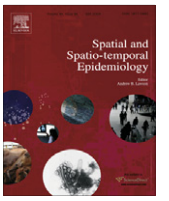
Spatial and Spatio-temporal Epidemiology 3 (2012) 31–38

Contents lists available at SciVerse ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste



Spatial error in geocoding physician location data from the AMA Physician Masterfile: Implications for spatial accessibility analysis

Sara McLafferty [a,*], Vincent L. Freeman [b,e], Richard E. Barrett [c], Lan Luo [d], Alisa Shockley [a]

[a] Department of Geography, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[b] Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, USA
[c] Department of Sociology, University of Illinois at Chicago, USA
[d] Blue Health Intelligence, Chicago, IL, USA
[e] Population Health, Behavior and Outcomes Program, Cancer Center, University of Illinois at Chicago, USA

ARTICLE INFO

Article history:
Available online 15 February 2012

Keywords:
Physicians
Geocoding
Accessibility
AMA Masterfile

ABSTRACT

The accuracy of geocoding hinges on the quality of address information that serves as input to the geocoding process; however errors associated with poor address quality are rarely studied. This paper examines spatial errors that arise due to incorrect address information with respect to physician location data in the United States. Studies of spatial accessibility to physicians in the U.S. typically rely on data from the American Medical Association's Physician Masterfile. These data are problematic because a substantial proportion of physicians only report a mailing address, which is often the physician's home (residential) location, rather than the address for the location where health care is provided. The incorrect geocoding of physicians' practice locations based on inappropriate address information results in a form of geocoding error that has not been widely analyzed. Using data for the Chicago metropolitan region, we analyze the extent and implications of geocoding error for measurement of spatial accessibility to primary care physicians. We geocode the locations of primary care physicians based on mailing addresses and office addresses. The spatial mismatch between the two is computed at the county, zip code and point location scales. Although mailing and office address locations are quite close for many physicians, they are far apart (>20 km) for a substantial minority. Kernel density estimation is used to characterize the spatial distribution of physicians based on office and mailing addresses and to identify areas of high spatial mismatch between the two. Errors are socially and geographically uneven, resulting in overestimation of physician supply in some high-income suburban communities, and underestimation in certain central city locations where health facilities are concentrated. The resulting errors affect local measures of spatial accessibility to primary care, biasing statistical analyses of the associations between spatial access to care and health outcomes.

1. Introduction

Geographic inequality in access to physicians is a significant public health concern in the United States. Research shows wide disparities in the supply and availability of physicians between rural and urban areas (Ricketts and Randolph, 2007), as well as inequalities within urban and metropolitan regions. Mapping and analyzing the uneven geographical distribution of physicians requires accurate, geocoded information about physicians' practice locations – the sites where health care services are provided. This article examines spatial errors that arise due to incorrect address information with respect to physician location data in the United States. Studies of spatial accessibility to physicians in the U.S. typically rely on data from

* Corresponding author.
E-mail addresses: smclaff@illinois.edu (S. McLafferty), Freem981@uic.edu (V.L. Freeman), barrett@uic.edu (R.E. Barrett), lan.luo@bluehealthintelligence.com (L. Luo), shockle2@illinois.edu (A. Shockley).

1877-5845/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sste.2012.02.004

the American Medical Association (AMA)'s Physician Masterfile. These data are problematic because a substantial proportion of physicians only report a mailing address, which is often the physician's home (residential) location, rather than the address of the location where health care is provided. The incorrect geocoding of physicians' practice locations based on inappropriate address information results in a form of geocoding error that has not been widely analyzed. Such error leads to inaccuracies in characterizing the spatial distribution of physicians and people's access to care; yet, the magnitude and characteristics of this error are unknown. We address these issues by analyzing geographic and socioeconomic patterns of error in primary care physician location data from the AMA Masterfile for the Chicago metropolitan region.

