Spatial and Spatio-temporal Epidemiology 3 (2012) 17–29

Contents lists available at SciVerse ScienceDirect

Spatial and Spatio-temporal Epidemiology

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

Investigating impacts of positional error on potential health care accessibility

Scott Bell [a,*], Kathi Wilson [b,1], Tayyab Ikram Shah [a,2], Sarina Gersher [a,3], Tina Elliott [a,3]

[a] Department of Geography and Planning, University of Saskatchewan Saskatoon, Canada SK S7M 5C8
[b] Department of Geography, University of Toronto Mississauga Mississauga, Canada ON L5L 1C6

ARTICLE INFO

Article history:
Available online 23 February 2012

Keywords:
GIS
Geocoding
Positional error
Health care accessibility

ABSTRACT

Accessibility to health services at the local or community level is an effective approach to measuring health care delivery in various constituencies in Canada and the United States. GIS and spatial methods play an important role in measuring potential access to health services. The Three-Step Floating Catchment Area (3SFCA) method is a GIS based procedure developed to calculate potential (spatial) accessibility as a ratio of primary health care (PHC) providers to the surrounding population in urban settings. This method uses PHC provider locations in textual/address format supplied by local, regional, or national health authorities. An automated geocoding procedure is normally used to convert such addresses to a pair of geographic coordinates. The accuracy of geocoding depends on the type of reference data and the amount of value-added effort applied. This research investigates the success and accuracy of six geocoding methods as well as how geocoding error affects the 3SFCA method. ArcGIS software is used for geocoding and spatial accessibility estimation. Results will focus on two implications of geocoding: (1) the success and accuracy of different automated and value-added geocoding; and (2) the implications of these geocoding methods for GIS-based methods that generalise results based on location data.

1. Introduction

Accessibility to health services at local and community scales is an important metric for measuring health care delivery in Canada and the United States. The concept of access to health care is multifaceted; it builds links between populations at risk (clients) and the delivery system (service providers) which vary across both space and place (Penchansky and Thomas, 1981). In measuring potential access to health services Geographical Information Systems (GIS) and spatial methods provide powerful analytic tools. The Three-Step Floating Catchment Area (3SFCA) method is a GIS-based procedure developed by Bell (forthcoming) to calculate potential (spatial) accessibility at the neighbourhood level as a ratio of primary health care (PHC) providers to population in urban settings.

Like other GIS based methods, measuring potential (spatial) access to health care requires locations of Primary Health Care (PHC) providers in global absolute geographic coordinates (Latitude/Longitude, Universal Transverse Mercator (UTM), etc.) and population information associated with enumeration areas (census areas or local neighbourhoods) (Bell et al., forthcoming; Luo, 2004; Luo and Wang, 2003; McGrail and Humphreys, 2009; Paez et al., 2010; Schuurman and BÉRubÉ, 2010). In Canada, census based population data is gathered by Statistics Canada every five years and is available at a variety of enumeration levels. One such enumeration unit, and the unit used in

* Corresponding author. Tel.: +1 306 966 5676; fax: +1 306 966 5680.
E-mail addresses: smb332@campus.usask.ca (S. Bell), kathi.wilson@utoronto.ca (K. Wilson), tayyab.shah@usask.ca (T. Shah), sag221@mail.usask.ca (S. Gersher), tme021@mail.usask.ca (T. Elliott).
[1] Tel.: +1 905 828 3864; fax: +1 905 828 5273.
[2] Tel.: +1 306 341 4145; fax: +1 306 966 5680.
[3] Tel.: +1 306 966 5133; fax: +1 306 966 5680.

1877-5845/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sste.2012.02.003

this study, is the dissemination area (DA) which provides good spatial resolution.

