Graphic Detail

Geographic Information Systems (GIS) organize and clarify the patterns of human activities on the Earth's surface and their interaction with each other. GIS data, in the form of maps, can quickly and powerfully convey relationships to policymakers and the public. This department of Cityscape includes maps that convey important housing or community development policy issues or solutions. If you have made such a map and are willing to share it in a future issue of Cityscape, please contact john.c.huggins@hud.gov.

Misalignment Between ZIP Codes and Municipal Boundaries: A Problem for Public Health

Richard C. Sadler
Assistant Professor, Division of Public Health, Michigan State University

Abstract

While useful for mail delivery, ZIP Codes are flawed as a geographic metric for public health research. This paper quantifies the magnitude of potential error inherent in using ZIP Codes as a unit of analysis in the state of Michigan. ZIP Codes are intersected with municipality boundaries in ArcGIS to determine the degree of misclassification. Results showed that 49 percent of the population had their municipality misclassified by their ZIP Code. This creates potentially huge errors when ZIP Code is the only geographic identifier, because actual exposure may vary from the exposure to which an individual is assigned based on ZIP Code. The Flint, Michigan, Water Crisis is a prime example of this error and the need to consider finer units of analysis whenever possible. Collaboration with experts in geographic information science is therefore essential for any public health research project where location is a factor.

Background

ZIP Codes are an arbitrary geographic designation assigned by the United States Postal Service for the purpose of delivering mail (Grubesic, 2008). As such, they do not correspond to political, social, or economic divisions, and have more heterogeneity than census units or neighborhood boundaries (Grubesic and Matisziw, 2006; Krieger et al., 2002). Even so, they are frequently used

uncritically as a unit of analysis in public health studies (Beyer, Schultz, and Rushton, 2008; Drewnowski, Rehm, and Solet, 2007; Gordon-Larsen et al., 2006; Acevedo-Garcia, 2001), owing to the ease with which this data can be gleaned from surveys or medical records.

Spatial error was front and center at the outset of the 2014 and ongoing Flint water crisis (Hanna-Attisha et al., 2016), when state officials initially downplayed the significance of an uptick in elevated blood lead levels among children because they were using ZIP Codes to define the exposed area rather than a more precise geographic metric (Sadler, 2016). In that particular case, using the city of Flint as the unit of analysis was much more accurate for defining the exposed population because the water system aligned almost exactly with the municipal boundary. Because of the possibility for error in other studies, the goal of this article is to quantify the magnitude of potential error inherent in using ZIP Codes as a unit of analysis.

Methods

Shapefiles containing Michigan's ZIP Codes (n = 987) and minor civil divisions, or municipalities (n = 1,517), were intersected in ArcGIS, a geographic information system, to determine the degree of overlap (or lack thereof) between the two. This operation led to the creation of 6,051 distinct ZIP Code-municipality pairs. The area of each region was calculated, and populations were assigned by joining each census block group centroid (n = 8,104) to the overlapping ZIP Code-municipality pair. For reference, 2,045 (34 percent) ZIP Code-municipality pairs were smaller than one-tenth of a square mile and 4,101 (68 percent) contained no census block group centroids, but all were included in the analysis.

Exhibit 1 illustrates the land area and population of three different "zones" (according to the type of overlap between ZIP Code and municipality). These zones include: (1) *match*: when the ZIP Code and municipality share the same name (whether they correspond to a city or another municipality); (2) *false match*: when the ZIP Code name is the same as that of two municipalities with similar names (for example, a ZIP Code named "Flint" covering "the City of Flint" and "Flint Township"); and (3) *non-match*: when the ZIP Code and municipality do not share a name (that is, the ZIP Code corresponds to another municipality).

