The Potential and Pitfalls of Geocoding Electronic Health Records

William R. Buckingham, PhD

ABSTRACT

Background: Geocoding electronic health records (EHRs) provides novel insights for clinicians, but it is important to understand and address key issues, including privacy and protection of patient records, in order to realize potential benefits.

Methods: This paper discusses the issues surrounding geocoding and illustrates potential benefits through 3 case studies of no-shows to clinical appointments, patient analysis for a merged clinic site, and multi-clinic patient overlap.

Conclusion: Geocoding EHRs provides a new contextual understanding for clinicians to understand patients and provide targeted interventions that patients can implement. While geocoding EHRs presents a need for high data security, the benefits outweigh the risks when proper protections are observed.

INTRODUCTION

Geocoding electronic health records (EHRs) offers novel and exciting benefits that allow clinicians and researchers to develop a place-based understanding of a patient's health environment as well as the assets and obstacles that are present for each patient. This understanding can allow the clinician to provide advice that can be directly implemented when it comes to chronic conditions such as obesity, asthma, and diabetes. Similarly, geocoded EHRs can allow clinicians to partner with public health officials to monitor infectious diseases such as influenza, a current focus for many health officials in light of the H1N1 scare of 2009. By geocoding EHR data, geographic analysis of health becomes possible at scales that are meaningful to both patient and physician. This paper will discuss the issues that surround geocoding EHRs, including the privacy protections that are a must for work of this nature. Subsequently, the paper suggests methods for handling geocoded data, both

• • •

Author Affiliations: Health Geographer, University of Wisconsin-Madison Applied Population Laboratory; PhD Candidate, Nelson Institute for Environmental Studies, UW-Madison.

Corresponding Author: William R. Buckingham, UW-Madison, 316D Agriculture Hall, 1450 Linden Dr, Madison, WI 53716; phone 608.262.9156; fax 608.262.6022; e-mail wrbuckin@wisc.edu.

in public presentation and research. The final section highlights 3 case studies to demonstrate the potential benefits when using geocoded EHRs,, and considers areas for expansion and improvement of the process.

GEOCODING BACKGROUND AND PITFALLS

Geocoding of health records has a history dating back decades, especially in public health-related endeavors. Vital records from state and local public health offices have been used to analyze birth data and birth disparities,[1,2] evaluating differences in gender, race, and inequality,[3-6] as well as general research practice in public heath.[7-9] However, these efforts have focused largely in the arena of vital records for the purpose of public health understanding. The use of geocoded EHRs in medical research has been largely absent. Nevertheless, the utility of EHRs to provide both context and depth to understanding the clinical population is encouraging. It is critical at this juncture to step back and define precisely what geocoding is and to discuss the issues surrounding geocoding.

EHRs present a complex use case when it comes to geo coding. While vital records have the ability to pinpoint a person at an address, it is common for identifying information such as the name of the person to be removed from the vital record. With EHRs, however, names as well as in-depth medical information are often a part of the record, making the records highly sensitive.

