

RESEARCH

Open Access



An electronic medical records study of population obesity prevalence in El Paso, Texas

Jennifer J. Salinas^{1*}, Jon Sheen¹, Navkiran Shokar², Justin Wright², Gerardo Vazquez² and Ogechika Alozie³

Abstract

Background: In this study, we determine the feasibility of using electronic medical record (EMR) data to determine obesity prevalence at the census tract level in El Paso County, Texas, located on the U.S.-Mexico border.

Methods: 2012–2018 Body Mass Index (BMI kg/m²) data from a large university clinic system in was geocoded and aggregated to a census tract level. After cleaning and removing duplicate EMR and unusable data, 143,524 patient records were successful geocoded. Maps were created to assess representativeness of EMR data across census tracts, within El Paso County. Additionally, maps were created to display the distribution of obesity across the same geography.

Results: EMR data represented all but one El Paso census tract. Representation ranged from 0.7% to 34.9%. Greatest representation were among census tracts in and around clinics. The mean EMR data BMI (kg/m²) was 30.1, this is approximately 6% less than the 36.0% estimated for El Paso County using the Behavioral Risk Factor Surveillance Study (BRFSS) estimate. At the census tract level, obesity prevalence ranged from 26.6 to 57.6%. The highest obesity prevalence were in areas that tended to be less affluent, with a higher concentration of immigrants, poverty and Latino ethnic concentration.

Conclusions: EMR data use for obesity surveillance is feasible in El Paso County, Texas, a U.S.-Mexico border community. Findings indicate substantial obesity prevalence variation between census tracts within El Paso County that may be associated with population distributions related to socioeconomics.

Keywords: Electronic medical records, Geographic information systems, Obesity, Body Mass Index, Mexican Americans

Background

Obesity prevalence has reached an epidemic level across the United States, significantly increasing since the 1990's [1–3]. This major risk factor for many chronic health conditions, such as diabetes and certain cancers, disproportionately affecting Latino populations [4, 5]. While these trends have been well-documented, little success

has been made in mitigating community or individual level risk factors.

Obesity surveillance has been typically conducted using the Center for Disease Control and Prevention (CDC) programs such as the National Health and Nutrition Evaluation Study (NHANES) and the Behavioral Risk Factor Surveillance System (BRFSS) [6, 7]. These and other data sources utilize national cross-sectional health survey methods to collect data on lifestyle behaviors among other health-related metrics. From these data sources, food environment, built environment,

*Correspondence: jennifer.salinas@ttuhsc.edu

¹ Department of Molecular and Translational Medicine, Texas Tech Health Sciences Center El Paso, 5001 El Paso Dr., El Paso, TX 79905, USA
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

segregation, poverty and other contextual risk factors for obesity have been well-established for Latino and other health disparate communities [8–13]. Though this large-scale study design is highly beneficial for national, state, and county level surveillance, it is still limited in providing insight into context of obesity at a smaller geographic scale, i.e. within county or city levels. As place matters in the context of an individual’s health, the more specific the unit of analysis, the better community-based prevention initiatives are able to target high rise areas within cities and counties [14, 15]. Relying on county-level data that is available can make prevention efforts for under-resourced communities futile, since they provide aggregated estimates, not taking into account important variations within a county or city.

Classically, surveillance at a more granular level like census tracts have been conducted in efforts to contain infectious disease spread, such as most recently during the COVID-19 pandemic [16]. GIS and other geographic tracking technologies have been used to track infectious disease within communities across international contexts [16]. These technologies have provided infectious disease researchers and public health departments on-the-ground and real-time information that has guided intervention and prevention programming to curb existing trends within the communities that they serve [17, 18]. In some cases, public health departments have been able to act on real time information and prevent mass exposure to influenza and other viruses. A localized approach such as this applied to the surveillance of obesity would allow for more specific, community-based interventions [18–20]. This may be particularly helpful in areas that have limited resources to conduct wide-scaled intervention efforts.

In 2009, the Health Information Technology for Economic and Clinical Health (HITECH) Act was signed into law as an effort to promote the widespread use of EMR in a “meaningful” way. Aside from the clinical and organizational advantages, EMR enables more available, combined aggregated data across populations leading to better, yet not perfect, health outcome surveillance and research production that benefits overall society [21–25]. Electronic medical records (EMR) is a source for objective data rather than self-reported data, and, could provide a better and low-cost obesity surveillance option for public health departments looking to provide targeted prevention measures.

