



eGEMs
Generating Evidence & Methods
to improve patient outcomes

The Spatial Distribution of Adult Obesity Prevalence in Denver County, Colorado: An Empirical Bayes Approach to Adjust EHR-Derived Small Area Estimates

David C. Tabano;^{i,ii} Kirk Bol;ⁱⁱⁱ Sophia R. Newcomer;^j Jennifer C. Barrow;ⁱ Matthew F. Daley, MDⁱ

ABSTRACT

Objectives: Measuring obesity prevalence across geographic areas should account for environmental and socioeconomic factors that contribute to spatial autocorrelation, the dependency of values in estimates across neighboring areas, to mitigate the bias in measures and risk of type I errors in hypothesis testing. Dependency among observations across geographic areas violates statistical independence assumptions and may result in biased estimates. Empirical Bayes (EB) estimators reduce the variability of estimates with spatial autocorrelation, which limits the overall mean square-error and controls for sample bias.

Methods: Using the Colorado Body Mass Index (BMI) Monitoring System, we modeled the spatial autocorrelation of adult (≥ 18 years old) obesity ($\text{BMI} \geq 30 \text{ kg m}^2$) measurements using patient-level electronic health record data from encounters between January 1, 2009, and December 31, 2011. Obesity prevalence was estimated among census tracts with ≥ 10 observations in Denver County census tracts during the study period. We calculated the Moran's I statistic to test for spatial autocorrelation across census tracts, and mapped crude and EB obesity prevalence across geographic areas.

Results: In Denver County, there were 143 census tracts with 10 or more observations, representing a total of 97,710 adults with a valid BMI. The crude obesity prevalence for adults in Denver County was 29.8 percent (95% CI 28.4-31.1%) and ranged from 12.8 to 45.2 percent across individual census tracts. EB obesity prevalence was 30.2 percent (95% CI 28.9-31.5%) and ranged from 15.3 to 44.3 percent across census tracts. Statistical tests using the Moran's I statistic suggest adult obesity prevalence in Denver County was distributed in a non-random pattern. Clusters of EB obesity estimates were highly significant ($\alpha=0.05$) in neighboring census tracts. Concentrations of obesity estimates were primarily in the west and north in Denver County.

CONTINUED

Conclusions: Statistical tests reveal adult obesity prevalence exhibit spatial autocorrelation in Denver County at the census tract level. EB estimates for obesity prevalence can be used to control for spatial autocorrelation between neighboring census tracts and may produce less biased estimates of obesity prevalence.

Background

Adult obesity prevalence, defined as the total number of individuals 18 years of age or older with a body mass index (BMI) of greater than or equal to 30 kg/m² among the overall adult population at risk, remains an important public health metric. The Centers for Disease Control and Prevention (CDC) estimate obesity prevalence in the United States to be 35.1 percent for adults 20 years old and older from 2010-2012.¹ Data to capture obesity prevalence estimates at state and local levels are typically derived from survey data in the form of self-reported height and weights, such as the CDC-sponsored, state-administered Behavioral Risk Factor Surveillance System (BRFSS)² Survey-derived measures can underestimate BMI and thus obesity prevalence among children and adults due to individuals underreporting weights and over-reporting heights.^{3,4} Obesity prevalence estimates for small geographies (e.g. census tracts) can be feasibly measured using BMIs calculated from objectively-measured heights and weights from electronic health record (EHR) data collected through routine care from health care providers and provider networks.⁵⁻⁷

When examined at the census tract level, particularly census tracts of varying population levels, obesity prevalence can be distributed in a non-random pattern. Such clustering of observations within

census tracts can induce correlation and impede the reliability of statistical tests, increasing the risk of type 1 error.⁸ Measuring obesity prevalence across census tracts areas should then account for spatial autocorrelation, or similar values of obesity prevalence across neighboring census tracts, as dependency among observations across a geographic area can violate statistical independence assumptions and bias estimates through incorrect probabilities of residuals of estimates and the coefficient estimates themselves.⁸⁻¹¹ Statistical methods to quantify and correct for bias due to spatial autocorrelation are applied in a wide array of health outcomes used to understand the ecology of a sample population across geographic areas, including obesity,^{5,6,12,13} diabetes,^{10,14} mortality risk^{15,16} and other comorbidities.¹⁷⁻²⁰

The primary objective of this paper is to use Empirical Bayes (EB) estimates to reduce the amount of spatial autocorrelation in obesity prevalence estimates with varying sample sizes across geographic areas. We will show that EB estimates can limit the overall mean square-error across geographies where occurrence of obesity prevalence are measured. We will compare crude obesity prevalence estimates to EB estimates across geographic areas. Finally, we will discuss the strengths and limitations of EB estimates for measuring obesity prevalence across census tracts.