2. Background

A large body of research has investigated the spatial distribution of physicians in the U.S. at the local, regional and national scales and examined the impacts of inequalities in spatial access to physicians on health and well-being. Researchers have documented significant disparities in physician supply among states and regions (Macinko et al., 2007; Shipman et al., 2011). They have also documented significant inequalities in supply between rural and urban areas and shown that these inequalities often place rural residents at a disadvantage in access to care. (Calman et al., 2006; Grumbach et al., 2003; King et al., 2009; Ricketts and Randolph, 2007; Weeks and Wallace, 2008). Research also shows wide gaps in availability within cities, with shortages of services in impoverished, inner city neighborhoods (Guagliardo et al., 2004). In these types of studies, researchers have also developed innovative measures of local spatial accessibility to physicians (Luo and Wang, 2003; Luo and Qi, 2009) and used these measures in investigating the associations between spatial access to physicians and health outcomes. For example, the relationship of local primary care physician supply with mortality and incidence of preventable health conditions has been analyzed for Medicare beneficiaries in the U.S. (Chang et al., 2010). Studies have also investigated whether or not poor spatial access to primary care physicians is linked to high rates of late-stage diagnosis for certain cancers (Wang et al., 2008; Dai 2010; McLafferty and Wang, 2009), and low utilization of preventive health services (Continelli et al., 2010). Summarizing the results of numerous studies of the relationship between primary care physician supply and health outcomes, Macinko et al. (2007) concluded that increasing the supply of primary care doctors would significantly improve health outcomes for the U.S. population.

Studies such as these require accurate, geocoded data on physician practice locations to measure geographic variation in physician supply. A key source of data for these types of investigations is the AMA's Physician Masterfile, a national-level data set representing virtually all Doctors of Medicine (M.D.s) in the U.S., including medical residents and foreign medical graduates. These data have been widely used in investigating geographic inequalities in physician supply and the associations with health care costs and health outcomes (Chang et al., 2010; Chen et al., 2010; Luo and Wang, 2003; Peterson et al., 2011; Ricketts and Belsky, 2012; Wang et al., 2009). The data also underpin important health care policy issues including the determination of health professional shortage areas and medically underserved areas (http://datawarehouse.hrsa.gov/pcsa2006.aspx?layout=print). For each physician, the Masterfile contains information on demographic characteristics, specialization, and practice type, as well as two types of address data: the preferred mailing address and the office address. Every record contains a preferred mailing address; however a fraction of records, typically between 10% and 25%, do not include an office address (Grumbach et al., 2003). In geocoding the AMA data, most researchers first use the office address, and then, for those lacking an office address, geocode based on the mailing address (Chang et al., 2010). In cases where the mailing address and office location do not correspond, this procedure results in a form of geocoding error, insofar as the geocoded location does not represent the place where health care is provided. Such errors affect estimates of spatial accessibility to physicians and local measures of physician supply. Some researchers have mentioned this geocoding problem (e.g. Conrad et al., 2008; Grumbach et al., 2003; Rosenthal et al., 2005), but none have systematically analyzed its geographic implications.

Geocoding error can arise during any phase of the geocoding process. The geocoding process involves linking address information for disease cases or health-related facilities to a geographic reference file that identifies locations on the earth's surface. Both of these files, the address file and the geographic reference file, as well as the processes by which they are linked, can result in geocoding error – a spatial mismatch between the geocoded location and its true position (Gregorio et al., 1999; Rushton et al., 2006). Errors can stem from inaccuracies in the address file including missing fields, misspellings, and post office box addresses (Rushton et al., 2006). Geographic base files are also a potential source of error as they may include inaccuracies that prevent the matching of addresses to correct geographic locations. Common problems in geographic base files include missing or mis-aligned streets or errors in street numbering data. Finally, the geocoding method itself can result in errors (Bonner et al., 2003; Duncan et al., 2011).

A key topic in the literature on geocoding is to document and analyze the extent and spatial characteristics of geocoding error. According to Zandbergen (2008), such error can be characterized by several indicators including match rates, positional accuracy (positional error), and repeatability. Match rate measures the percent of observations that can be assigned to a geographic location, emphasizing the completeness of the geocoding process. Positional error characterizes the spatial separation and orientation between geocoded and true locations. A variety of metrics have been used in assessing positional error including the average distance between geocoded and true point locations, the accuracy of assignments to zones such as census tracts, and the spatial orientation and clustering of errors (Bonner et al., 2003; Zandbergen, 2008). Finally, repeatability describes the consistency of geocoding results based on different algorithms and geographic reference files (Goldberg and Cockburn, 2010). In this research, we emphasize