The purpose of this study is to assess the feasibility and applicability of EMR data obtained from university and county safety-net outpatient electronic medical records to provide census tract-level obesity estimates and distributions across El Paso County, Texas. El Paso, Texas is a unique context in that it is predominantly Mexican

American (82%), is socioeconomically disadvantage relative to other cities its size, and has a high prevalence of obesity-related diseases ²⁶. This analysis will provide a more detailed picture of the distribution of obesity within the county, facilitating better targeted efforts to reduce obesity.

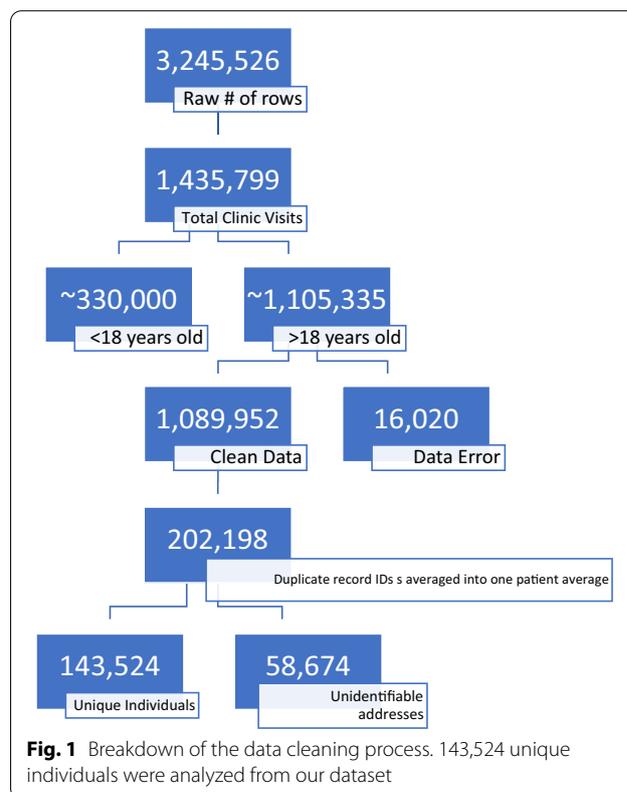
Methods

Design and setting

2012–2018 adult patient data was extracted from the Electronic Medical Records (EMR) systems from Texas Tech University Health Sciences Center El Paso Clinics and University Medical Center El Paso outpatient clinics in El Paso, Texas. Analysis was completed in 2018–2019. The data process can be seen in Fig. 1. Raw data for over 3.2 million observations were cleaned and prepared for analysis.

Data preparation

Data cleaning processes are documented in a previous paper [27]. Briefly, duplicates and inconsistencies were either removed or corrected when possible. Any case with unverifiable addresses or incomplete height or weight were removed. Finally, EMR-based Body Mass Indices that seemed out of range (above 100) were recalculated using height and weight or removed completely when



data was missing. Patients were assigned census tracts using street addresses. Any unidentifiable addresses were removed. Figure 1 provides an overview of the cleaning process and how we arrived at 143,524 participants from the 3,245,526 raw case files at the onset of the cleaning process. There were a total of 161 possible census tracts considered for this analysis. Patient representation per census tract was determined by a proportion of patients per census tract by total census tract population. This research was approved by the Texas Tech University Health Sciences Center El Paso Institutional Review Board (IRB) for Human Subjects Research. All methods and procedures were performed in accordance with the IRB guidelines. Since this was a secondary data analysis of existing electronic medical records, requirement for signed informed consent was waived by the Institutional Review Board for Human Subjects Research at Texas Tech University Health Sciences Center El Paso.

Analysis

Population representation was estimated using total number of patient records divided by total population for each census tract. Average Body Mass Index (BMI (kg/m^2)) for each of the 161 census tracts was calculated using patient EMR data. Census tract obesity prevalence (BMI (kg/m^2) 30 or above) was determined for each of the 161 census tracts by dividing the percent of patients with a BMI of 30 or above by the total number of patients represented in each census tract. Census tract patient record representation and obesity prevalence were entered into ArcGIS (ESRI) and mapped for El Paso County.

