

Childhood obesity in a diverse school-aged population: place, race and time

Joe Brew | Dept. of Epidemiology | University of Florida | joebrew@gmail.com | 352.318.4553 | databrew.github.io

Dear Sir / Madam,

Thank you for your consideration of my proposal, “Childhood obesity in a diverse school-aged population: place, race and time”. I believe that this project fits well into the Robert Wood Johnson Foundation’s focus on inequalities in childhood obesity, and has the potential to guide policy which will help children “to lead the healthiest lives they can.”

Introduction

The prevalence of obesity among American school-aged children has grown in recent decades at an alarming rate (figure 1)[1]. Obesity has a steep social gradient, and areas that are poor or socially marginalized have above average rates in the USA.[2] Because obesity is closely correlated with socioeconomic status, which is in turn associated with residential location, studying geography’s independent association with childhood obesity is a methodological challenge.

Alachua County, Florida, offers the ideal location to confront this challenge (figure 2). The Alachua County Public Schools system[3] collects a wealth of data on the health of its more than 30,000 students, and maintains databases with reliable information going back over a decade. Furthermore, Gainesville (Alachua’s largest city) has recently established an “open data portal,” allowing for quick and easy access to information on public safety, infrastructure/transportation, and government spending.[4] The combination of these two datasets allows for the discovery of child adiposity’s strongest associations; and the historical element facilitates exploration of obesity’s causal dimensions.

Alachua is a diverse county both geographically (rural/urban) and racially (46% of Alachua public school students are non-white)[5]. Socioeconomically, Alachua’s 41 public schools include multiple highly-ranked magnet programs, which has the effect of attracting students from families who might otherwise send their child to a private institution.

Though numerous explanations have been offered for the steep increase in the *prevalence* of childhood obesity in recent decades[6], and health inequalities in childhood obesity have been explored at the macro level[7][8], little is known about what affects childhood obesity *within* a community. This study proposes to examine those factors with the hope of elucidating causal pathways to childhood obesity, and in doing so, inform community-level interventions.

Aims

The primary purpose of this project is to assess and quantify variance in childhood obesity. Questions to be addressed include:

1. Which social, economic, demographic and geographic features are most closely associated with childhood obesity?
2. What role does residential environment (neighborhood) play in the development of childhood obesity over time, after adjustment for relevant social confounders?
3. Of childhood obesity’s causal factors, which are “modifiable” at the community-level, and how confident can policy-makers be that “built environment” interventions will have their intended effect?

Statement of need

Among Alachua County schoolchildren, the obesity rate is 18.14% and the overweight rate is 37.74%. Additionally, inequality in obesity is high: the prevalence of obesity is highest among black (22.08%) and multiracial

(26.67%) students, and students who qualify for free/reduced lunch are 67% more likely to be obese than those who don't.[5]

Assessing the causes of childhood obesity at the community level is urgent for two reasons. First, the lack of research in the area leaves community and public health practitioners with little ground to stand on when it comes to “best practice” in non-clinical and population-level interventions. Second, given that even the most ambitious (and least realistic) federal governments policy changes are unlikely to seriously reduce the prevalence of childhood obesity in the next 20 years[9], it is crucial that communities

Methodology

This project consists entirely of secondary data collection and analysis. Student data will consist of health screening records (age, weight and height in first, third, sixth and ninth grade), residential location, free/reduced lunch status, and race/ethnicity. Neighborhood features will be constructed using a combination of aggregated student data (e.g. characteristics of nearest peer neighbors), as well as data from the Gainesville open data portal and the United States census. Features to be constructed include infrastructure amenities (distance to nearest park, percent of fastest route to school with bike lanes and/or sidewalks), food options (ratio of grocery to convenience stores), crime rate, and economic growth (rate of new business opening and new home constructions per 1,000 residents). “Neighborhoods” will be defined as census blocks. Analysis will be iterative, but will have three main phases.