The actual geocoding procedure involves utilizing the address and a zone delimiter (often a ZIP code) to interpolate the location of the record on a street segment in a geographic information systems (GIS) database and place a point on the correct side of the street. It is common to use multiple geo coding engines to cross-validate the data points and to capture locations that may not be identifiable with a single dataset. Once the points are geocoded, the researcher often undertakes 2 basic tasks. The first is to append contextual data, often
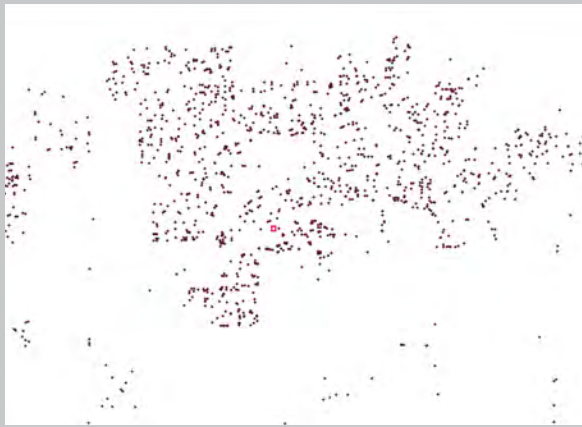
Figure 1. Geocoded patient points with no identifying geographic feature to protect privacy.
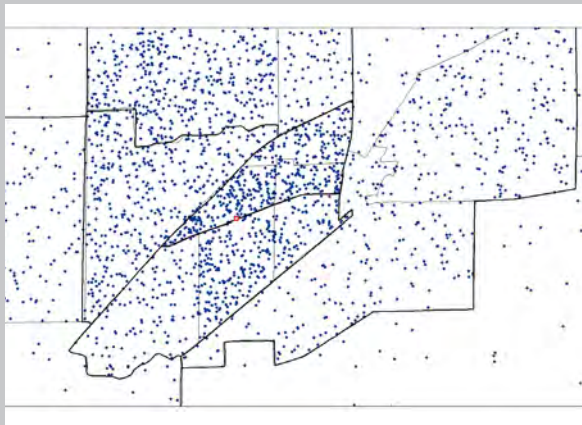
Figure 2. Random patient point locations with census block (thin lines) and census block group (thick lines) boundaries.
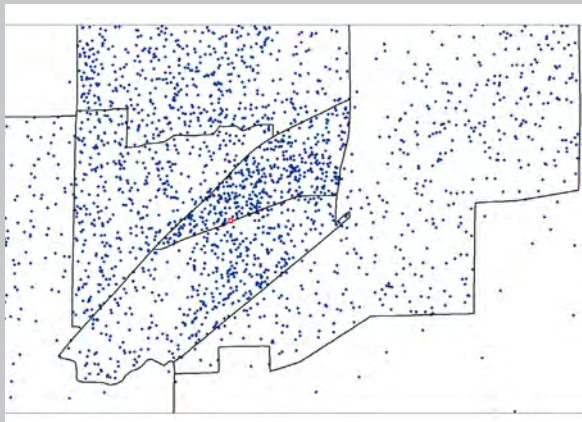
Figure 3. Random patient point locations with census block group boundaries.

from the US Census Bureau. The second is to create maps that enable the visualization of both the population distribution and the underlying contextual association. Both of these steps involve privacy concerns that the researchers must address.

Connecting detailed location with medical history requires a strong security regime. In the majority of cases, researchers separate the location information from the medical history and maintain 2 files on separate systems to protect confidentiality. A limited file provides the basis for the geocoding information and the census block group is appended for reconnection to the clinical data. The census block group is chosen because the US Census Bureau has defined the block group as the lowest unit of analysis available with non-physically identifying features published (eg, data about income or education as opposed to data about gender or age). Steps are then taken to ensure that the point locations are not accessible when publishing maps that require a connection between the patient location and the medical condition of interest. To accomplish this, a 3-step process is put into place. First, the geocoded records are aggregated to the census block (Figure 1). This is done to ensure a general distribution correctness (ie, avoiding areas such as lakes where people obviously do not live). Second, the block-level data are presented at random point locations within the block (Figure 2). This randomization removes the strict point location correlation with a person and begins the process of masking the actual location of the patient. Finally, the block boundaries are completely removed, any street data also is removed, and only the block group boundaries are presented (Figure 3). This masking maintains some geographic correctness but reduces the potential that someone could pinpoint a patient. This process is used only when point representation is critical on the map. In most other instances, the geocoded data are summarized into block or block group totals and these totals are presented on a choropleth map providing no point location issues.