Methods

We estimated adult crude and EB obesity prevalence estimates using the Colorado BMI Monitoring System, an electronic health record (EHR) based network comprised of multiple healthcare providers with patients residing in Denver County, Colorado.²¹ The Colorado BMI Monitoring System includes EHR data from January 1, 2009, to December 31, 2011, from Kaiser Permanente Colorado, Denver Health, Children's Hospital Colorado, and High Plains Community Health Center. These data-contributing sites represent a diverse mix of commercially-insured, low-income and homeless patient populations across managed care, safety-net hospital and community clinic providers.

Objectively measured heights and weights obtained during routine care were extracted from the EHRs of each individual site, along with other clinical and demographic characteristics including age, race, ethnicity and gender, geocoded location based on residence address, and insurance coverage at the time of the encounter. Encounters were de-duplicated within each site, and those without measures of height and weight were removed. Data was securely transferred to the Colorado Department of Public Health and Environment (CDPHE), then combined across sites. CDPHE geocoded home addresses to the census tract level and removed addresses from the data. CDPHE applied the CDC BMI SAS[®] macro²² to calculate patient-level BMIs across the system. Biologically-implausible values, defined as extreme values of height or weight for adults at least 18 years of age (heights less than 48 inches, heights greater than 84 inches, weight less than 50 lbs., and weights greater than 700 lbs.),²² were omitted from the system. Additionally, pregnant women, defined as women with an International Statistical Classification of Diseases, Ninth Edition (ICD-9) diagnosis or procedure code for pregnancy or delivery during

the study period, were excluded from the system. Data were then organized by measure date and the most recent BMI record for each individual within the study period was assigned.²³ The Colorado BMI Monitoring System was reviewed and approved by the Kaiser Permanente Colorado and Colorado Multiple Institutional Review Boards. Written consent for inclusion into the BMI Monitoring System was not required.

Empirical Bayes Estimates

Unlike traditional Bayesian estimates, for EB estimates (also known as Stein estimation, penalized estimation and random-coefficient or ridged regression),^{24,25} prior data for estimates comes from the underlying available data itself, not *a priori* from prior information.²⁵⁻²⁷ EB estimates are assumed to vary randomly across data from their respective prior frequency distributions such that the EB posterior estimates represent the frequency confidence intervals themselves. EB estimates provide researchers with the convenience of utilizing existing data to estimate variability in parameter estimates themselves, particularly among data inputs with varying sample sizes of prior estimates.²⁸

The prevalence of obesity for a given geographic area, defined as the number of obese individuals in a geographic area divided by the total population at risk of obesity in the same geographic area, can lead to instability in the variance of obesity prevalence across geographic areas. The variance of the obesity prevalence estimate depends inversely on population at risk; i.e., as the population decreases, the variance of the expected value of the obesity prevalence estimate increases. Smaller sample sizes within geographic areas have larger variance compared to larger sample sizes. EB estimates use "prior" information to reduce the variability (from the overall mean (global mean) prevalence estimate) of the prevalence estimate across geographies,

leveraging “priors” from the global mean prevalence estimate across all census tracts.

EB estimates reduce variability using the inverse function of variance.^{26,29} For areas with lower variance, higher weight is assigned to the observed prevalence estimate. Geographic areas with high variance are weighted less in the calculation of the observed prevalence estimate. Taken together, the application of the global mean mitigates the challenges that arise from large variation of variance estimates due to differences in population sample sizes across geographic areas.

Data Aggregation and Analysis

For this analysis, we used the adult patient population from the Colorado BMI Monitoring System with a most recent valid BMI measure between January 1, 2009, and December 31, 2011, and a geocode based on residence address in Denver County, Colorado.²³ We aggregated the individual, patient-level data to estimate the number of adults ≥ 18 years of age with a valid BMI record in each census tract, the mean BMI for adults by census tract, and number of adults with a BMI ≥ 30 kg/m² in each census tract. We defined coverage as the number of adults in a given census tract with a valid BMI from the Colorado BMI Monitoring System divided by the estimated total number of adults in the census tract from the United States Census 2010 population estimates. We calculated the crude obesity prevalence for each census tract by dividing the total number of obese adults by the total number of adults with a valid BMI in each census tract. Obesity prevalence was estimated among census tracts with ≥ 10 observations in Denver County census tracts during the study period.^{23,30}

We calculated the EB estimate of the obesity prevalence across census tracts in Denver County. We utilized a spatially-naïve EB estimate to reduce the variability of extreme values across census tracts

with the global mean estimate. We employed the Queen’s contiguity matrix³¹ in the spatial analysis and mapping of the EB obesity estimate, which defines neighboring census tracts for all tracts sharing a border in every direction.

