Spatial and Spatio-temporal Epidemiology 3 (2012) 243–253

Contents lists available at SciVerse ScienceDirect

Spatial and Spatio-temporal Epidemiology

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

The impact of place and time on the proportion of late-stage diagnosis: The case of prostate cancer in Florida, 1981–2007

Pierre Goovaerts [a,*], Hong Xiao [b]

[a] BioMedware, Inc., 121 W. Washington St., 4th Floor – TBC, Ann Arbor, MI 48104, USA
[b] Florida A&M University, Tallahassee, FL, USA

ARTICLE INFO

Article history:
Received 29 August 2011
Revised 20 February 2012
Accepted 2 March 2012
Available online 13 March 2012

Keywords:
Cluster analysis
Boundary analysis
Binomial kriging
PSA screening
Urbanization

ABSTRACT

A suite of techniques is introduced for the exploratory spatial data analysis of geographical disparities in time series of health outcomes, including 3D display in a combined time and geography space, binomial kriging for noise filtering, space–time boundary analysis to detect significant differences between adjacent geographical units, and spatially-weighted cluster analysis to group units with similar temporal trends. The approach is used to explore how time series of annual county-level proportions of late-stage prostate cancer diagnosis differ across Florida. The state-average proportion of late-stage diagnosis decreased 50% since 1981. This drop started in the early 1990s when prostate-specific antigen (PSA) test became widely available and several parts of Florida underwent fast urbanization. Boundary analysis revealed geographical disparities in the impact of the screening procedure, in particular as it began available. The gap among counties is narrowing with time, except for the Big Bend region where the decline is much slower.

1. Introduction

Interpretation of cancer incidence and mortality rates in a defined population requires an understanding of multiple complex factors that likely change through time and space, and interact with the different types and scales of places where people live. These factors include the prevalence of risk factors in the population, changes in the use of medical interventions to screen and treat the disease, and changes in how data are collected and reported. Analyzing temporal trends in cancer incidence and mortality rates can provide a more comprehensive picture of the burden of the disease and generate new insights about the impact of various interventions (Potosky et al., 2001). The analysis of temporal trends outside a spatial framework is however unsatisfactory, since it has long been recognized that there is significant variation among US counties and states with regard to the incidence of cancer (Cooper et al., 2001). Visualizing, analyzing and interpreting these geographical disparities should bring important information and knowledge that will benefit substantially cancer epidemiology, control and surveillance.

Despite the significant work accomplished in health data visualization and analysis this last decade, spatial and temporal data are still displayed in separate views and so one does not capitalize on the human visual processing engine to extract knowledge from the spatial interconnectedness of information over time and geography. For example, Geographic Information System (GIS) products, such as GeoDA (Anselin et al., 2006) or ESRI ArcView, show events in the single dimension of space on a map. In each case, only a thin slice of a multidimensional picture is represented. Recently Goovaerts (2010a) proposed the use of time as a third dimension to display time series of cancer mortality and incidence data. The 3D view of time series of health outcome maps makes it easier to comprehend spatiotemporal relationships because there is no disconnect between the temporal and spatial dimensions as

* Corresponding author.
E-mail addresses: goovaerts@biomedware.com, goovaerts@terraseer.com (P. Goovaerts), hong.xiao@famu.edu (H. Xiao).

1877-5845/$ - see front matter
http://dx.doi.org/10.1016/j.sste.2012.03.001

opposed to a combination of 2D map and linked time line plots or an animation.

During the analysis of large space–time–attribute datasets, users may have difficulty perceiving, tracking and comprehending numerous visual elements that change simultaneously both in space and time, such as yearly time series for 67 counties in Florida. A common solution adopted in the spatial domain is to group or cluster geographical units with similar properties (Guo, 2008). Including additional information, such as the geographical locations of observations, in the classification creates clusters that are spatially compact and more easily interpretable. One popular approach in soil sciences is spatially weighted classification that is based on a dissimilarity matrix that accounts for distances in both the attribute and geographical spaces (Caeiro et al., 2003; Simbahan and Dobermann, 2006). This approach has however been applied mainly to stratify isopleth maps of interpolated values and it has always been used outside a temporal framework. In this paper, we introduce a dissimilarity measure to assess differences between time series of health outcomes in both the geographical and attribute spaces. To account for the instability of rates recorded in sparsely populated counties (small number problem), the dissimilarity measure is computed after noise-filtering using binomial kriging (Goovaerts, 2009b). This approach is similar to the practice of computing local Moran's I on rates that are first noise-filtered using empirical Bayes smoothers (Anselin et al., 2006).

