, Gennarelli RL, Lyons SA, et al. Using small-area analysis to estimate county-level racial disparities in obesity demonstrating the necessity of targeted interventions. *Int J Environ Res Public Health*. 2013;11(1): 418–428.
31. Battese GE, Harter RM, Fuller WA. An error-components model for prediction of county crop areas using survey and satellite data. *J Am Stat Assoc*. 1988;83(401):28–36.
32. Ghosh M, Rao J. Small area estimation: an appraisal. *Stat Sci*. 1994;9(1):55–93.
33. Myers CA, Slack T, Martin CK, et al. Regional disparities in obesity prevalence in the United States: a spatial regime analysis. *Obesity*. 2015;23(2):481–487.
34. Zhang X, Holt JB, Yun S, et al. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the Behavioral Risk Factor Surveillance System. *Am J Epidemiol*. 2015;182(2): 127–137.
35. Panczak R, Held L, Moser A, et al. Finding big shots: small-area mapping and spatial modelling of obesity among Swiss male conscripts. *BMC Obes*. 2016;3(1):10.
36. Koh K, Grady S, Darden J, et al. Adult obesity prevalence at the county level in the United States, 2000–2010: downscaling public health survey data using a spatial microsimulation approach. *Spat Spatiotemporal Epidemiol*. 2018;26:153–164.
37. Li X, Staudt A, Chien LC. Identifying counties vulnerable to diabetes from obesity prevalence in the United States: a spatiotemporal analysis. *Geospat Health*. 2016;11(3):299–306.
38. Benjamin Neelon SE, Burgoine T, Gallis JA, et al. Spatial analysis of food insecurity and obesity by area-level deprivation in children in early years settings in England. *Spat. Spatiotemporas. Epidemiol*. 2017;23:1–9.
39. Agovino M, Crociata A, Sacco PL. Proximity effects in obesity rates in the US: a spatial Markov chains approach. *Soc Sci Med*. 2019;220(May 2018):301–311.
40. Cataife G. Small area estimation of obesity prevalence and dietary patterns: a model applied to Rio de Janeiro city, Brazil. *Heal Place*. 2014;26:47–52.

41. Klompas M, Cocoros NM, Menchaca JT, et al. State and local chronic disease surveillance using electronic health record systems. *Am J Public Health*. 2017;107(9):1406–1412.
42. Zhang Z, Zhang L, Penman A, et al. Using small-area estimation method to calculate county-level prevalence of obesity in Mississippi, 2007–2009. *Prev Chronic Dis*. 2011; 8(4):A85.
43. Mills CW, Johnson G, Huang TTK, et al. Use of small-area estimates to describe county-level geographic variation in prevalence of extreme obesity among US adults. *JAMA Netw Open*. 2020;3(5):e204289.
44. Eberth JM, Hossain MM, Tiro JA, et al. Human papillomavirus vaccine coverage among females aged 11 to 17 in Texas counties: an application of multilevel, small area estimation. *Women's Heal Issues*. 2013;23(2):1–25.
45. Khan S, McLain A, Olatosi B, et al. Small area estimation of county-level U.S. HIV-prevalent cases. *Ann Epidemiol*. 2020; 48:30–35.e9.
46. Zhang X, Holt JB, Lu H, et al. Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *Am J Epidemiol*. 2014;179(8): 1025–1033.
47. Lin Y, McLain A, Probst J, et al. Health-related quality of life among adults 65 years and older in the United States, 2011–2012: a multilevel small area estimation approach. *Ann Epidemiol*. 2017;27(1):52–58.
48. Ortega Hinojosa AM, Davies MM, Jarjour S, et al. Developing small-area predictions for smoking and obesity prevalence in the United States for use in environmental public health tracking. *Environ Res*. 2014;134:435–452.
49. Hindmarsh D, Steel D. Creating local estimates from a population health survey: practical application of small area estimation methods. *AIMS Public Health*. 2020;7(2): 403–424.