Geocoding is increasingly used in health studies to map sites of service providers and participants. There are many aspects of geocoding that require attention in order to ensure a sufficient match rate to generate location data that is reliable. For instance, the positional accuracy of geocoded locations depends on the geocoding techniques employed and reference data used. In general, an automated address match geocoding procedure is used to convert each address to a pair of geographical coordinates. The automated part of the geocoding process can be accomplished in several ways, these include range interpolation, areal unit interpolation, and rooftop geocoding; in this research we use range interpolation as the primary automated method. Range interpolation involves interpolating address locations along a street segment; such estimations of position can introduce positional error in the geocoded point (Goldberg, 2008; Zandbergen, 2009). This is traditionally followed by a manual intervention/interactive geocoding process that involves examining possible location candidates for an address from a digital street file (Goldberg et al., 2008). In this research, the term value-added refers to those manual interventions or measures taken after the initial automated geocoding process; such methods are generally accepted as increasing the validity and accuracy of the output. The primary objective of this research is to investigate the positional error that results from geocoding PHC provider addresses using six different geocoding methods. These methods include: (1) automated Postal Code geocoding with Desktop Mapping Technologies Inc. (DMTI) data; (2) value-added matching with Postal Code data; (3) automated range interpolation using DMTI street data; (4) value-added matching with DMTI data; (5) automated range interpolation using ESRI Tele Atlas street data (bundled with ArcGIS 10); and (6) value-added matching with ESRI Tele Atlas/street data. The secondary objective of our study is to investigate the impact of positional error on measures of accessibility estimated using the three-Step Floating Catchment Area (3SFCA) method (Bell et al., forthcoming).

2. Background

Geocoding has different meanings depending on its application (Goldberg et al., 2007). In health research, geocoding is used as a means of transforming textual geographic descriptions into explicitly georeferenced data that can be used for spatial analyses (Goldberg, 2008). Geocoding has an established role in health research, whether estimating supply and demand of various health services (Schuurman and BÉRubÉ, 2010) or dealing with disease patterns and distribution (Bruneau et al., 2008). Address match geocoding is the process of transforming addresses in local and relative coordinate systems (such as street addresses or postal codes), which are not themselves amenable to GIS-based spatial analysis, into a format which assigns coordinates in an global absolute coordinate system (such as latitude and longitude or UTM). Geocoding is largely an automated process and is at least partially dependent on the quality of the reference data used to estimate an addresses location (Zandbergen, 2009, 2011); therefore, uncertainty exists in all geocoding output. In the context of health research, there are important implications of such errors. What is little understood is the extent to which such geocoding errors manifest themselves in higher order applications of the geocoded results.

2.1. Common geocoding errors

There are four main types of error related to address match geocoding (Zandbergen, 2009). The four types of error include:

1. Error arising from geocoding an incorrect address. An address can be incorrect in a variety of ways, two commons errors include: (1) a typo in the number component of the address, and (2) an error in the street designation (wrong designation or an incorrect interpreted abbreviation).
2. Errors related to inaccurate interpolation along a street segment. Since it is rare that street addresses are evenly distributed across a street segment such errors are common but tend to be smaller in urban areas where street segments are shorter (Bakshi et al., 2004; Cayo and Talbot, 2003); even within urban areas there can be variability in such errors, for instance, in areas dominated by multi-family housing units, condominiums, or apartment buildings (Ward et al., 2005; Zimmerman and Li, 2010). Interestingly, commercial areas are also susceptible to higher error resulting from this type as there tend to be fewer commercial entities per street segment than in residential areas, resulting in sometimes arbitrary address numbering along the segment (Zandbergen, 2008).
3. Error resulting from the geocoded point being placed at an incorrect perpendicular distance from the street segment (i.e. incorrect side offsets); again, this type or error is generally minimal.
4. Error in the placement of the reference data's street segments within the road network can produce misplaced location that are difficult to reconcile without local knowledge; in this type of error the address is correctly located along the street segment but the street segment is not in a location congruous with its position in the street network.

2.2. Geocoding with areal units

A secondary geocoding method involves the geocoding of points using areal unit reference data. Examples of such geocoding include zip codes, postal codes, and other land allocation systems. While some of the potential errors listed above can extend to areal unit geocoding, such geocoding has its own set of challenges. In Canada, postal codes can either be used to geocode independently (Bruneau et al., 2008) or to supplement missing or erroneous street address information (Schuurman and BÉRubÉ, 2010). Here, it is important to describe the difference between Canadian postal codes and US Zone Improvement Program (ZIP) Codes. In the United states a ZIP Code