1. Exclusive analysis of outcome (BMI percentile for age) by geography using Kulldorff’s method for detecting “clusters” or “hotspots”.[10] Though more commonly used for the identification of communicable disease and other non-chronic condition outbreaks, the Kulldorff scan statistic will be useful in identifying areas where children are much more likely to be obese; this, in turn, will inform how model features should be constructed in the second phase.
2. Feature construction, using qualitative insights gained in phase 1, will be the most laborious phase. Neighborhood characteristics will be compiled into multiple indexes (such as walkability, community cohesiveness, urban/rural-ness, safety, marginalization, etc.), each with multiple variations. In addition to community-level feature construction, student-level data will also be engineered so as to reflect changes over time (e.g. a student coming off free lunch or changing residences). Though a surplus of predictors might be problematic using traditional methodology, the “machine learning” component of the analysis in the third phase means that a multitude of similar predictors will be advantageous to fitting a model.
3. Random forest classification will be used to model adiposity outcomes. Random forest methodology (as opposed to traditional linear / frequentist statistics) is preferred in that its use of bootstrap aggregation (“bagging”) allows for better predictor selection as well as the ranking of predictor importance.

The model will be “trained” on data from the 2003-2004 to 2013-2014 school years and validated on data from the 2014-2015 school year. Geospatial mapping of model misprediction will inform where the model is strongest and weakest, and allow for posterior predictor adjustment, if applicable.

Conclusion

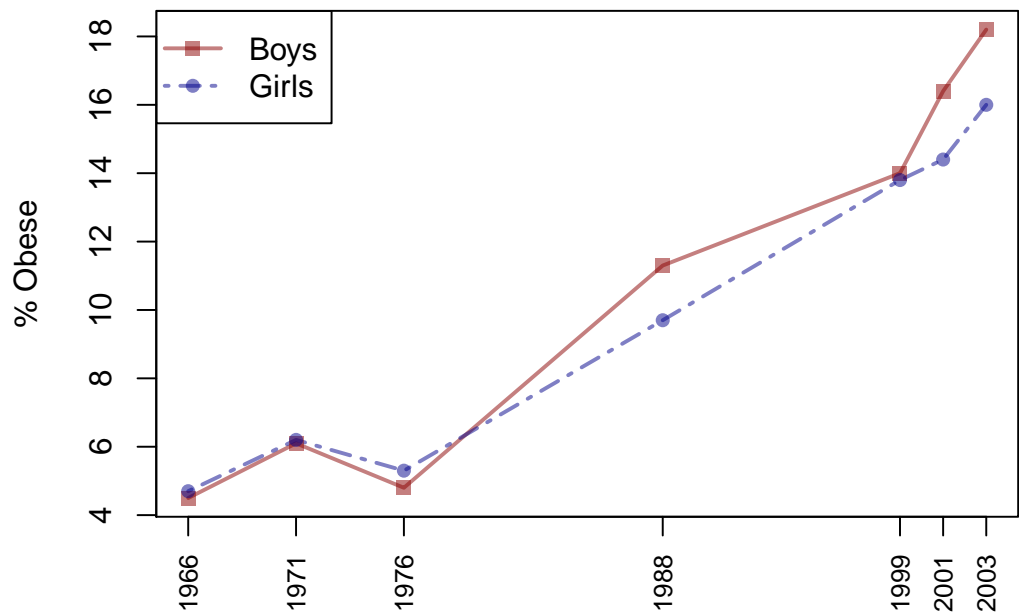
We live in a time of a rapid increase in childhood obesity, a plethora of new and innovative computing methods, and a wealth of data. Through the use of the latter two, this project aims to address the former. The results of this study, both in terms of the methods used as well as the correlations uncovered, will inform both further research as well as community-level policy and practice.

Thank you for your consideration. Please do not hesitate to contact me if you have any questions.

