

Who's missing?

Assessing geographic bias in understanding low-acuity frequent utilizers of the Emergency Department

Caitlyn Grubbs¹, Jacqueline Curtis, PhD², Soropat Vijitakula, MD, MPH³, Andrew Curtis, PhD²



View presentation

¹ Case Western Reserve University, Master of Public Health Program | ² Case Western Reserve University, Department of Population & Quantitative Health Sciences | ³ University Hospitals, Department of Emergency Medicine

Background

This project is part of a larger on-going study with University Hospitals Cleveland Medical Center looking at low acuity high utilization of the emergency department (ED). Despite the growing number of studies investigating geographic patterns of frequent ED utilization, there has been a lack of awareness of the potential geographic bias in results, especially due to geocoding error. This study identifies the geographic bias that may exist within this population and may affect any analysis that may examine geographic influence.

Population

The study population consisted of patients 18 years and older who accessed the emergency department of University Hospitals for health care between January 1, 2018 through December 31, 2020 and were noted as having a low acuity which is defined as having an Emergency Severity Index (ESI) of 3V, 4, or 5). The ESI is a five-level emergency department triage algorithm.

ESI Level	Basic Definition
1	Patient requires immediate life-saving intervention
2	Patient is in a high-risk situation, is disoriented, in severe pain or vitals are in danger zone
3	If multiple resources are required to stabilize the patient, but vitals are not in danger zone
4	If one resource is required to stabilize the patient
5	If patient does not require any resources to be stabilized

Table 1. Definitions of ESI Levels.

This was further narrowed to encounters of patients who had 2 or more encounters within the study period. Those addresses for each encounter were assigned "Matched" if they geocoded to their point level location or "Unmatched" if the addresses were not geocoded to their point level location. Therefore, the study population is 17,774 (Figure 1).