We compared the crude obesity prevalence to the EB obesity prevalence graphically, and statistically using a one-sample t-test. We generated maps of crude and EB obesity prevalence estimates for Denver County. We calculated the Moran’s I statistic³² for the crude obesity prevalence and the EB obesity prevalence to estimate the degree of spatial autocorrelation across census tracts in Denver County in each estimate. A statistically-significant Moran’s I of 1.0 indicates perfect spatial autocorrelation across census tracts (i.e., obesity prevalence estimates in a given census tract are completely dependent on the prevalence estimates of neighboring census tracts and are distributed in a non-random pattern). A statistically-significant Moran’s I of -1.0 implies no spatial autocorrelation across census tracts and perfectly random distribution of obesity prevalence.

Data aggregation of Colorado BMI Monitoring System data was performed using SAS[®] 9.2. Geocoded addresses were created using Tele Atlas, U.S. Census, Environmental Systems Research Institute (ESRI) (Pop2010 fields) and Bowes Centrus[®] Desktop v6.01, utilizing the TomTom[®] address point database. Coverage and obesity estimates, statistical tests and maps were calculated and generated using GeoDa[™] 1.4.6.

Results

Table 1 summarizes the data collected from the Colorado BMI Monitoring System for this analysis. There were 143 census tracts in Denver County, Colorado, based on 2010 United States Census Bureau geographic definitions. There were $n=97,710$ adults ≥ 18 years of age with at least one valid



BMI measure and geocoded home addresses in Denver County census tracts within the January 1, 2009-December 31, 2011, study period in the four-site Colorado BMI Monitoring System. Based on United States Census 2010 population estimates, the Colorado BMI Monitoring System sample represented approximately 20.7 percent of the total adult population for Denver County. There were 31,275 adults classified as obese (BMI \geq 30 kg m⁻²) in the 2009-2011 study period. Coverage of the BMI Monitoring System population in Denver County census tracts ranged from 3.7percent to 60.2 percent.

Table 2 summarizes results for the Denver County obesity prevalence estimates. Crude obesity

prevalence for adults was 29.8 percent (95% CI 28.4-31.1%) and ranged from 12.8 to 45.2 percent across individual census tracts. EB obesity prevalence was 30.2 percent (95% CI 28.9-31.5%) and ranged from 15.3 to 44.3%. The Moran's I for crude obesity prevalence was 0.7142 ($p \leq 0.001$) and the Moran's I for the EB obesity prevalence was 0.7307 ($p \leq 0.001$), suggesting spatial autocorrelation in adult crude and EB obesity prevalence estimates in Denver County and that obesity is geographically distributed in a non-random pattern. The mean (standard error) difference in crude and EB obesity prevalence estimates was -0.0046 (0.0003) (one-sample t-test, p -value= <0.001).

Table 1. Summary of BMI Monitoring System for Denver County, CO, Adult Population 2009-2011

DENVER COUNTY	
Colorado BMI Monitoring System Population with valid BMI \geq 18 years old	97,710
U.S. Census 2010 Population Estimates	471,392
Estimated Coverage*	0.2073
Range of Coverage Across Individual Census Tracts	(0.0373, 0.6021)
Total Obese (BMI \geq 30 kg/m ²)	31,275

*Coverage defined as Colorado BMI Monitoring System adult population with valid BMI divided by U.S. Census 2010 Population \geq 18 years old

Table 2. Adult Obesity Prevalence Estimates for Denver County, CO, 2009-2011

	CRUDE OBESITY PREVALENCE (%)	EB OBESITY PREVALENCE (%)	DIFFERENCE BETWEEN MEANS (TWO MEANS, ONE-SAMPLE T-TEST)
Mean (se)	29.8** (0.09)	30.2** (0.08)	-0.0046*** (0.0003)
95% CI	(28.4, 31.2)	(29.0, 31.5)	(-0.0051, -0.0041)
Range	(12.8, 45.2)	(15.3, 44.3)	
Moran's I Statistic	0.7142***	0.7307***	

*, **, *** denotes significance at the 90th, 95th, and 99th percentile, respectively

Figure 1 summarizes the absolute difference of the EB and crude obesity prevalence estimates by BMI Monitoring population across Denver County census tracts. One census tract was omitted from the figure due to insufficient coverage. As the BMI Monitoring population increases, the difference between the crude and EB obesity prevalence estimate decreases.

Maps of obesity prevalence by census tract in Denver County are shown in Figure 2. Crude obesity prevalence estimates showing higher prevalence of obesity were concentrated in the west and north in the county. EB obesity prevalence estimates reveal a similar pattern, but after accounting for variance across census tracts, some census tracts had lower obesity prevalence.