Results

Figure 2 displays the percentage of census tract total population is represented by the EMR patient records. The percent of EMR patients per census tract ranged from 0.7% to as high as 34.9%. Patients were most represented in the Northeast, Downtown/Lower Valley and Far Eastside of El Paso County and in and around clinic catchment areas. Patients were least represented in the Fort Bliss area, the large North-Central area of the map. Fort Bliss generally provides healthcare on base through William Beaumont hospital, rather than community-based clinic. The lack of representation in that area are consistent with what we would expect.

Figure 3 presents EMR record obesity prevalence by census tract across El Paso County. Prevalence ranged from 26.6% to as high as 57.6%. Census tracts with the highest obesity prevalence were located in the Lower Valley, Far Eastside and Northwest El Paso County. This area corresponds geographically to areas where Latino ethnicity, poverty and immigrant concentration is the highest. On the other hand, the Westside of El Paso County had

the pocket with the lowest prevalence, this area is also the most affluent of the County.

Discussion

Latinos are disproportionately impacted by obesity. Strategies for curbing high prevalence rates among Latinos have generally been informed by data collected at a state or federal level. The continued focus on surveillance at such a high level of aggregation has provided little insight into city or county risk factors that are actionable in addressing current obesity trends. This study provides evidence of feasibility of electronic medical record obesity data as a tool to surveil population-level obesity within small geographic units. Findings from this study also demonstrate the uneven distribution of obesity within small units of a city or county which is not as well represented in national surveillance data at a county level.

Latino communities have been described as obesogenic by multiple previous studies [28, 29]. However, most of these studies have used large-scale sampled data that often only represent urban residing populations or ethnic enclaves within a heterogeneous ethnic structure. This approach has limited the ability to infer factors responsible for obesity in communities where Latinos are the majority. El Paso County Texas is predominantly Mexican American-Latino. While the overall estimated prevalence of obesity is 34.9% based on BRFSS estimates, this study's findings suggest that there may be substantial variation of the degree of obesity that may co-vary with ethnic concentration and socioeconomic status by census tract (cite other paper). Our study's findings suggest that known areas of El Paso County that are more heavily concentrated with Latinos, immigrants and lower average socioeconomic status, may also have a higher burden of obesity, relative to the 34.9% overall estimate from the BRFSS. This variation may misrepresent that extent of obesity in areas that may be more heavily populated with Latinos at the same time as being socioeconomically diverse. Our research findings are based on analysis using EMR from a university-base and county outpatient clinic system and may not represent the true county obesity prevalence, since selection in insurance status or presence of chronic conditions may have biased our findings. Therefore, it would be important to replicate this analysis using a pool of EMR from multiple providers. This is an important direction for research given the high burden of obesity in Latino communities and limited effectiveness in curbing trends through currently available interventions.

In this paper, we demonstrated a high feasibility of using a EMR data, to analyze health outcomes across a large population. EMR databases allow for quick

