Childhood Obesity and CA Public Schools’ Neighborhood Environment: An Application of Machine Learning and GIS for Variable Selection and Analysis.

Alberto M. Ortega Hinojosa, PhD, MPH1; Michael Jerrett, PhD2

1IMPAQ International, LLC, Oakland, CA; 2UCLA Fielding School of Public Health, Los Angeles, CA

Obesity is a pressing worldwide public health concern. Because it is a major risk factor for chronic diseases, including certain cancers, stroke, diabetes, and cardiovascular disease, understanding how to reverse current weight gain trends is imperative. Of particular interest is understanding the effect of pervasive, modifiable environmental exposures.

This research seeks to increase our understanding of the relationship between potentially modifiable attributes of the school environment and childhood obesity.

Methods

Design: Observational study using longitudinal and repeated cross-sectional data.

Exposures: Individual demographics; school level demographics, academic performance, and instruction; school neighborhood land use, safety, air quality, and food access; and census tract demographics, transportation, safety, and food access.

Outcome: Childhood obesity (BMI > 95th percentile)

Analysis: To examine the relationship of social and physical environmental attributes of the school environment and childhood obesity, we first analyzed the relative importance of these attributes using an ensemble learning method (Random Forest). Second, we analyzed the influence of the top ranked variables with one-level random effects binomial regression.

Data Sources

- CA Department of Education
- U.S. Geological Survey
- ESRi’s Business Analyst
- The U.S. Census and other public sources
- Physical Fitness Test (PFT) data

Sample: >5 million student records (from >8,000 unique schools) of California Public School 5th, 7th and 9th grade students, 2003-2007.

Findings

Race and gender, school and census tract socioeconomic variables, and school neighborhood variables ranked as the most important contributors based on the Random Forest.

A small subset of all available variables had a high Random Forest importance value, that is, a large influence on the model.

Census tract violent crime density and school level percent English learners, socioeconomic disadvantage, and diversity index were positively associated with obesity.

School level Academic Performance Index, PE participation, and parental mean education attained and census tract per capita income were negatively associated with obesity.

No physical attributes of the school neighborhood such as distance to convenience store and parkland met the importance criterion.

There is notable variability in observed obesity not explained by our model, suggesting that there are other important unmeasured factors, such as the home environment.

Conclusion

These analyses suggest that socioeconomic attributes influence obesity prevalence and identify several potential modifiable attributes of the school and surrounding environment (e.g., PE participation, academic performance index, and crime) that may be used to combat the obesity epidemic. This is the first statewide analysis to study the school environment, and childhood obesity is unique in its large sample size and comprehensiveness of environmental attributes considered. Unique to this analysis is the use of machine learning methods and GIS to understand the relative importance of variables and to assist in choosing a covariate subset from a large col-linear variable set.

These findings reinforce the importance of gender, race and SES in childhood obesity susceptibility and in the targeting of obesity prevention strategies.

Limitations:

- Important environments, such as the home, were not available
- Only public schools were available
- Data quality
  - Age was to the nearest year
  - CDE Databases
  - Environment is constantly changing
- Only longitudinal data at the school level
- Only three grade levels (5th, 7th, and 9th)

Acknowledgements

This study was funded by research contract 200-2010-37394 from the Centers for Disease Control and Prevention.

We would like to acknowledge the following for their help in making this research possible:

- John Balmes, MD, UC Berkeley School of Public Health; UCSF School of Medicine
- Molly M. Davies, PhD, UC Berkeley School of Public Health
- Jennifer Falbe, ScD, MPH, UC Berkeley School of Public Health

IMPAQ evaluates & enhances pubic programs and polices. Learn more at impaqintl.com.

References