Data points of the Moran's I statistic for individual census tracts for crude and EB obesity prevalence estimates and spatial lag (average across neighboring census tracts) of crude and EB obesity prevalence estimates for Denver County are plotted linearly in the Moran scatterplots in Figure 3. The graphical representation highlights the wider dispersion (distribution of point estimates from the mean) of spatial autocorrelation contained in the crude obesity estimates across individual census tracts compared to the tighter dispersion of spatial autocorrelation in the EB estimates across the same census tracts. High values on the vertical axis are generally associated with high values on the horizontal axis. Similarly, low values on the vertical axis are associated with low values on the horizontal axis. Overall spatial autocorrelation increases slightly

Figure 1. Difference (Empirical Bayes - Crude) Obesity Prevalence Estimates by Census-Tract level BMI Monitoring Population in Denver County, CO, 2009-2011

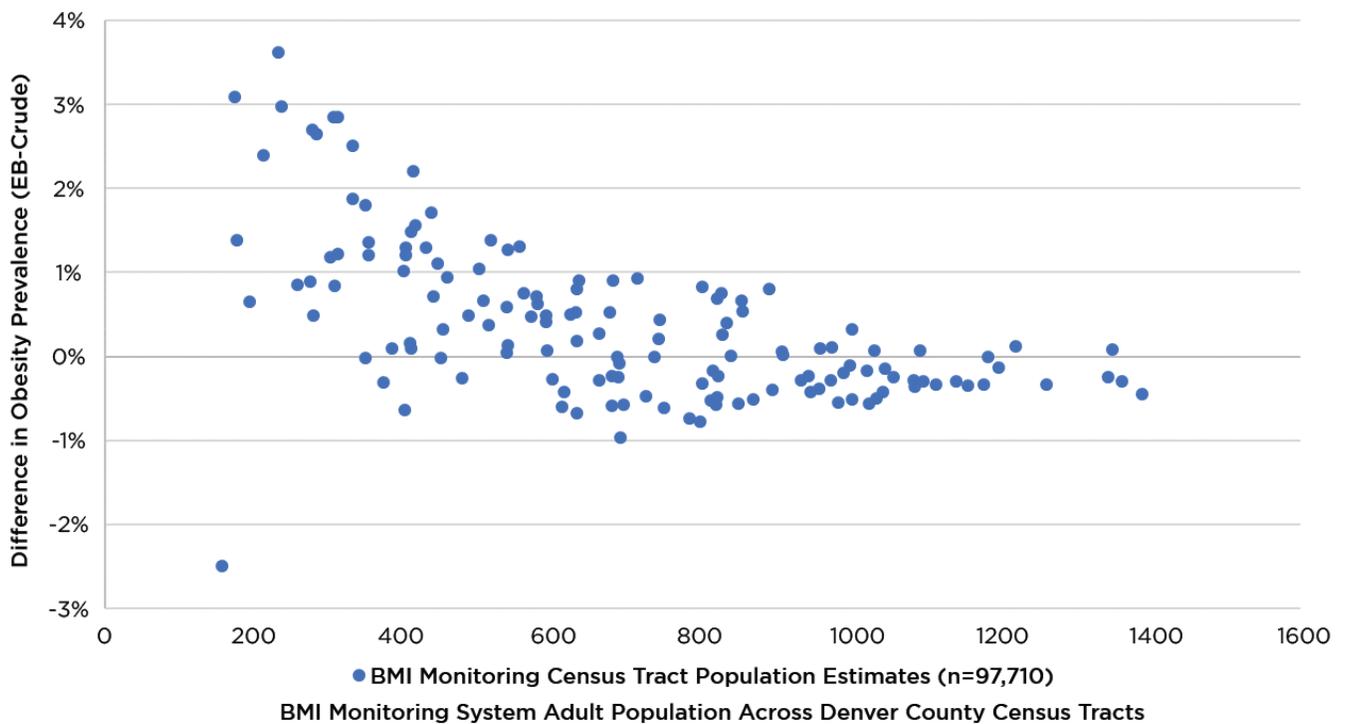




Figure 2. Denver County Obesity Prevalence Estimates by Census Tract, 2009-2011

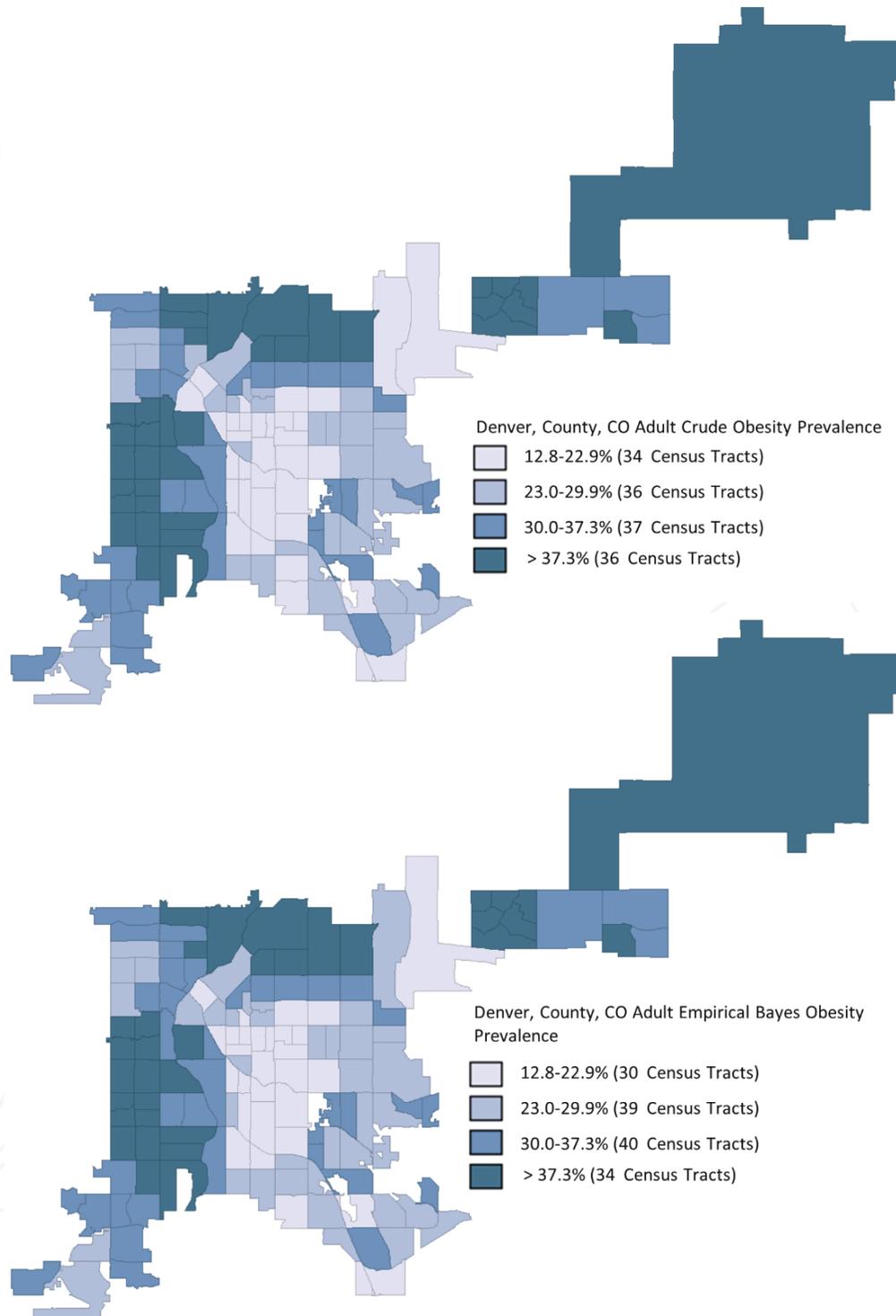
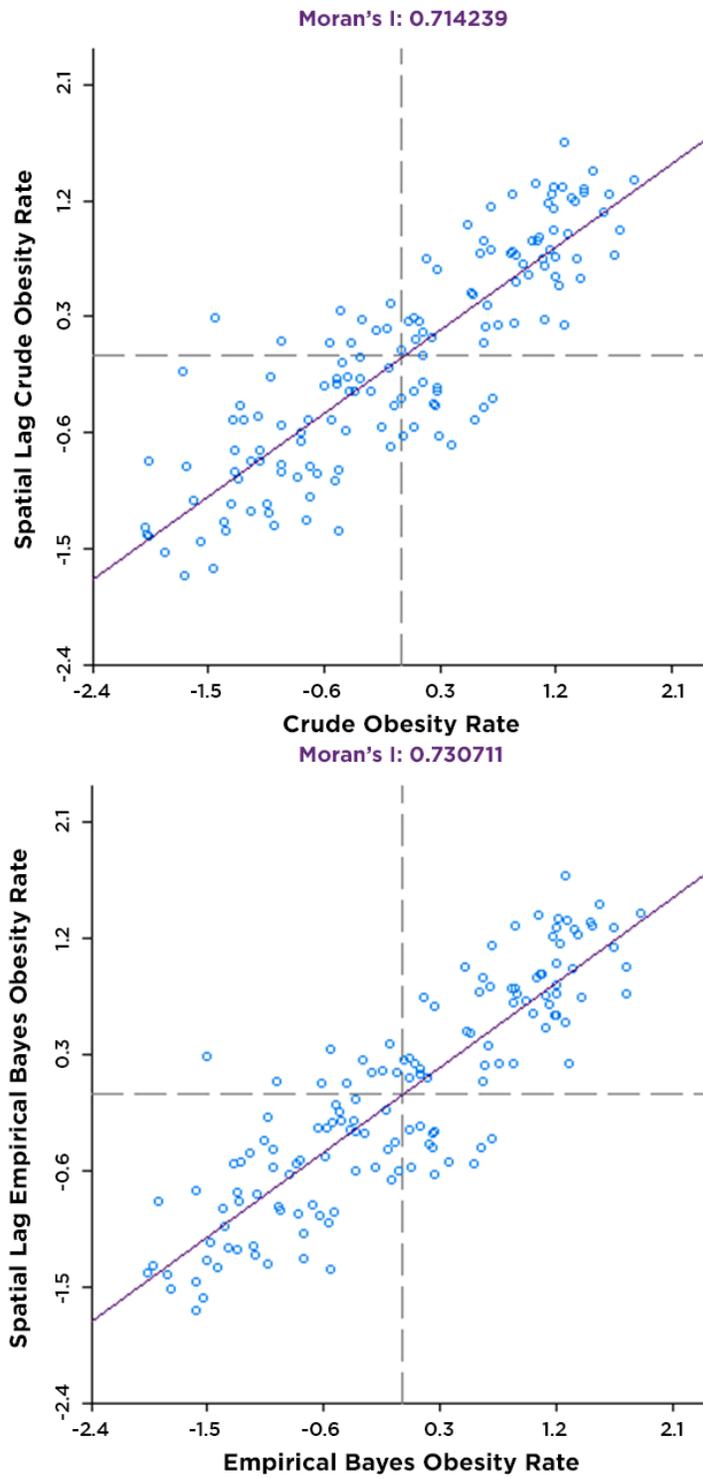