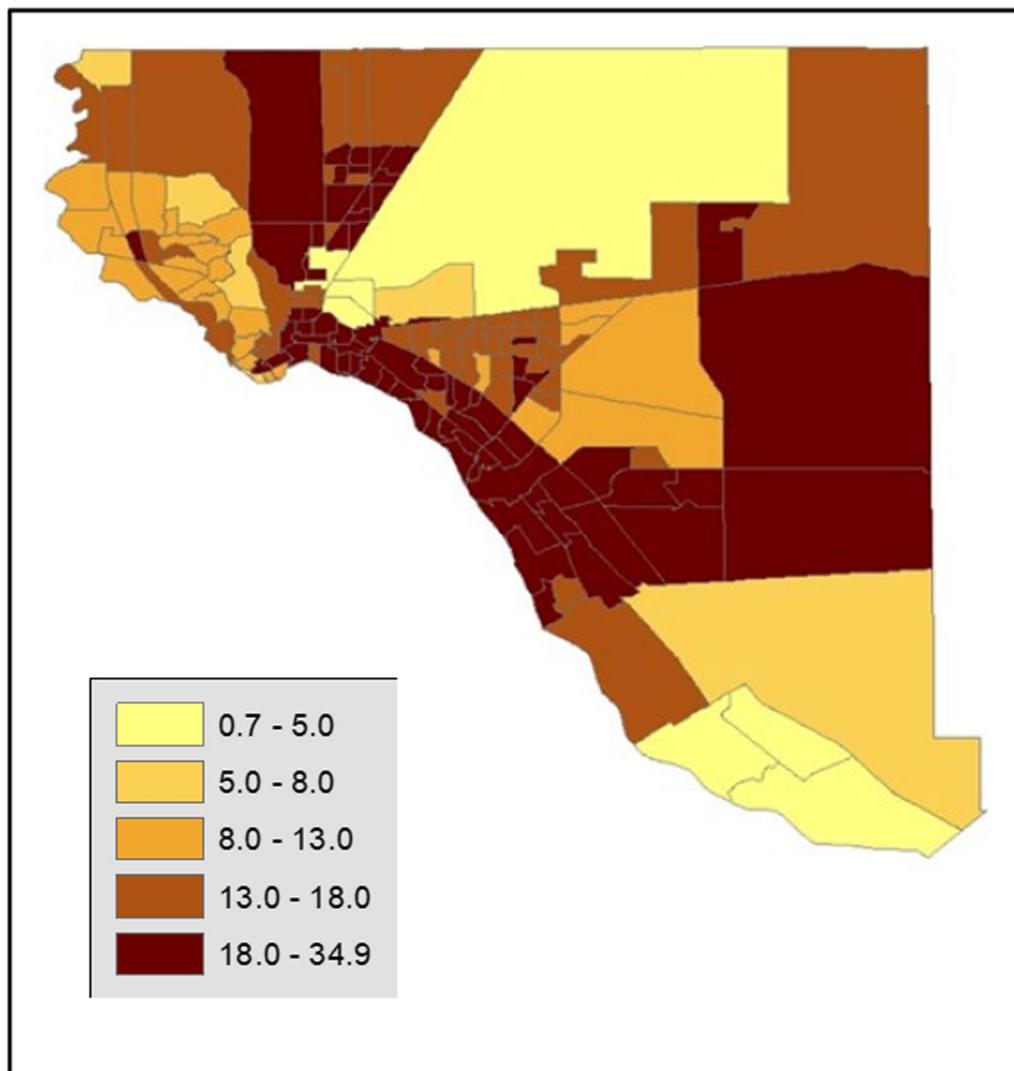


Fig. 2 Percent of census tract total represented by EMR patient records

extraction and analysis of large quantities of data, and do not require an abundance of resources and manpower, as discussed in Funk et al.'s study [17]. We were able to analyze nearly a fifth of the population -143,524 unique adults in the El Paso area of over 800,000 adults²⁶. While there are shortcomings still with respect to selection bias, this paper's intent was to demonstrate feasibility, potential use in disease surveillance in areas that otherwise lack this capacity, and provide basis for future surveillance and intervention work.

Using readily available BMI (kg/m^2) data within university-based and county health clinics allowed us to assess the distribution of obesity within El Paso County. Few studies have previously looked at obesity prevalence with small geographic units such as cities or counties, which

is suggested by Jia et al. [14]. Roth et al. [10] successfully explored linking EMR and community data with a large sample size to study factors associated with obesity, but used a zip-code level of analysis. The study by Shafiri et al. [31] is one of the few studies that used both EMR for a large sample and the census tract level for their analysis of the built environment and ethnic disparities in childhood obesity. Funk et al. demonstrated the ease of studying over 380,000 patients from a university-based healthcare system and showed that the results are comparable to NHANES [17]. Our study was an attempt to replicate this approach in a context of disadvantage and high ethnic homogeneity. The previous finds from Shafiri et al., and Funk et al. coupled with the results from the present study demonstrate the feasibility of EMR use in