50. US Census Bureau. 2016 Methodology Report. National Survey of Children's Health. 2016. https://www.census.gov/content/dam/Census/programs-surveys/nsch/tech-documentation/methodology/2016-NSCH-Methodology-Report.pdf. Accessed February 9, 2020.
51. Himes JH. Challenges of accurately measuring and using BMI and other indicators of obesity in children. *Pediatrics*. 2009;124(suppl 1):3–22.
52. Akinbami L, Ogden C. Childhood overweight prevalence in the United States: the impact of parent-reported height and weight. *Obesity*. 2009;17(8):1574–1580.
53. Li W, Kelsey JL, Zhang Z, et al. Small-area estimation and prioritizing communities for obesity control in Massachusetts. *Am J Public Health*. 2009;99(3):511–519.
54. US Environmental Protection Agency. National Walkability Index. 2010. https://www.epa.gov/smartgrowth/smart-location-mapping#walkability. Accessed September 28, 2020.
55. Kuczmarski R, Ogden C, Guo S. 2000 CDC growth charts for the United States: methods and development. National Center for Health Statistics. *Vital Health Stat 11*. 2002;11(246): 1–190.
56. Cressie N. *Statistics for Spatial Data. Revised*. New York, NY: Wiley; 1993.
57. Banerjee S, Carlin B, Gelfand A. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall; 2004.
58. Stone M. Cross-validation choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B*. 1974;36(2):111–147.
59. Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J R Stat Soc Ser B*. 1977;39(1):44–47.
60. Allen DM. The relationship between variable selection and data augmentation and a method for prediction. *Dent Tech*. 1974;16(1):125–127.
61. Goodman MS. Comparison of small-area analysis techniques for estimating prevalence by race. *Prev Chronic Dis*. 2010; 7(2):A33.
62. Tofighi D, MacKinnon DP. Monte Carlo confidence intervals for complex functions of indirect effects. *Struct Equ Model A Multidiscip J*. 2016;23(2):194–205.
63. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57(1):289–300.
64. SAS Institute Inc. SAS Software, *Release 9*.4. Cary, NC: SAS Institute Inc; 2013.
65. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
66. ESRI. *ArcGIS Desktop, release 10*. Redlands, CA: ESRI; 2011.
67. Kramer M, Raskind I, van Dyke M, et al. Geography of adolescent obesity in the U.S., 2007–2011. *Am J Prev Med*. 2016;51(6):898–909.
68. Tarasenko Y, Chen C, Smalley K, et al. Rural-urban differences in perceptions of child overweight among children and adolescents, their guardians and health care professionals in the United States. *J Rural Heal*. 2016;32(1): 63–71.
69. Ogden C, Fryar C, Hales C, et al. Differences in obesity prevalence by demographics and urbanization in US children and adolescents, 2013–2016. *JAMA*. 2018;319(23): 2410–2418.
70. Zheng S, Holt N, Southerlund J, et al. Prevalence of and risk factors for adolescent obesity in Tennessee using the 2010 Youth Risk Behavior Survey (YRBS) data: an analysis using weighted hierarchical logistic regression. *Biom Biostat Int J*. 2016;4(6):234–242.
71. Kaczynski A, Eberth J, Stowe E, et al. Development of a national childhood obesogenic environment index in the United States: differences by region and rurality. *Int J Behav Nutr Phys Act*. 2020;17(1):83.
72. Johnson JA 3rd, Johnson A. Urban-rural differences in childhood and adolescent obesity in the United States: a systematic review and meta-analysis. *Child Obes*. 2015; 11(3):233–241.
73. He J, Fan X. How accurate is using parent-reported height and weight for screening children and adolescents for overweight and obesity? Meta-analyses at both population and individual levels. *Child Obes*. 2018;14(5):302–315.
74. Weden M, Brownell P, Rendall M, et al. Parent-reported height and weight as sources of bias in survey estimates of childhood obesity. *Am J Epidemiol*. 2013;178(3): 461–473.