By disassociating the location and medical condition information in the EHR, patient confidentiality is maintained while allowing for the geographic context to be brought to bear on the question at hand. For presenting and visualizing data, this disassociation is not possible; however, following strict masking processes as described above allows the researcher to overcome the privacy concerns and protect patient confidentiality.
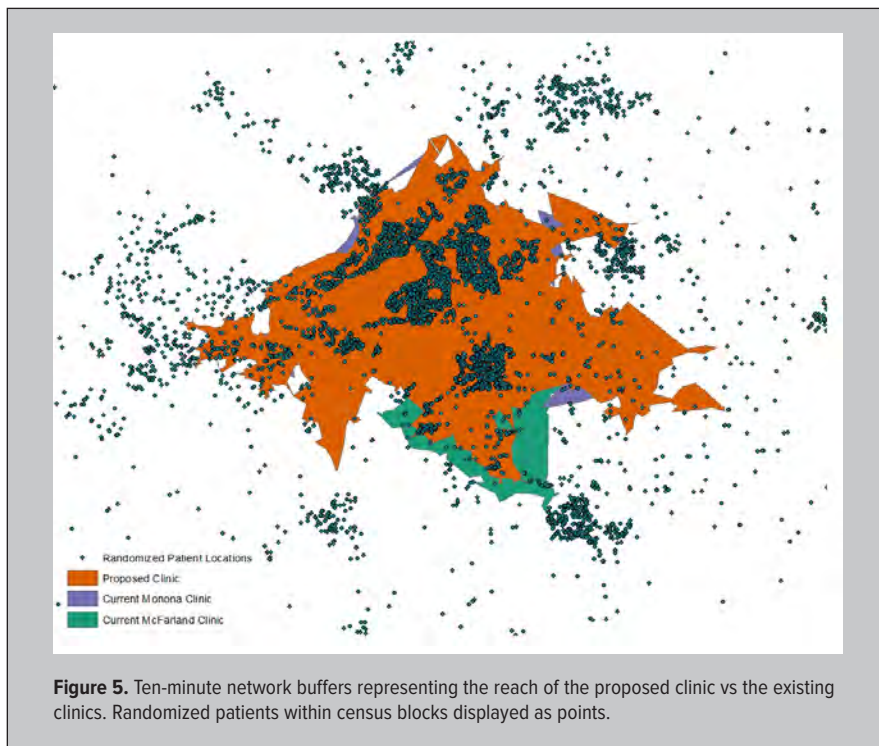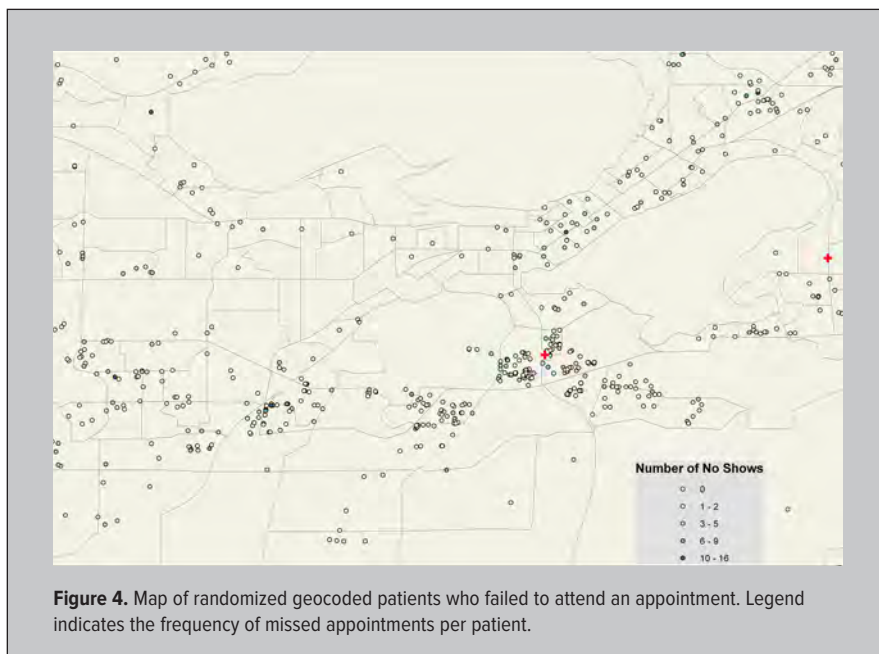
CASE STUDIES WITH GEOCODED EHRs

Using the privacy preserving methods mentioned above, 3 case studies are described below demonstrate how the use of geocoded EHRs can be used to improve both service delivery and contextual understanding of the primary care doctor in treating patients.

Understanding Where Patients (Don't) Come From

At the Wingra Family Medical Center in Madison, Wisconsin, the occurrence of missed appointments is a daily issue. Missed appointments influence not only the patient's health, but also the clinic's ability to serve the patient population. While the effects of missed appointments are recognized, understanding how to combat missed appointments is difficult.

It was from this jumping-off point that the University of Wisconsin-Madison Department of Family Medicine and the Applied Population Laboratory began an analysis of no-shows to appointments at the Wingra clinic. EHRs were used to pull a set of records detailing only the address of the patient and the number of times the patient missed an appointment. By reducing the necessary information to just these 2 components, the privacy of the individuals was maintained. The EHRs were geocoded and the locations were randomized within block groups. The resulting points were then coded based on the number of times an individual record failed to appear for an appointment. The classification for the resulting color scheme was created using ArcGIS software[10] and applying a modified Jenk's Natural Breaks classification scheme.

Two immediate trends were noticeable from the resulting map (Figure 4). First, a cluster of no-shows appeared in the 3-block groups immediately adjacent to the clinic. This may be in part due to a self-selection by residents in these areas to pursue health care at the Wingra clinic. Second, the spread of patients was city-wide (the dataset was restricted to Madison). However, areas in southwest Madison and the northern sections of the city of Fitchburg appeared as areas of a high preponderance of missed visits. These findings were presented to the faculty at the Wingra clinic, where the intent was to develop strategies to facilitate better attendance at appointments, either through transportation arrangements or scheduling changes. In this case, the geocoded records not only inform the location of the no-shows but also clarify potential solutions based on geography.