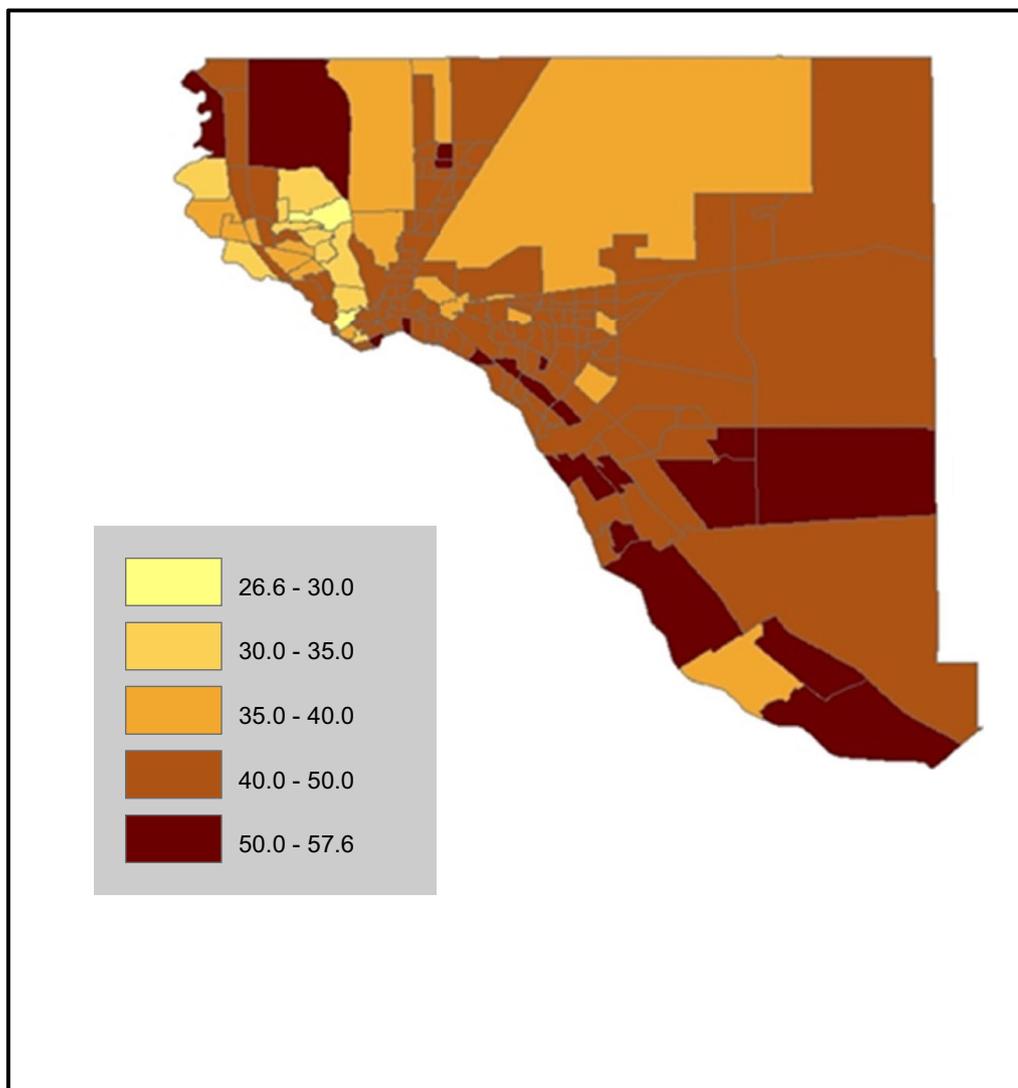


Fig. 3 % EMR patient records with 30 or greater BMI (kg/m^2) by census tract

estimating obesity and other chronic diseases at a much more granular level than currently available national estimates at a county level. Future studies are needed to determine the reliability of EMR data to estimate population-level obesity prevalence. The public health implications for this type of use are not only limited to obesity, but to other related health conditions traceable within a patient's medical record.

This study has a number of strengths. First, the EMR data represented approximately 21% of the overall El Paso population. In some census tracts, the proportion was close to 35%. Secondly, this secondary data source has measured data and not self-reported data like many similar studies. Finally use of existing medical records is

a cost effective way to conduct surveillance for obesity and other health outcomes. The findings from this study have a number of limitations that should be noted. First, EMR data were only obtained from an university clinic system clinics and while in many cases the patients represented a large proportion of a given census tract, it may not represent fully the El Paso County population. For example, there was variation in the percentage of census tract population the EMR represented from as low as 0.7% to as high as 35%. It is likely in census tracts where representation was low, there may be substantial variance in the data that may contribute to over or underestimate of obesity. Additionally, patients that visit the clinics on a regular basis, may be sicker and therefore, we may be

overestimating the true census tract-level obesity prevalence. This study was intended to demonstrate feasibility in a community that is underrepresented and carries a high burden of chronic diseases related to obesity [29]. Future work using multiple EMR data sources, would reduce potential bias and improve county-wide estimates. Furthermore, analysis in other Latino dominate communities would need to be conducted to determine feasibility and applicability in other settings.