Figure 3. Moran Scatterplot of I-Statistic Comparisons of Obesity Prevalence Estimates





from the crude obesity prevalence estimates in Denver County to the EB obesity prevalence estimates, but the autocorrelation among individual census tracts was reduced in the EB estimates.

Discussion

This paper presented the use of EB estimation to reduce spatial autocorrelation in obesity prevalence estimates across small geographies with different sample sizes. We estimated adult crude and EB obesity prevalence estimates in Denver County, Colorado using EHR-derived BMI data from the Colorado BMI Monitoring System. We compared and quantified the differences in crude and EB estimates, and showed that EB estimates can limit the errors in the residual estimates across geographies where occurrence of obesity prevalence are measured.

The crude adult obesity prevalence estimate derived from the Colorado BMI Monitoring System for Denver County was 29.8 percent; the EB obesity prevalence estimate was 30.2 percent. The difference between the two obesity prevalence estimates was statistically significant, revealing EB obesity prevalence for adults were non-random in Denver County at the census tract level. Clusters of EB obesity were highly significant ($\alpha \leq 0.05$) in neighboring census tracts of high obesity prevalence. The Moran's I statistic for the EB obesity prevalence estimate showed that a high degree of spatial autocorrelation exists within Denver County, quantifying the degree to which obesity prevalence in neighboring census tracts were correlated across Denver County. The results suggest autocorrelation of obesity prevalence at the census tract level exists and should be accounted for to limit bias in calculated obesity estimates.

While comparisons of estimates derived from the BRFSS and Colorado BMI Monitoring System cannot be made directly due to sample size and data

collection methods, assessing the reasonableness of obesity prevalence estimates derived from the Colorado BMI Monitoring System with an established alternative is important to validate this novel approach. Estimates of obesity prevalence from the 2009-2010 BRFSS adult obesity estimates for Denver County (19.6%; 95% CI [16.8- 22.4]),² but may underestimate obesity due to phone survey respondents' underreporting weights and over-reporting heights.³ Similarly, the Colorado BMI Monitoring System may overestimate BMI and obesity prevalence due to selection bias of patient populations of the data-contributing sites and the use of EHR data as the source. Data may misrepresent the obese population in Denver County, as the data can only reflect patients who utilize healthcare services at data-contributing sites. Patients of federal hospitals (e.g., Veteran's Affairs) or other commercial insurance providers in Denver County may not be represented in these estimates.