Exhibit 1

Breakdown of Land Area and Population by ZIP Code-Municipality Match Type

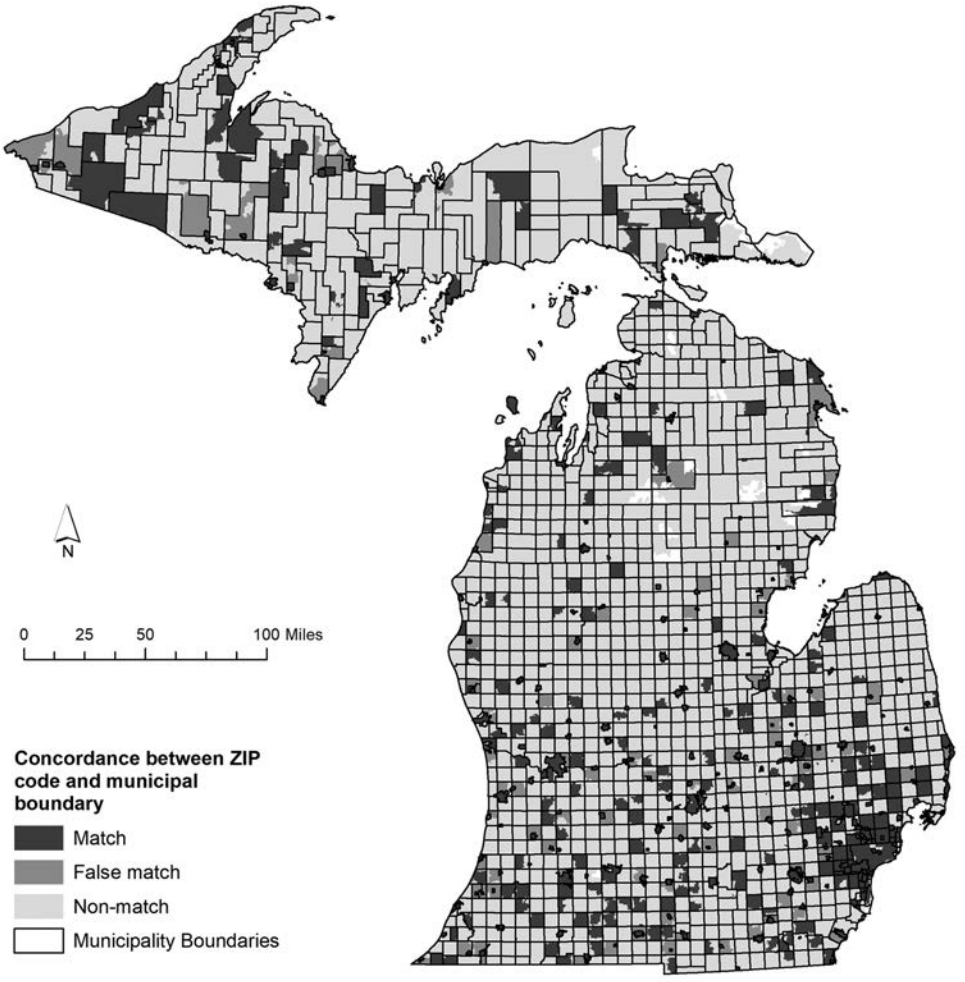
Match Type	n	Area (Square Miles)		Population	
Match	582	8,197	14%	5,071,822	51%
False Match	136	3,349	6%	679,177	7%
Non-Match	5333	45,958	80%	4,098,740	42%

Results

Exhibit 2 illustrates the match-non-match patterns spatially, with the "match" category shown in dark gray, the "non-match" category shown in light gray, and the "false match" category shown in medium gray. A quick examination reveals that most matching occurs in center cities throughout the state and townships, especially near Detroit in the southeast.