Joe Brew

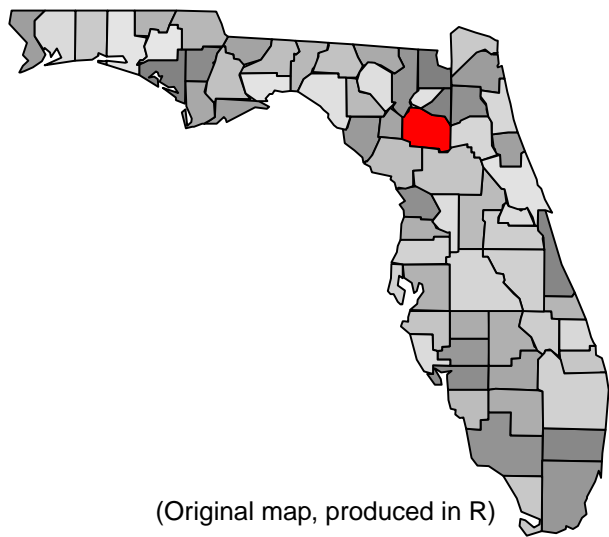
Figures

Figure 1: Obesity among American children[11]



(Original chart, data from NIH)

Figure 2: Alachua County, Florida[12]



(Original map, produced in R)

References

- [1] C Ogden, M Carroll, and K Flegal B Kit. Prevalence of childhood and adult obesity in the united states, 2011-2012. *JAMA*, 311(8):806, feb 2014. doi: 10.1001/jama.2014.732. URL <http://dx.doi.org/10.1001/jama.2014.732>.
- [2] GM Budd and JA Peterson. CE. *AJN, American Journal of Nursing*, page 1, 2014. doi: 10.1097/01.naj.0000457410.83963.4b. URL <http://dx.doi.org/10.1097/01.NAJ.0000457410.83963.4b>.
- [3] *Alachua County Public Schools*. URL <http://www.sbac.edu/pages/ACPS>.
- [4] *City of Gainesville Open Data Portal*. URL <https://data.cityofgainesville.org/>.
- [5] J Brew, J Baker, and K White. Quantifying overweight and obesity among alachua county adults and children. *Community Health Improvement Plan - Internal Documents*, jan 2014.
- [6] J Bishop, R Middendorf, R Babin, and W Tilson. Childhood obesity. *ASPE Research Brief*, page 1, 2005. URL http://aspe.hhs.gov/health/reports/child_obesity/index.cfm.
- [7] Childhood obesity: socioeconomic inequalities and consequences for later cardiovascular health. *Longitudinal and Life Course Studies*, jan 2013. doi: 10.14301/llcs.v4i1.217. URL <http://dx.doi.org/10.14301/llcs.v4i1.217>.
- [8] A Drewnowski. Obesity, diets, and social inequalities. *Nutrition Reviews*, 67:S36–S39, may 2009. doi: 10.1111/j.1753-4887.2009.00157.x. URL <http://dx.doi.org/10.1111/j.1753-4887.2009.00157.x>.
- [9] AH Kristensen, TJ Flottemesch, MV Maciosek, JJ, G Barclay, M Ashe, EJ Sanchez, M Story, SM Teutsch, and RC Brownson. Reducing childhood obesity through u.s. federal policy. *American Journal of Preventive Medicine*, 47(5):604–612, nov 2014. doi: 10.1016/j.amepre.2014.07.011. URL <http://dx.doi.org/10.1016/j.amepre.2014.07.011>.
- [10] M Kulldorff. A spatial scan statistic. *Commun. Statist. - Theory Meth*, June 1997.
- [11] Working Group on Future Research Directions in Childhood Obesity Prevention and Treatment. Future research: Childhood obesity prevention and treatment. *NIH*, aug 2007. URL <http://www.nhlbi.nih.gov/research/reports/2007-child-obesity/index.htm#postmodel>.
- [12] RA Becker, AR Wilks, R Brown, and TP Minka. *Maps package: Draw Geographical Maps*, 2014. URL <http://CRAN.R-project.org/package=maps>. R package version 2.3-9.