Additional analyses can be conducted to further identify spatial autocorrelation within the BMI Monitoring System. EB estimates help identify if prevalence estimates across geographic areas contain spatial autocorrelation and whether estimates are distributed at random or in a non-random pattern. Demographic-specific obesity prevalence estimates and associated Moran's I statistics can be compared to determine which particular demographic strata (e.g., age groups, gender, race and ethnicity) may be contributing more or less to spatial autocorrelation across census tracts for a given geographic area. Socioeconomic status (SES) and environmental data can be modeled at the census tract level to further determine the extent of autocorrelation due to these additional variables. Several studies have found SES and environmental exposures to explain large portions of variation in obesity prevalence across census tracts.^{5,33-35}

The Moran's I statistic can be employed in studies as a useful tool for estimating correlations across census tracts. If neighboring census tracts are highly correlated but not accounted for, obesity prevalence estimates may be incorrect. Policy decisions and community-level interventions may be made from inaccurate estimates, which can in turn hinder the impact of such public health efforts. Public health entities and community policy makers can use EB estimation and the Moran's I statistic to infer variability of obesity prevalence, as well as SES and environmental exposures within clusters of high or low obesity prevalence, that may be correlated with obesity.

Limitations

There are several limitations to the use of EB estimates to calculate obesity prevalence, and to the Colorado BMI Monitoring System for measuring obesity prevalence over a large population and geographic area. The BMI Monitoring System does not employ a patient-master index for de-duplicating patients across data-contributing providers. Rather, patients having insurance coverage for another data-contributing site at the time of most-recent height and weight measure were reallocated to that site.

Weighting obesity prevalence estimates at the census tract level by the global mean obesity prevalence makes understanding the EB estimate difficult and not necessarily accessible for the public, a key aim for the Colorado BMI Monitoring System. Conversely, public consumers of obesity prevalence estimates may be more interested in the accuracy of the estimates themselves (i.e., the accuracy of obesity prevalence estimates relative to the "true" obesity prevalence) and less on the derivation of the estimates themselves. EB estimates do provide an estimate of obesity prevalence with reduced spatial autocorrelation for modeling the association between obesity, SES and environmental risk factors

across neighboring geographic areas, reducing bias in estimates and interpretation of correlations.⁵

Other measures of spatial autocorrelation were not considered in this study, including autoregressive parameter specification or simultaneous autoregressive (SAR) modeling,³⁶ as well as spatial filtering.³⁷ These techniques are often used in assessing correlation between exposure and outcome measures across geographies, but can be employed as an alternative to EB estimation.

Conclusion

EB estimates of obesity prevalence can reduce bias of estimates across geographic areas with different sample sizes using the data within the sample to generate prior estimates, providing estimates of obesity prevalence that are less prone to bias due to spatial autocorrelation. EB estimates can help researchers discern whether prevalence estimates are distributed across geographies in non-random patterns. EHR data can provide a rich source of information to measure disease prevalence across populations and geographies. Additional analyses of demographic, SES and environmental data can further define spatial variance and autocorrelation across census tracts.

References

1. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. *Jama*. 2014;311(8):806-14.
2. Colorado Health Information Dataset, Behavioral Risk Factor Surveillance System 2009-2010 [Internet]. [cited February 3, 2016]. Available from: http://www.chd.dphe.state.co.us/cohid/topics.aspx?q=Behavioral_Risk_Factors.
3. Ezzati M, Martin H, Skjold S, Vander Hoorn S, Murray CJ. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *Journal of the Royal Society of Medicine*. 2006;99(5):250-7.
4. Gordon NP, Mellor RG. Accuracy of parent-reported information for estimating prevalence of overweight and obesity in a race-ethnically diverse pediatric clinic population aged 3 to 12. *BMC pediatrics*. 2015;15:5.



