Spatial and Spatio-temporal Epidemiology 1 (2010) 207–218

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

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

Geographic boundary analysis in spatial and spatio-temporal epidemiology: Perspective and prospects

Geoffrey M. Jacquez *

BioMedware, Inc., 3526 W. Liberty Rd., Suite 100, Ann Arbor, MI 48103, USA

ARTICLE INFO

Keywords:
Geographic boundary analysis
Space-time change
Health outcomes
Environmental exposures
Health-environment association

ABSTRACT

Geographic boundary analysis is a relatively new approach that is just beginning to be applied in spatial and spatio-temporal epidemiology to quantify spatial variation in health outcomes, predictors and correlates; generate and test epidemiologic hypotheses; to evaluate health-environment relationships; and to guide sampling design. Geographic boundaries are zones of rapid change in the value of a spatially distributed variable, and mathematically may be defined as those locations with a large second derivative of the spatial response surface. Here we introduce a pattern analysis framework based on Value, Change and Association questions, and boundary analysis is shown to fit logically into Change and Association paradigms. This article addresses fundamental questions regarding what boundary analysis can tell us in public health and epidemiology. It explains why boundaries are of interest, illustrates analysis approaches and limitations, and concludes with prospects and future research directions.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

"A boundary is that which is an extremity of anything"

Euclid's *Elements*: Book 1

Why employ boundary analysis? Three reasons are paramount. First, boundaries are where the values of a variable are changing rapidly, and are often of direct scientific interest since they are zones of dynamic geographic change (e.g. edges of neighborhoods defined by socio-economic status, employment and deprivation; zones of population mixing in population genetics; the edges of disease clusters in public health; places where environmental exposures are changing and so on). In spatial and spatio-temporal epidemiology boundary analysis may be used to identify edges of populations homogeneous in health outcomes, covariates and/or risk factors. This is useful when identifying study populations, targeting groups for health interventions, when siting health screening facilities, and for exploring relationships between environmental exposures and health outcomes (Jacquez and Greiling, 2003).

Second, boundary analysis allows us to better define sample populations, increasing our ability to resolve underlying functional relationships. It is difficult to accurately assess odds ratios, fit models and assess health-environment relationships within homogeneous areas – both exposed and not-exposed groups are required in order to find an effect. A common mistake in geographic sampling design is to focus on those sub-populations with a high risk in the health outcome of interest. In these instances we shouldn't be surprised by an inability to reveal underlying health-environment relationships, since the range of variability needed to resolve them is lacking. Consider for example Fig. 1, left, which shows no relationship between the values of exposure and health outcome variables sampled from within an area homogeneous in the values of these variables – e.g. away from geographic boundaries in the health outcome. By placing samples *across* such boundaries the analyst is better able to capture the full

* Tel.: +1 734 913 1098; fax: +1 734 913 2201.
E-mail address: jacquez@biomedware.com

1877-5845/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.sste.2010.09.003

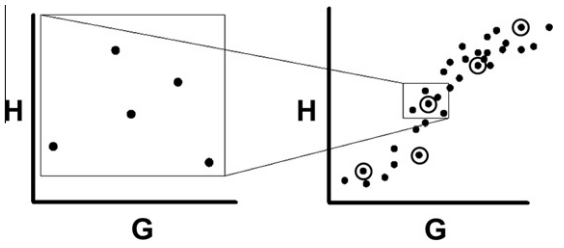

Fig. 1. Failure to sample a representative range of variability can lead one to miss health (H) environment (G) relationships. Sampling from within a geographic cluster (map not shown) will yield homogeneous values such as shown in the rectangle (left). Sampling across geographic boundaries (map not shown), which are zones of rapid change in values, results in a sample drawn from the full range of variability, as shown by the observations in circles (right). In practice one first identifies boundaries on the map of the variables and then samples across them (e.g. on both sides of and within the boundaries themselves). These graphs illustrate that zones of rapid change identified by geographic boundary analysis can be used to guide sample design.

range of variability in the variables, and to detect the functional relationship (Fig. 1, right).