Exhibit 2

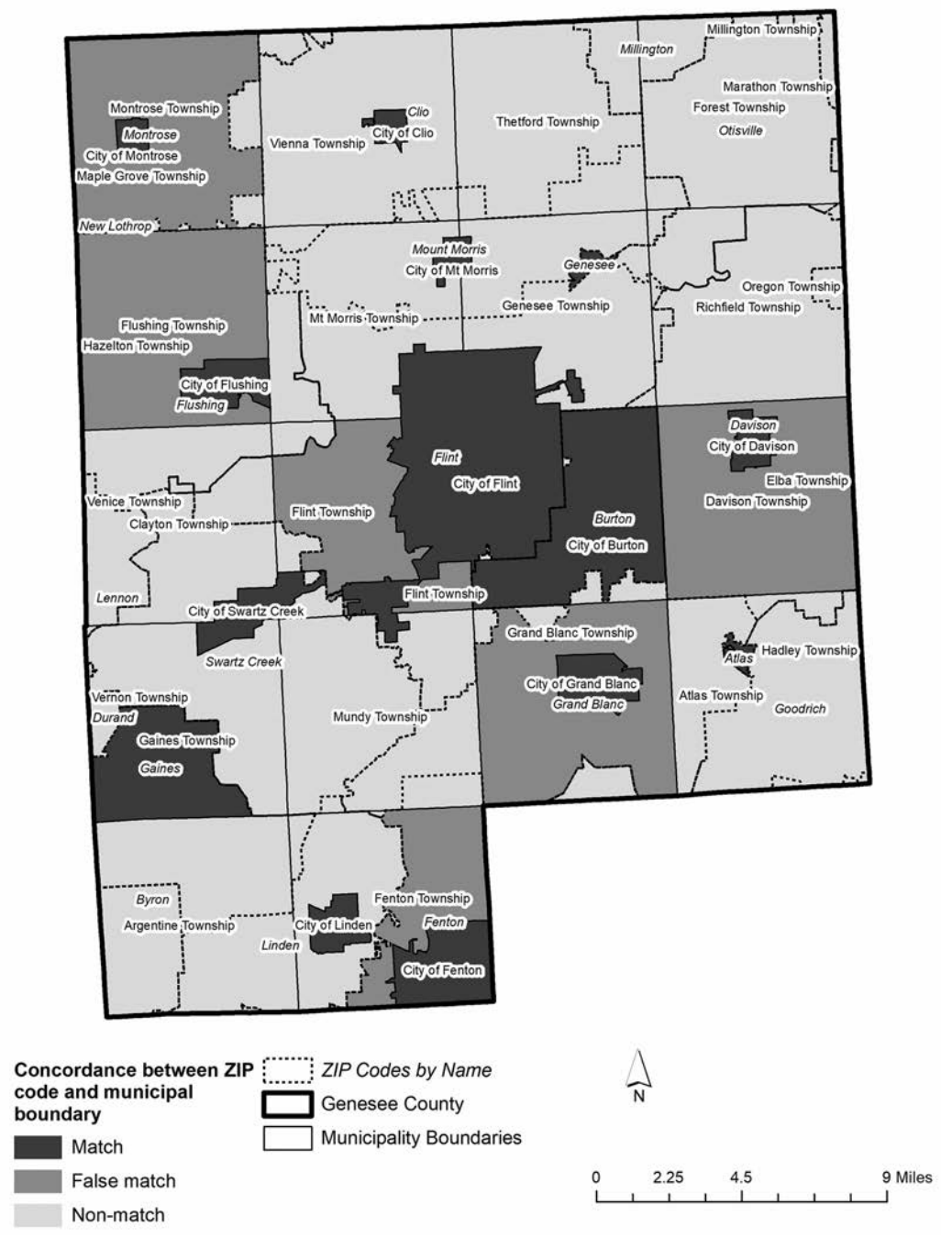
Concordance Between ZIP Code and Municipal Boundary Map Showing Michigan's ZIP Codes, Municipalities, and Overlap Type



The distribution of non- and false matches varies across the state. To give a better impression of this variation, a closer look at Genesee County (where Flint is located) is shown in exhibit 3. Municipalities are labeled in regular text, and corresponding ZIP Code names are labeled in italicized text. Note that matches are found in only a small amount of land area, and many false match areas exist.

Exhibit 3

Concordance Between ZIP Code and Municipal Boundary Map Showing Higher Degree of Overlap in Genesee County



Overall, only 14 percent of Michigan's land area matches ZIP Code and municipality name. While this small area includes 51 percent of the population, it means that 49 percent of the population is misrepresented in some way by their ZIP Code.

From a public health reporting perspective, the "false" and "non-match" categories are troublesome for different reasons. The "false match" category includes people living in townships with the same name and ZIP Code as the nearby center city (for example, Flint-Flint Township), which would lead many people to assume those residents live in the city if not specified. This category includes 6 percent of the state's land area and 7 percent of its population. Referencing exhibit 3 again, these false matches tend to occur where a township has the same name as the central city.