5. Roth C, Shivade CP, Foraker RE, Embi PJ. Integrating population- and patient-level data for secondary use of electronic health records to study overweight and obesity. *Studies in health technology and informatics*. 2013;192:1100.
6. Drewnowski A, Rehm CD, Arterburn D. The geographic distribution of obesity by census tract among 59 767 insured adults in King County, WA. *International journal of obesity* (2005). 2014;38(6):833-9.
7. Tabano, David C., Elizabeth Cole, Erin Holve, and Arthur J. Davidson. "Distributed Data Networks That Support Public Health Information Needs." *Journal of Public Health Management and Practice* (2017).
8. Legendre P. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*. 1993;74(6):1659-73.
9. Diniz-Filho JAF, Bini LM, Hawkins BA. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*. 2003;12(1):53-64.
10. Congdon P. Spatial variation in attributable risks. *Spatial and spatio-temporal epidemiology*. 2015;12:39-52.
11. Kühn I, Dormann CF. Less than eight (and a half) misconceptions of spatial analysis. *Journal of Biogeography*. 2012;39(5):995-8.
12. Farhadian M, Moghimbeigi A, Aliabadi M. Mapping the obesity in iran by bayesian spatial model. *Iranian journal of public health*. 2013;42(6):581-7.
13. Huang R, Moudon AV, Cook AJ, Drewnowski A. The spatial clustering of obesity: does the built environment matter? *Journal of human nutrition and dietetics : the official journal of the British Dietetic Association*. 2015;28(6):604-12.
14. Drewnowski A, Rehm CD, Moudon AV, Arterburn D. The geography of diabetes by census tract in a large sample of insured adults in King County, Washington, 2005-2006. *Preventing chronic disease*. 2014;11:E125.
15. Castillo-Carniglia A, Kaufman JS, Pino P. Geographical distribution of alcohol-attributable mortality in Chile: a Bayesian spatial analysis. *Addictive behaviors*. 2015;42:207-15.
16. Fox L, Serre ML, Lippmann SJ, Rodriguez DA, Bangdiwala SI, Gutierrez MI, et al. Spatiotemporal approaches to analyzing pedestrian fatalities: the case of Cali, Colombia. *Traffic injury prevention*. 2015;16(6):571-7.
17. Hsu CY, Chang SS, Lee ES, Yip PS. "Geography of suicide in Hong Kong: spatial patterning, and socioeconomic correlates and inequalities". *Social science & medicine* (1982). 2015;130:190-203.
18. Tsiko RG. Bayesian spatial analysis of childhood diseases in Zimbabwe. *BMC public health*. 2015;15:842.
19. Zayeri F, Kavousi A, Najafimehr H. Spatial analysis of Relative Risks for skin cancer morbidity and mortality in Iran, 2008-2010. *Asian Pacific journal of cancer prevention : APJCP*. 2015;16(13):5225-31.
20. Tsai PJ. Application of Moran's test with an empirical Bayesian rate to leading health care problems in Taiwan in a 7-year period (2002-2008). *Global journal of health science*. 2012;4(5):63-77.
21. Colorado BMI Monitoring System [October 31, 2016]. Available from: http://www.cohealthmaps.dphe.state.co.us/colorado_bmi_monitoring_system/.
22. A SAS Program for the 2000 CDC Growth Charts (ages 0 to <20 years): Centers for Disease Control and Prevention; [October 31, 2016]. Available from: <http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>.
23. Centers for Disease Control and Prevention, 2006. Principles of epidemiology in public health practice: an introduction to applied epidemiology and biostatistics.
24. Lee PM. *Bayesian statistics: an introduction*: John Wiley & Sons; 2012.
25. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology*. 2006;35(3):765-75.
26. Krutchkoff RG. Empirical Bayes Estimation. *The American Statistician*. 1972;26(5):14-6.
27. Robbins H. The empirical Bayes approach to statistical decision problems. *Herbert Robbins Selected Papers*: Springer; 1985. p. 49-68.
28. Lwin T, Maritz J. Empirical Bayes approach to multiparameter estimation: with special reference to multinomial distribution. *Annals of the Institute of Statistical Mathematics*. 1989;41(1):81-99.
29. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics*. 1991:283-94.
30. Colorado State Department of Public Health and Environment. Guidelines for working with small numbers. <http://www.cdph.state.co.us/cohid/smnumguidelines.html>
31. Anselin L. *Exploring spatial data with GeoDaTM: a workbook*. Urbana. 2004;51:61801.
32. Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. *Geographical analysis*. 2006;38(1):5-22.
33. Cesare MD, Bhatti Z, Soofi SB, Fortunato L, Ezzati M, Bhutta ZA. Geographical and socioeconomic inequalities in women and children's nutritional status in Pakistan in 2011: an analysis of data from a nationally representative survey. *The Lancet Global health*. 2015;3(4):e229-39.
34. Drewnowski A, Aggarwal A, Tang W, Moudon AV. Residential property values predict prevalent obesity but do not predict 1-year weight change. *Obesity (Silver Spring, Md)*. 2015;23(3):671-6.
35. Drewnowski A, Moudon AV, Jiao J, Aggarwal A, Charreire H, Chaix B. Food environment and socioeconomic status influence obesity rates in Seattle and in Paris. *International journal of obesity* (2005). 2014;38(2):306-14.
36. Griffith DA, Layne LJ, Ord J, Sone A. *A casebook for spatial statistical data analysis: a compilation of analyses of different thematic data sets*: Oxford University Press on Demand; 1999.
37. Haining R. Bivariate correlation with spatial data. *Geographical Analysis*. 1991;23(3):210-27.