Towards Estimating Childhood Obesity Prevalence Using Electronic Health Records

Timothy D. McFarlane

Epidemiology, Indiana University Richard M Fairbanks School of Public Health, Indianapolis, Indiana, United States

Objective

To discuss the use of electronic health records (EHRs) for estimation of overweight and obesity prevalence in children aged 2 to 19 years and to compare prevalence between the convenience sample obtained from EHRs to prevalence adjusted for potential selection bias.

Introduction

Although recent data suggests childhood obesity prevalence has stabilized, an estimated 1 in 3 U.S. children are overweight or obese [1]. Further, there is variation by racial and ethnic groups, location, age, and poverty [2], resulting in a need for local data to support public health planning and evaluation efforts.

Current methods for surveillance of childhood weight status rely on self-report from community-based surveys. However, surveys have long time intervals between data collection periods, are expensive, and are not often able to produce precise small-area estimates. EHRs have been increasingly proposed as an alternative or supplement to community surveys. Childhood weight and height is collected as a part of routine care, and leveraging these data from EHRs may provide rapid and locally precise estimates of childhood weight status. A concern for the use of EHRs is the potential for selection bias. EHRs represent only those seeking healthcare and may not generalize to the population. Additionally, the type of clinical visit (e.g., wellness vs. acute) may affect the prevalence estimates and the likelihood of collecting height and weight data in the EHR. Thus, in addition to EHRs being a convenience sample, there may be additional selection biases based on the type of visit and whether height and weight was measured and recorded. The current work sought to quantify the effect of visit type on childhood overweight and obesity prevalence and generate weights to adjust prevalence for potential EHR-related selection bias.

Methods

Two years (2014-2015) of EHR data were obtained from the Indiana Network for Patient Care, a community health information exchange. Data included clinical encounters of patients living in the eight-county metropolitan area of Indianapolis, Indiana. BMI was calculated using recorded height and weight from the most recent encounter. Encounters were screened for valid BMI entries by examining records in the $0-5^{th}$ and $95-100^{th}$ percentiles. BMI results were validated using the following procedure: censoring records with one encounter; removing encounters with implausible values ($5 < BMI < 100$); calculating the mean BMI across remaining encounters; calculating the percent difference from the mean BMI for each encounter; and removing encounters with BMI results greater or less than 10% from the mean BMI. Records which could not be validated were censored and treated as missing height and weight. Using the age- and sex- specific Centers for Disease Control and Prevention growth charts, patients were classified as underweight ($0-5^{th}$ percentiles), normal weight ($5-85^{th}$ percentiles), overweight ($85-95^{th}$ percentiles), and obese ($>95^{th}$ percentile).

Wellness visits were identified using the following ICD-9-CM or ICD-10-CM diagnosis codes: V20.2, V70.0, V70.9; and Z00.121, Z00.129, Z00.00, Z00.01. To adjust for potential selection bias, two stabilized inverse probability weights (SIPW) were constructed. First, to account for potential selection bias induced by visit type and, second, to account for potential selection bias due to censoring (i.e., missing height and weight data). The SIPW were generated using logistic regression models to calculate the predicted probabilities for visit type and uncensored observations as a function of the covariates race, ethnicity, age, gender, and insurance. The SIPW weights were specified as depicted below, where W=1 is a wellness visit, L=observed covariates, and C=0 is uncensored for each child, i. [Figure 1]

The final weight (SW_{Final}) was applied to the sample to create a pseudo-population in which there is no association between covariates, L and visit type and which has the same distribution of covariates, L, as the censored individuals not included in the

pseudo-population, thus making censoring occur at random, given the observed covariates. Under the assumption of exchangeability and no unmeasured or residual confounding, the pseudo-population will no longer have selection bias due to differences in visit type and missing data.