Figure 4. Map of randomized geocoded patients who failed to attend an appointment. Legend indicates the frequency of missed appointments per patient.

Figure 5. Ten-minute network buffers representing the reach of the proposed clinic vs the existing clinics. Randomized patients within census blocks displayed as points.

Using EHRs to Site a Merging Clinic

In 2009, 2 clinics in suburban Madison within the UW Health network were targeted for merging into a single location. Again, the use of geocoded EHRs was brought to bear, this time to assess patient access. Patient lists from both clinics were
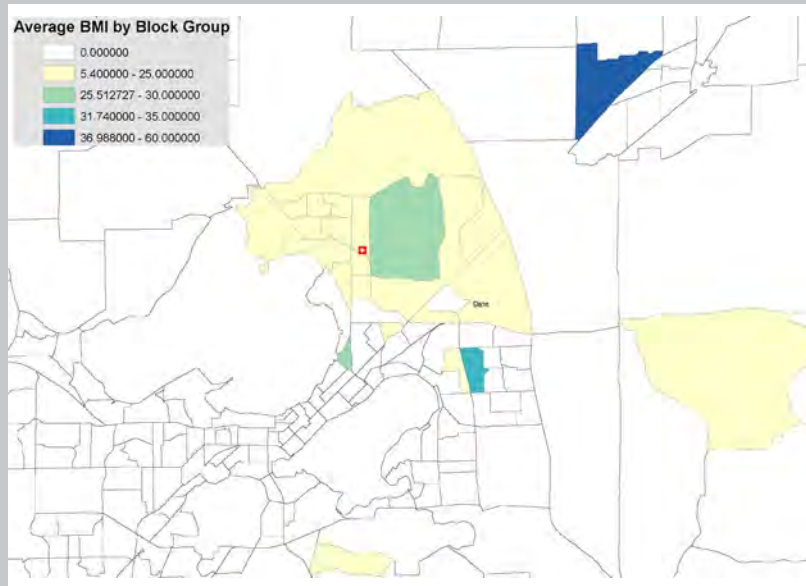
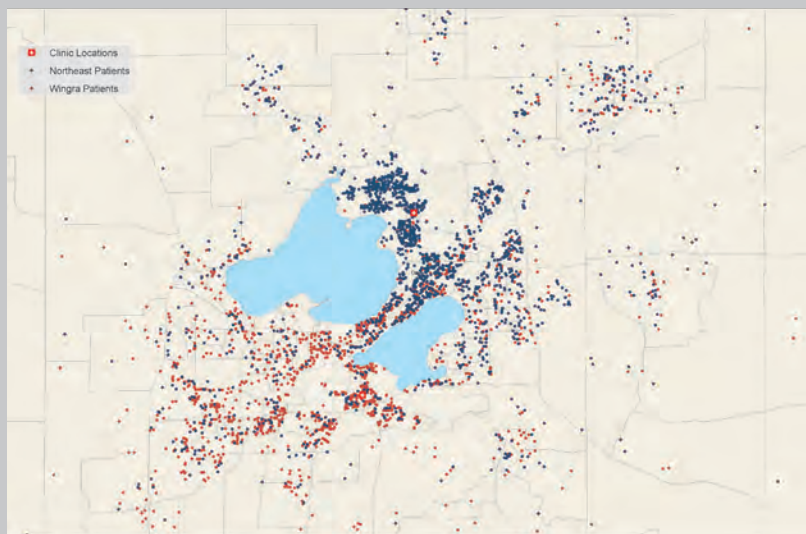**Figure 6.** Average BMI value by block group based on aggregation of geocoded patient records.

Figure 7. Map illustrating the overlap between 2 clinical populations within the Madison, Wisconsin metropolitan area.

each of the clinics to evaluate the potential gain or loss based on the proposed location (Figure 5). In discussions, there had been some concern that walkability and neighborhood ties would be lost with the new clinic site. And indeed, within a 5-minute drive more patients were near the old clinics than the new site. However, within a 10-minute drive, the new clinic site captured a greater volume of patients than the other 2 sites combined. These findings were presented at a meeting with the majority of clinicians from each site to allow for discussion and to illustrate the benefits and drawbacks of the proposal. Ultimately, the project went forward and the merged clinic, now known as the Yahara Clinic, was opened in 2011 in roughly the location proposed in the initial drive-time analysis.