Third, boundary analysis allows us to relax unrealistic and/or unfounded assumptions regarding the form of the functional relationships between measures of human health and its predictors. Tests for boundary overlap require that the variables whose association is being assessed covary only to the extent that change in one results in change in the other, and are less stringent about the form of the relationships between the variables. In practice boundary overlap may be assessed in several ways, including minimum average distance between health and environment boundaries (Jacquez, 1995), area intersection operations (Maruca and Jacquez, 2002), and the direct overlap of the boundaries themselves. None of these approaches make assumptions regarding the functional form of the underlying health environment-relationship. Contrast overlap analysis with approaches such as the Pearson product–moment correlation coefficient, which assumes a linear dependence between the variables. Boundary overlap does not make assumptions regarding the form of the model of dependence. This is a critical assumption to relax since relationships of biological interest are often non-linear and may not even be monotonic.

Boundary analysis informs spatial pattern analysis, which is classified for convenience into Value, Change, and Association questions. These three questions are similar to those identified as important to ask of an atlas map by epidemiologists (see Pickle, 2009); these in turn are similar to Bertin's classification of visualization tasks (Bertin, 1974). *Value* questions have to do with the values of the variables surveyed, and how they are arranged in geographic space. Value questions are explored using disease mapping through spatial point distributions, choropleth maps (Richards et al., 2010) and related techniques. This in many ways is the point of departure for spatial epidemiology, with examples such as Snow's Cholera map (Snow, 1855) and disease Atlases [see (Pickle, 2009) for a review]. Value questions are the domain of disease clustering, which seeks to identify spatially contiguous areas of high or low disease occurrence. This includes techniques for case-control data (Cuzick and Edwards, 1990), case count and population at risk data (Takahashi et al., 2004; Tango and Takahashi, 2005; Kulldorff et al., 2006) and disease rates (Rushton et al., 2004).

Change questions have to do with higher order properties of spatial response surfaces, such as gradients (how values change through geographic space). Boundary analysis is the dual of cluster analysis, in that the former seeks to identify geographic areas where the health outcome (e.g. disease risk) is changing rapidly (e.g. where the spatial response surface has large derivatives), while the latter seeks to identify local populations with high relative risks (e.g. where the derivative is near zero and disease risk is high). Methods for detecting boundaries date back at least to 1951 (Womble, 1951), and include geostatistical (Goovaerts, 2008), Bayesian (Lu and Carlin, 2005), wavelet (Csillag et al., 2001), distribution-based (Jacquez et al., 2008), difference (Monmonier, 1973), as well as distribution-free approaches (Hall, 2008). Several methodological reviews are available for readers who wish to become more familiar with these techniques (Fortin, 1994; Jacquez et al., 2000; Kent, 2006).

Association questions seek to relate spatial pattern in one variable or set of variables to the pattern in another set of variables, and include diverse methods such as boundary overlap (Jacquez, 1995), map area intersection (Sadahiro and Umemura, 2001; Maruca and Jacquez, 2002; Robertson et al., 2007), spatial regression modeling (Mantel, 1967; Greenland and Robins, 1994; Dormann et al., 2007; Fotheringham, 2009), geostatistical analysis (Goovaerts, 2009) and Bayesian disease mapping (Ma et al., 2007; Lawson and Banerjee, 2008). As noted above, tests for boundary overlap evaluate association by determining the extent to which features on spatial response surfaces coincide.

This paper is a perspective on some of the issues and problems in boundary analysis in public health. It begins with a description of technological and societal trends, alternative approaches to pattern recognition, and then focuses on statistical approaches that support probabilistic assessment of how unusual a pattern is under a specified null hypothesis. Value, Change, and Association questions are then described in detail. This perspective is illustrated with a motivating example: the pattern of leukemia incidence in eight counties in New York State. This pattern is related to the location of sites contaminated with TriChloroethylene (TCE), using a step-wise approach involving Value, Change and Association questions. The importance of pattern recognition in extracting knowledge from the burgeoning information stream made possible by emerging technologies is described, as is the role of pattern analysis in scientific inquiry. The author concludes with a discussion of current needs such as improved null spatial models, and speculates on the future of the field.

2. Technological and societal trends

Recent advances in remote sensing are providing hyperspectral imagery at the sub-1 meter scale for most

Download English Version:

https://daneshyari.com/en/article/1064409

Download Persian Version:

https://daneshyari.com/article/1064409

Daneshyari.com