Results

The sample consisted of 130,626 unique individuals between the ages of 2 and 19 years, of which 92,755 (71%) had at least one recorded height and weight result. Of the 10,184 records screened for BMI results, 5,242 (51%) were validated using measurements from previous encounters. The final sample consisted of 87,804 records with a valid BMI result (67%) and 42,822 records censored due to missing data (33%). Compared to the U.S. Census, the EHR sample over- represented older girls (e.g., 31.2% vs. 41.2% 15-19 year-old girls) and under-represented younger girls (e.g., 34.3% vs. 29.5% for 5-9 year-old girls). Wellness visits were associated with censoring due to missing data; only 3% of censored encounters were wellness visits compared to 33% of uncensored encounters [$P(\chi^2_1 > 14437 = < 0.0001)$].

In the unweighted sample, the overall prevalence of overweight or obesity was 36.5%. The overweight or obesity prevalence was lower among wellness visits2 (33.9%) than other visits (37.8%; $P(\chi_1 > 124.2 =< 0.0001$). Similarly, wellness visits had lower prevalence estimates when stratified by sex, race, age, ethnicity, and insurance (Table 1). After weighting the sample by SW_{Final}, the overall prevalence of overweight or obesity was 36.2% and the difference between wellness (35.1%) and other visits (36.7%) was attenuated, though statistically significant [$P(\chi^2 > 22.2 =< 0.001$). Likewise, the differences between wellness and other visits in the weighted pseudo-population were attenuated when stratified by covariates, compared to unweighted analyses (Table 1). While the SIPW method demonstrated some adjustment for selection bias due to visit type and censoring due to missing data, the adjustment was incomplete, likely as a result of unmeasured and imperfectly measured covariates.

Conclusions

Wellness visits were associated with lower childhood overweight and obesity prevalence and were more likely to have weight and height measurements recorded in the EHR than other visit types. Adjusting prevalence for EHR-related selection bias using stabilized inverse probability weights may produce more valid estimates but the lack of social determinant data in EHRs results in imperfect adjustment. Future work should integrate individual- or community-level social determinants of health data into the weighting models.

References

1. Skinner AC, Skelton JA. 2014. Prevalence and trends in obesity and severe obesity among children in the United States, 1999-2012. *JAMA Pediatr.* 168(6), 561-66. PubMed https://doi.org/10.1001/jamapediatrics.2014.21

2. Ogden CL, et al. 2018. Differences in Obesity Prevalence by Demographics and Urbanization in US Children and Adolescents, 2013-2016. *JAMA*. 319(23), 2410-18. PubMed https://doi.org/10.1001/jama.2018.5158

$$\text{SIPW for wellness visit: } SW_{Wi} = \frac{P(W_i = 1)}{P(W_i = 1 | L)}$$

$$\text{SIPW for without a wellness visit: } SW_{Wi} = \frac{1 - P(W_i = 1)}{1 - P(W_i = 1 | L)}$$

$$\text{SIPW for censoring due to missing data: } SW_{Ci} = \frac{P(C_i = 0 | W)}{P(C_i = 0 | W, L)}$$

$$\text{Final SIPW: } SW_{Final} = SW_{Wi} \; x \; SW_{Ci}$$

Figure 1.

Table 1. Overweight and Obesity Prevalence Differences Between Wellness and Other Clinical Visit Types.

	Unweighted %△ *	Weighted %△ *
Overall	-3.9	-1.6
Sex **Female** **Male**	 -4.6 -3.0	 -1.2 -2.1
Race **White** **Black/African American** **Asian** **Other** **Unknown**	 -5.3 -1.7 0.0 -7.1 -3.7	 -2.3 -0.6 0.5 -6.1 -0.2
Ethnicity **Hispanic or Latino** **Non-Hispanic of Latino** **Unknown**	 -0.5 -3.8 -4.2	 0.5 -1.6 -3.5
Age **2-5 years** **6-11 Years** **12-19 years**	 -2.8 -2.7 -3.2	 -2.2 -1.4 -1.6
Care Setting **Emergency** **Inpatient** **Outpatient** **Other**	 4.6 13.2 -3.0 -52.1	 6.6 17.3 -1.2 -48.7
Insurance **Commercial** **Medicaid/Medicare** **Other** **Unknown**	 -2.9 -3.4 -2.7 10.5	 -2.4 -1.7 -2.1 5.6

*%△ = (Prevalence % Among Wellness Visits) - (Prevalence % Among Other Visits)