A natural complement to the clustering of geographical units is provided by boundary analysis since the edge of a cluster necessarily implies a boundary (Jacquez et al., 2008). Yet, boundary detection allows a finer analysis than cluster detection because only two entities are considered at a time, leading to the detection of significant changes or edges that might go undetected when neighboring rates are averaged. In recent years, substantial insights and benefits have accrued by using geographic boundary analysis to study spatial patterns of cancer health outcomes. The identification of zones of rapid change has allowed researchers to focus scientific and epidemiological inquiry on those areas where mortality and/or incidence are changing rapidly, and to then evaluate whether these transition zones tend to occur near boundaries in putative environmental exposures (Jacquez and Grieling, 2003). The application of boundary analysis to space–time health data poses however two challenges: (1) the need to account for the instability of rates recorded in sparsely populated counties, and (2) the incorporation of the time dimension in this intrinsically spatial technique. These two aspects are here tackled by the repetition across time of the geostatistical boundary analysis introduced by Goovaerts (2010b).

Prostate cancer is the most frequently diagnosed non-skin cancer and the second leading cause of male cancer-related death in the US. Prostate cancer mortality and late-stage diagnosis started declining after 1991 (Smart, 1997; Chu et al., 2002). According to some studies, this decline in mortality is due to early detection (prostate-specific antigen (PSA) screening) although screening for prostate cancer is still controversial (Farkas et al., 1998; McDougall et al., 2000; Coldman et al., 2003; Shaw et al., 2004). Other studies showed that men who are diagnosed with and treated aggressively for localized prostate cancer have higher survival rates compared to men diagnosed with advanced-stage cancer (Wong et al., 2006). Although prostate cancer-related incidence and mortality have declined recently, striking geographical and racial/ethnic differences in incidence and mortality persist in the United States. For example Jemal et al. (2005) showed that non-metro counties generally had higher death rates and incidence of late-stage disease and lower prevalence of PSA screening (53%) than metro areas (58%), despite lower overall incidence rates. Their analysis was however conducted for a single time period (1995–2000) and based on state-level data.

This paper explores how the county-level proportions of prostate cancer diagnosed late among patients 65 years and older changed yearly over the period 1981–2007 in Florida. This exploratory spatial data analysis of aggregated data is a preliminary step toward the quantification of the relative contribution of contextual (neighborhood-level) and compositional (individual-level) factors through multi-level regression. The approach rely on techniques that are either new or were recently introduced in the field of health geostatistics and medical geography (Goovaerts, 2009a). Although a county-level analysis might seem rather crude and limits the interpretation of results because of potentially wide heterogeneity within a county, the present study represents a substantial improvement over most analyses of temporal trends which are usually aspatial and conducted at the National level or for a single cancer registry. In addition, county-level analysis allowed the use of a fine temporal resolution (i.e. year) which would not be possible for finer spatial resolutions because of rate instability caused by the small number problem.

2. Data and methods

Number of cases of prostate cancer and associated stage at diagnosis recorded yearly from 1981 through 2007 for non-Hispanic white males within each county of Florida were downloaded from the Florida Cancer Data System website. Proportions of late-stage diagnosis were computed for each year and county using only cases 65 years and over to minimize the impact of disparities in age distribution across Florida and attenuate the impact of variability in health coverage since all cases are covered by Medicare. One potential problem associated with the analysis of time series of areal data is temporal changes in the definition of administrative units used to report the results. This was not the case in the present study since no county has been deleted or created in Florida since 1925. In addition, out of the 144 boundaries that exist between adjacent Florida counties, only four slightly changed between 1981 and 2007. Two of these changes consisted in a shift of the boundary over water bodies (e.g. from the east bank to the middle of a river), so without any impact on the county population.

The rates of late-stage diagnosis were processed using binomial kriging (Goovaerts, 2009b) to filter the noise