Conclusions

Use of EMR data for surveillance of obesity prevalence within El Paso County, TX is feasible and may provide a better snapshot of the distribution of obesity within the county than BRFSS estimates. With the potential of using EMR data for obesity and other chronic conditions, public health officials have an opportunity to engage in precision surveillance, identify subpopulations who might be at greatest risk. Including population data such as education, health insurance, mean family income, and other sociodemographic factors could lead to more effective targeted prevention efforts. However, further investigation is needed to determine data quality of these fields in the available EMR databases. This study provides evidence of its potential utility in understanding the distribution of obesity within a Latino community and is a great starting point for further examination of community risk factors for obesity.

Acknowledgements

Not applicable.

Authors' contributions

JJS is the principal investigator of the study, conceptualizing the project, overseeing data analysis and manuscript development. JS analyzed and interpreted electronic medical record data, as well as, contributing to manuscript writing. OA provided the access to the research team and editing. NS, JW, GV contributed to the writing and the interpretation of the findings. All authors read and approved the final manuscript.

Funding

This study was funded by an evidence-based prevention grant from the Cancer Prevention and Research Institute of Texas (CPRIT) (PP180026).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This analysis was approved by the Texas Tech University Health Sciences Center Institutional Review Board (IRB) (# E18089). All methods were performed in accordance with the relevant guidelines and regulations of the IRB. Since this was a secondary data analysis of existing electronic medical records, requirement for signed informed consent was waived by the Institutional Review Board for Human Subjects Research at Texas Tech University Health Sciences Center El Paso.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Molecular and Translational Medicine, Texas Tech Health Sciences Center El Paso, 5001 El Paso Dr., El Paso, TX 79905, USA. ²Department of Family and Community Medicine, Texas Tech Health Sciences Center El Paso, El Paso, TX, USA. ³Del Sol Medical Center, El Paso, TX, USA.

Received: 8 February 2021 Accepted: 8 February 2022

Published online: 22 February 2022

References

- Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in obesity and severe obesity prevalence in US youth and adults by sex and age, 2007–2008 to 2015–2016. *JAMA*. 2018;319(16):1723–5.
- Grabner M. BMI trends, socioeconomic status, and the choice of dataset. *Obes Facts*. 2012;5(1):112–26.
- Sturm R, Hattori A. Morbid obesity rates continue to rise rapidly in the United States. *Intern J Obes*. 2013;37(6):889–91.
- Pi-Sunyer FX. The obesity epidemic: pathophysiology and consequences of obesity. *Obes Res*. 2002;10(2):97–104.
- American Cancer Society. Cancer facts & figures for Hispanics/Latinos 2018–2020. American Cancer Society, Inc. 2018.
- Flegal KM, Ogden CL, Fryar C, Afful J, Klein R. Comparisons of self-reported and measured height and weight, BMI, and obesity prevalence from national surveys: 1999–2016. *Obesity (Silver Spring)*. 2019;27(10):1711–9.
- Forrest KYZ, Leeds MJ, Ufelle AC. Epidemiology of obesity in the Hispanic adult population in the United States. *Family Comm Health*. 2017;40(4):291–7.
- Fitzpatrick KM, Shi X, Willis D, et al. Obesity and place: chronic disease in the 500 largest U.S. cities. *Obes Res Clin Pract*. 2018;12(5):421–5.
- Lundeen EA, Park S, Pan L, O'Toole T, Matthews K, Blanck HM. Obesity prevalence among adults living in metropolitan and nonmetropolitan counties - United States, 2016. *MMWR*. 2018;67(23):653–8.
- Roth C, Foraker RE, Payne PR, Embi PJ. Community-level determinants of obesity: harnessing the power of electronic health records for retrospective data analysis. *BMC Med Inf Decis Mak*. 2014;14:36.
- Yu C, Woo A, Hawkins C, Iman S. The impacts of residential segregation on obesity. *J Phys Act Health*. 2018;15(11):834–9.
- Krishna A, Razak F, Lebel A, Davey Smith G, Subramanian SV. Trends in group inequalities and interindividual inequalities in BMI in the United States, 1993–2012. *Am J Clin Nutr*. 2015;101(3):598–605.
- Kershaw KN, Albrecht SS. Metropolitan-level ethnic residential segregation, racial identity, and body mass index among U.S. Hispanic adults: a multilevel cross-sectional study. *BMC Public Health*. 2014;14:283.
- Jia P, Cheng X, Xue H, Wang Y. Applications of geographic information systems (GIS) data and methods in obesity-related research. *Obes Rev*. 2017;18:400–11.
- Ward ZJ, Long MW, Resch SC, et al. Redrawing the US obesity landscape: bias-corrected estimates of state-specific adult obesity prevalence. *PLoS ONE*. 2016;11(3):e0150735.
- Calvo R, Deterding S, Ryan R. Health surveillance during covid-19 pandemic. *BMJ*. 2020;369:m1373.
- Willis SJ, Cocoros NM, Randall LM, Ochoa AM, Haney G, Hsu KK, DeMaria A Jr, Klompas M. Electronic health record use in public health infectious disease surveillance, USA, 2018–2019. *Curr Infect Dis Rep*. 2019;21(10):32.
- Moon KA, Pollak J, Hirsch AG, Aucott JN, Nordberg C, Heaney CD, Schwartz CS. Epidemiology of Lyme disease in Pennsylvania 2006–2014 using electronic health records. *Ticks Tick Borne Dis*. 2019;10(2):241–50.
- Peterson KE, Hacek DM, Robicsek A, Thomson RB Jr, Peterson LR. Electronic surveillance for infectious disease trend analysis following a quality improvement intervention. *Infect Control Hosp Epidemiol*. 2012;33(8):790–5.
- Bernardo CO, González-Chica DA, Chilver M, Stocks N. Influenza-like illness in Australia: a comparison of general practice surveillance