The non-match category poses similar problems—it contains 80 percent of the state's land areas and 42 percent of the population. When a non-match includes a reference to a center city, estimates for public health problems, such as a "city's" disease prevalence, crime incidence, or another phenomenon, could be skewed by including outlying areas. Many of these non-matches include rural areas where township names do not correspond to ZIP Code names. Although these rural non-matches would pose slightly fewer problems for public health reporting, it is worth raising this issue in general, because stakeholders and policymakers need to be aware of this source of error.

Conclusions

The importance of this misclassification to public health should not be forgotten: while ZIP Codes can absolutely be useful tools for public health planning and policymaking, they can misrepresent health statistics when looking at phenomena that may not coincide well with ZIP Code boundaries (as described in Sadler and Lafreniere, 2017), including those related to municipal services.

The example of the Flint water crisis—and the results of this article showing how misaligned ZIP Codes are from municipal boundaries—suggest that geographers, epidemiologists, and others with expertise in geographic information science should be closely consulted on any public health research where the location may play an important determining factor.

Acknowledgments

This publication was supported by a grant from the National Institute on Minority Health and Health Disparities under award number U54MD011227.

Author

Richard C. Sadler is an Assistant Professor in the Division of Public Health at Michigan State University. He can be contacted via phone at (810) 600-5674 or via e-mail at sadlerr@msu.edu.

References

Acevedo-Garcia, Dolores. 2001. "ZIP Code-Level Risk Factors for Tuberculosis: Neighborhood Environment and Residential Segregation in New Jersey, 1985-1992," *American Journal of Public Health* 91 (5): 734.

Beyer, Kirsten M., Alan. F. Schultz, and Gerald Rushton. 2008. *Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice*, edited by G. Rushton, M. P. Armstrong, J. Gittler, B. R. Greene, C. E. Pavlik, M.M. West, and D. L. Zimmerman. Boca Raton, FL: CRC Press.

Drewnowski, Adam, Colin D. Rehm, and David Solet. 2007. "Disparities in Obesity Rates: Analysis by ZIP Code Area," *Social Science & Medicine* 65 (12): 2458–2463.

Gordon-Larsen, Penny, Melissa C. Nelson, Phil Page, Barry M. Popkin. 2006. "Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity," *Pediatrics* 117 (2): 417–424.

Grubesic, Tony H. 2008. "ZIP Codes and Spatial Analysis: Problems and Prospects," *Socio-economic Planning Sciences* 42 (2): 129–149.

Grubesic, Tony H., and Timothy C. Matisziw. 2006. "On the Use of ZIP Codes and ZIP Code Tabulation Areas (ZCTAs) for the Spatial Analysis of Epidemiological Data," *International Journal of Health Geographics* 5 (1): 58.

Hanna-Attisha, Mona, Jenny LaChance, Richard Casey Sadler, and Allison Champney Schnepp. 2016. "Elevated Blood Lead Levels in Children Associated with the Flint Drinking Water Crisis: A Spatial Analysis of Risk and Public Health Response," *American Journal of Public Health* 106 (2): 283–290.

Krieger, Nancy, Pamela Waterman, Jarvis T. Chen, Mah-Jabeen Soobader, Subu V. Subramanian, and Rosa Carson. 2002. "ZIP Code Caveat: Bias Due to Spatiotemporal Mismatches Between ZIP Codes and U.S. Census–Defined Geographic Areas—The Public Health Disparities Geocoding Project," *American Journal of Public Health* 92 (7): 1100–1102.

Sadler, Richard Casey. 2016. "How ZIP Codes Nearly Masked the Lead Problem in Flint," *The Conversation*. http://theconversation.com/how-zip-codes-nearly-masked-the-lead-problem-in-flint-65626.

Sadler, Richard Casey, and Don J. Lafreniere. 2017. "You Are Where You Live: Methodological Challenges to Measuring Children's Exposure to Hazards," *Journal of Children and Poverty* 23 (2): 189–198.