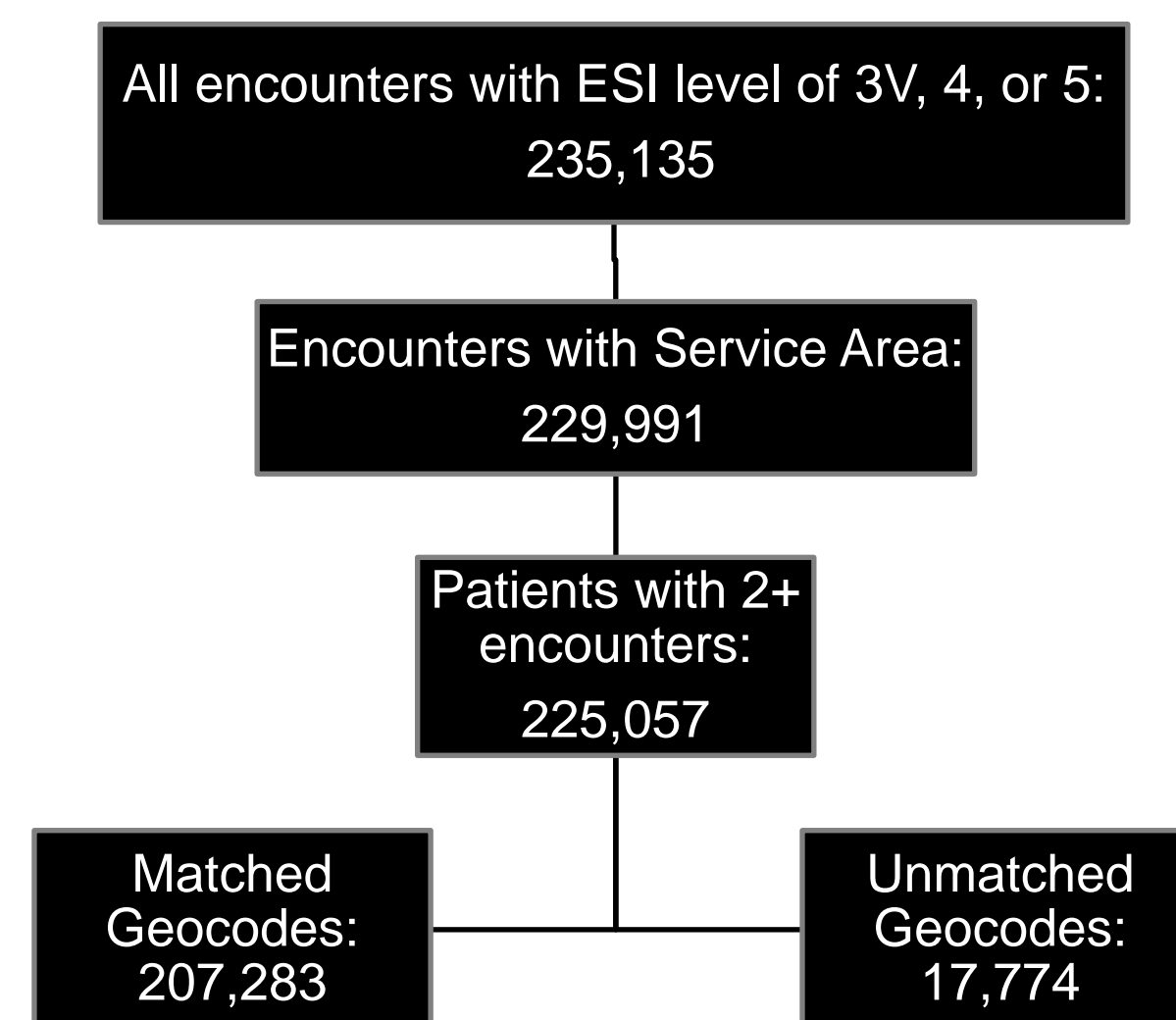


Figure 1. Study population flowchart.

Learning Objectives

- Understand the geographic bias that exists when using a geocoder
- Examine the geographic and demographic characteristics of the population that frequently utilized the ED for non-emergent reasons
- Identify if sources of geographic bias exists within the ED health care data
- Create a systematic method to identify and assess for geographic bias within a health dataset

Activities

- Conduct a literature review regarding high utilization of the ED
- Conduct a literature review of geographic bias in research
- Create maps utilizing a Geographic Information System (GIS) to examine the geographic bias
- Clean data to better examine patterns within the GIS

Methods

- Zip Codes within the study area were converted to Zip Code Tabulation Areas (ZCTA) using a crosswalk file.
- The proportion of unmatched addresses were calculated for each Zip Code. These proportions by ZCTA were mapped using a choropleth map using ZIP Code Tabulation Area Cartographic Boundary File from the Census Bureau (Figure 2).
- The number of encounters that were 10 or less total per ZCTA were mapped as well to identify ZCTA s that were identified as high proportion of unmatched geocodes, but also have small number of total encounters. These are not shown to protect patient privacy.
- Lastly, the addresses were examined for reasons why they would be unmatched in a geocoder to identify if patterns or systematic reasons exist.

Results

- Successful geocodes were available for 207,283 (92.1%) of encounters for patients with 2 or more visits within the study period. Most of the Unmatched encounters are still seen in rural parts of the study area (Figure 2)