Describing the Population of Clinics Within a Provider Network

The final case study centers on the use of geocoded EHRs to help assess the distribution of clinical diagnoses within the clinic population for the purpose of developing actionable recommendations for the affected patients. Geocoding and mapping the distribution and prevalence of clinical data such as high A1C values, high BMI, and location of diabetic patients was a first step. Once the data were mapped, family physicians at the clinic could evaluate where high values of these conditions exist and begin to develop intervention schemes to offer solutions to these problems. For example, the high average body mass index (BMI) values in conjunction with the clinical population distribution allowed clinicians to understand where exercise opportunities may or may not exist. Also, data may be sorted by race and ethnicity information obtained through initial patient registration.

In Figure 6, the green-shaded area represents high average BMI values for the aggregated patient population. Both of these areas are somewhat isolated due to either industrial features or high economic, social, or racial contrasts between neighborhoods. These observations provided the clinicians a place-based understanding of this issue and allowed them to begin seeking

geocoded and placed on a networked road dataset, although these images were never published even to ephemeral computer images. The existing clinic locations also were placed on the road network, as was a hypothetical location in the vicinity of a possible new clinic. Each of the clinics—the 2 existing locations and the potential new site—were then analyzed on the road network to construct 5-, 10- and 20-minute drive shapes from the clinics. Again, researchers used ArcGIS to conduct this network analysis and produce the 9 drive-time areas.

These shapes were then intersected with the patients from

local opportunities for patients to combat isolation and poor BMIs with geographically targeted programming.

Joining clinical data from multiple locations provides the benefit of a more complete geographic picture of the patient and health landscape. Geocoding the patients in a complete network and mapping separate clinics together (Figure 7) makes it apparent how difficult it is for clinicians to account for geography in a clinical setting. Interventions must be planned at a health system level to be effective for an area, as patients overlap considerably yet visit different clinics, where 2 different recommendations are possible despite identical geographic conditions.

CONCLUSION

None of the case studies illustrated above provides a complete assessment of the effect of the geocoded EHR. In each case ,the collaboration ended once the data were presented, and the ultimate use of these data was not reported. Unsatisfying as this may be, the use of geocoded EHRs is encouraging for analysis of factors ranging from clinic siting to geographic barriers to healthy lifestyles. While this type of analysis is more prevalent in the public health sphere than in the clinical arena, the ability to understand the geographic constraints on a patient may help a physician prescribe a more effective means of intervention in a given diagnosis. Geocoding EHRs is not without challenge—privacy is a paramount concern that requires vigilance from both researcher and clinician at all times. However, the potential benefit to the patient outweighs the risks, so long as

good custodianship is practiced. Providing a clinician with a spatial perspective can lead to better service delivery and a better prescription for combating chronic and infectious disease.

REFERENCES

1. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health.* 2006;83(6):1041-1062.

2. O'Campo P, Xue X, Wang M, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. *Am J Public Health.* 1997;87(7):1113-1118.

3. Krieger N. Putting health inequities on the map: social epidemiology meets medical/health geography——an ecosocial perspective. *GeoJournal.* 2009;74(2):87-97.

4. Krieger N, Chen JT, Waterman P, Rehkopf DH, Subramanian SV. Race/ethnicity, gender and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures—the public health disparities geocoding project. *Am J Public Health.* 2003;93(10):1655-1671.

5. Krieger N, Waterman P, Lemieux K, Zierler S, Hogan JW. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health.* 2001;91(7):1114-1116.

6. McLafferty S. Immigrant Reproductive Health Disparities: A GIS Analysis [presentation]. HRSA Maternal and Child Health DataSpeak Web Conference; 2008 Feb 20. http://mchb.hrsa. gov/researchdata/mchirc/dataspeak/pastevent/february2008/resources/index.html. Accessed April 27, 2012.

7. Rushton G, Elmes G, McMaster R. Consideration for improving geographic information systems research in public health. *URISA Journal.* 2000;12(2):31-49.

8. Cromley EK, McLafferty S. *GIS and Public Health.* New York: Guilford Press, 2002.

9. O'Carroll PW. Introduction to public health informatics. In: O'Carroll PW, Yasnoff WA, Ward ME, Ripp LH, Martin EL, eds. *Public Health Informatics and Information Systems.* New York: Springer-Verlag, 2003.

10. ArcGIS [computer program]. Version 10.0. Redlands, CA: Esri; 2010.

advancing the art & science of medicine in the midwest

WMJ

Visit www.wmjonline.org to learn more.