system with electronic medical records. *Influenza Respir Viruses*. 2020;14(6):605–9.

21. Zozus MN, Richesson, R, Hammond WE, et al. Acquiring and Using Electronic Health Record Data. NIH Collaboratory: <https://rethinkingclinicaltrials.org/resources/acquiring-and-using-electronic-health-record-data/#bibliography>. Published on 2015. Accessed 2 April, 2020.
22. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*. 2011;4:47–55.
23. Leonardi C, Simonsen NR, Yu Q, Park C, Scribner RA. Street connectivity and obesity risk: evidence from electronic health records. *Am J Prev Med*. 2017;52(1S1):S40–S47.
24. Flood TL, Zhao YQ, Tomayko EJ, Tandias A, Carrel AL. Electronic health records and community health surveillance of childhood obesity. *Am J Prev Med*. 2015;48(2):234–40.
25. Baer HJ, Cho I, Walmer RA, Bain PA, Bates DW. Using electronic health records to address overweight and obesity: a systematic review. *Am J Prev Med*. 2013;45(4):494–500.
26. Census Bureau. American Community Survey 2016. <https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2015/>. Published in 2015. Accessed 04/21/2020.
27. Salinas JJ, Sheen J, Carlyle M, Shokar NK, Vazquez G, Murphy D, Alozie O. Using electronic medical record data to better understand obesity in hispanic neighborhoods in El Paso, Texas. *Int J Environ Res Public Health*. 2020;17(12):4591.
28. Guerrero A, Ponce N, Chung P. Obesogenic Dietary practices of Latino and Asian subgroups of children in California: an analysis of the California Health Interview Survey, 2007–2012. *Am J Public Health*. 2015;105:e105–12.
29. Salinas JJ, Rocha E, Abdelbary BE, Gay JL, Sexton K. Impact of Hispanic ethnic concentration and socioeconomic status on obesity prevalence in Texas counties. *IJERPH*. 2012;9(4):1201–15.
30. Park Y, Neckerman K, Quinn J, Weiss C, Jacobson J, Rundle A. Neighbourhood immigrant acculturation and diet among Hispanic female residents of New York City. *Public Health Nutr*. 2011;14(9):1593–600.
31. Sharifi M, Sequist TD, Rifas-Shiman SL, Melly SJ, Duncan DT, Horan CM, Smith RL, Marshall R, Taveras EM. The role of neighborhood characteristics and the built environment in understanding racial/ethnic disparities in childhood obesity. *Prev Med*. 2016;91:103–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