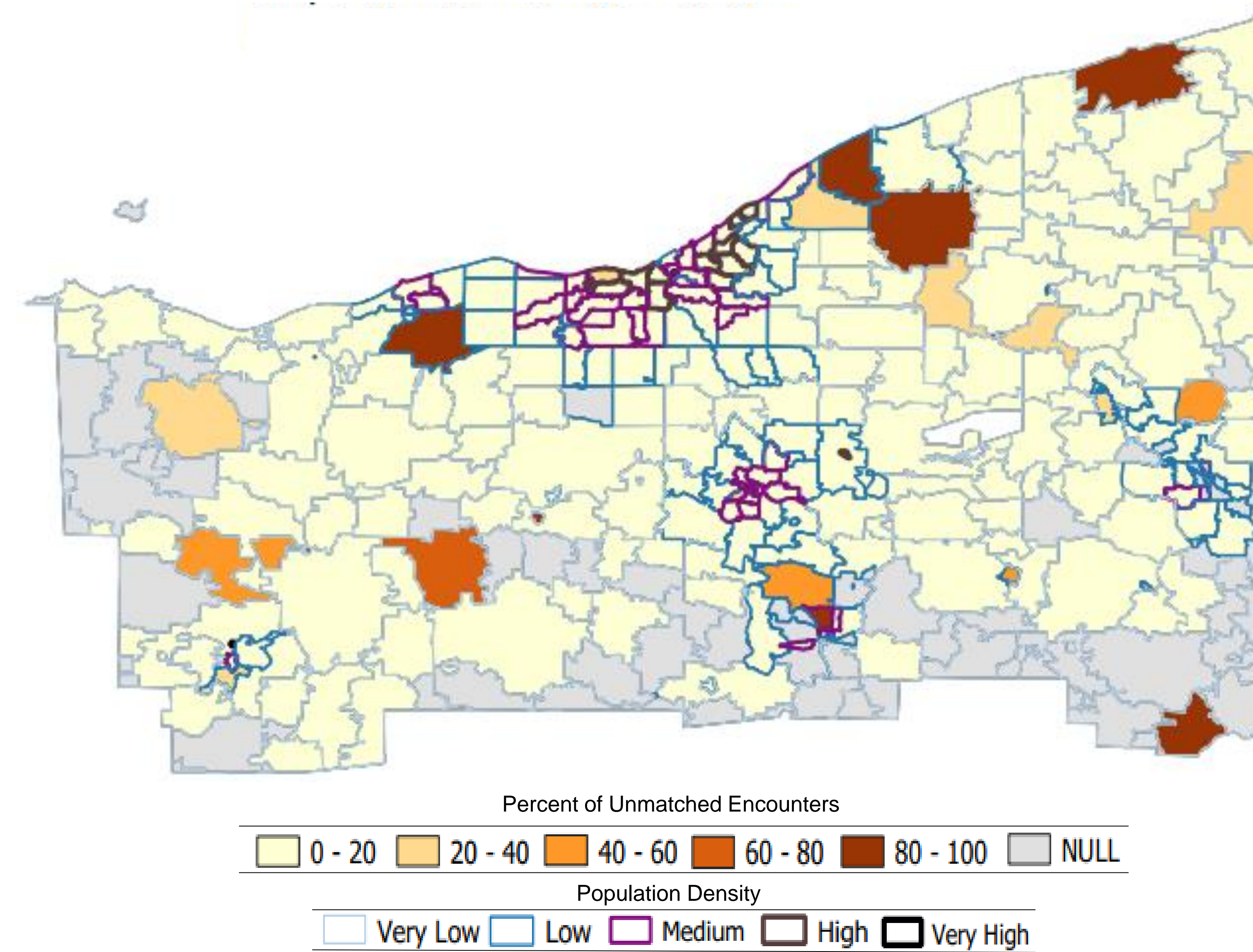


Figure 2. Population density and proportion unmatched for each ZCTA in the study area.

- Unmatched address include misspelling of city, use of post office box as provided by the patient, as well as unverifiable mailing addresses ("Bad Address") and homelessness designations (Table 2). However, there is no systematic, consistent reasons for exclusion of the addresses within the patient population.

Types of Addresses Unmatched	Count
Apartments	5220
P.O. Box	2103
Unknown/ "Bad Address"	426
Homeless	60
General Delivery	9
Misspelling of cities	1809
Other	8147

Table 2. Reasons for Unmatched geocodes.

Deliverables

- Maps indicating underrepresentation of non-emergency frequent ED use due to unverifiable patient addresses.
- A report showing the geographic bias in healthcare data
- Recommendations for future health geographic bias reviews

Lessons Learned

- Cleaning and manipulation of data is much more time consuming than actual map creation
- How to utilize QGIS, software used and accesses by UH for this project
- Creation of choropleth maps using census data and cartographic boundary files
- There are various reasons an address may not geocode; these are not consistent in datasets and require deep digging in the data to determine the prevalence and impact on the overall dataset

Public Health Implications

- There is a lack of input consistency when addresses are obtained within healthcare settings
- The proportion of unmatched geocodes are heavily representative of the rural populations. These individuals may be under-represented in geographic ED healthcare research due absence of verifiable data when working with geocoded addresses. This is seen in previous studies as well.
- As GIS is more frequently used to analyzed geographic impacts of health, knowing and understanding geographic bias that exists within health data is essential to accurately examining the geographic impact of public health.

Recommendations

- Guidelines for reducing inconsistency in address input to improve overall geocoding quality of health data:
- For counties, cities, and Zip Codes, create a drop-down autofill template to prevent the need for abbreviations and misspellings of commonly used inputs.
- For Apartment and other multi level addresses, have a predesignated fill box so these second lines can be easily removed prior to geocoding
- For PO Boxes, have a mailing address option and a resident address option. Since PO Boxes do not ever geocode nor correspond to a physical home address.

References

Baldovin, T., Zangrando, D., Casale, P., Ferrarese, F., Bertinello, C., Buja, A., Marcolongo, A., & Baldo, V. (2015). Geocoding health data with Geographic Information Systems: a pilot study in northeast Italy for developing a standardized data-acquiring format. *Journal of preventive medicine and hygiene*, 56(2), E88–E94.

Krieger N, Waterman PD, Chen JT, Soobader MJ, Subramanian SV, Carson R. Zip code caveat: Bias due to spatiotemporal mismatches between zip codes and U.S. census-defined geographic areas -- The Public Health Disparities Geocoding Project. *Am J Public Health*. 2002;92:1100–1102.

Oliver, M. N., Matthews, K. A., Siadaty, M., Hauck, F. R., & Pickle, L. W. (2005). Geographic bias related to geocoding in epidemiologic studies. *International Journal of Health Geographics*, 4(1), 29.