The emerging role and benefits of boundary analysis in spatio-temporal epidemiology and public health

Geoffrey M. Jacquez, MS, PhD [President, BioMedware][1] and **Pierre Goovaerts, PhD [Chief Scientist]**[1]

Geoffrey M. Jacquez: jacquez@biomedware.com; Pierre Goovaerts: goovaerts@biomedware.com

[1] BioMedware, Inc, 3526 W. Liberty Rd., Suite 100, Ann Arbor, MI 48103

Keywords

geographic boundary analysis; space-time change; health outcomes; environmental exposures; health-environment association

1. Introduction

This special issue on geographic boundary analysis in spatio-temporal epidemiology and public health seeks to expand awareness on the contributions and potential caveats of using geographic boundary analysis to study human health outcomes, correlates and determinants. Epidemiology has a strong clinical tradition focused on the individual, on the characteristics of the population from which that individual arises, and on the risk factors that individuals may have been exposed to during critical times over his or her life course. The emphasis is on the construction of insightful hypotheses whose validity may be formally evaluated and on the identification of factors that may increase or reduce disease risk. Quantitative methods play a role in *estimation*, to provide accurate and unbiased estimates of the incidence and prevalence of health events and associated risk factors; in *hypothesis testing*, to evaluate whether sub-groups with quantitatively different exposures have concomitant differences in disease risk; and in *modeling*, to provide a formal basis for prediction and model-based theory formulation and testing (Kuh and Ben-Shlomo 1997; Koopman, Chick et al. 2000).

The geography of the individual and of the population whence that individual comes underpins many aspects of the life course that are of direct epidemiological interest including social networks, socio-economic status and deprivation, accessibility to resources, and many of the ingredients of living that are determinants of individual health status and outcomes. From this perspective, an increased understanding of the locations, extent, and strength of geographic boundaries can directly strengthen our understanding of the spatio-temporal epidemiology of disease and hence our ability to design, conduct and assess the effectiveness of public health interventions. Consider some examples.

Correspondence to: Geoffrey M. Jacquez, jacquez@biomedware.com.

Obesity and overweight

Overweight, obesity and inactivity are a rapidly growing problem in many post-industrial nations. The magnitude of the problem is astonishing. Since the 1980's in the United States obesity has increased steadily across all states, genders, age groups, ethnicities, and education levels, with 31% of all US adults obese in 2004 (Mokdad AH 1999; Flegal KM 2002; Frank, Andresen et al. 2004). This "obesity epidemic" has strong spatial and temporal patterns, with southern states serving as the initial nexus, followed by rapid increases in obesity and overweight across the Southeast, Midwest and western states in ensuing decades. Recent studies have focused on how the local neighborhood environment, access to fast foods, absence of accessibility to grocery stores, travel patterns, and the availability of opportunities for routine exercise serve as predictors of the prevalence of overweight and obesity in local populations (Frank, Andresen et al. 2004). The built environment and other neighborhood characteristics appear to be risk factors, with perceptions of "unsafe neighborhood" associated with increased risk of overweight and obesity. If we accept that overweight and obesity have strong spatial and temporal structure that varies locally (e.g. neighborhoods) as well as regionally (e.g. states), questions such as "where are the neighborhood boundaries" and "are the borders of states meaningful in terms of risk of obesity and overweight" become directly relevant.

Alcohol abuse

Recent studies have demonstrated that alcohol abuse is a growing problem on college campuses, and that alcohol availability – such as the density of outlets selling alcohol, and their proximity to campus – are one of the predictors of binge drinking (Scribner, Ackleh AS et al. 2009). The definition of alcohol availability from the perspective of the student drinker may be measured in many ways, based (for example) on travel time from student residence to nearest alcohol outlet. For larger student groups it may make sense to measure accessibility based on physical proximity of the campus to those districts where the bars and liquor stores are found. But where are the margins of such areas, and how do we quantify their boundaries?

Infectious diseases

Space-time boundaries are of direct interest in the study of epidemics, as boundaries are defined by the epidemic wave or front of the epidemic. With such boundaries defined, it then is possible to evaluate their velocity, and to predict when the disease will reach a given at-risk population. Such analyses have been accomplished for the spread of raccoon rabies in the mid-west and eastern United States, and to the historic spread of the bubonic plague from the Middle East and across Europe.

Vector-borne disease risk mapping relies on the identification of habitat suitable for the vector, and the role of suitable habitat in the vectors life history and opportunities for infection transmission to human hosts. The extent to which vector habitat overlaps or intersects with the space-time paths of susceptible humans underpins infection transmission. The boundaries and extent of suitable vector habitat is critical for the prediction of vector-borne disease risks.

Age, demography, socio-economic structure and neighborhoods

Age structure, demographic characteristics and socio-economic status are often important covariates for infectious as well as chronic diseases. They commonly exhibit strong geographic pattern, and are organized into self-similar neighborhoods with common housing types, incomes and ethnicity. Neighborhoods often have well-defined boundaries. Some are fiat boundaries imposed by legislative or bureaucratic action (such as land use zoning); some are circumscribed by features such as roads and rivers; while still others reflect changing urban landscapes and socio-economic trends (such as gentrification). The definition of neighborhood boundaries can be a critical step in the construction of spatially explicit models such as multilevel models (Subramanian 2010).

Chronic diseases

Cancers, cardiovascular disease and other chronic diseases can differ dramatically in incidence and mortality at national, regional and local levels (Jemal, Kulldorff et al. 2002; Fang, Kulldorff et al. 2004; Pickle 2009; Greiling, Jacquez et al. 2010). Across contiguous areas change in disease incidence and mortality may be absent, gradual or abrupt, with a good amount of geographic variability explicable wholly or in part by underlying changes in covariates and risk factors that are themselves spatially structured.

Understanding the locations, extent and boundaries of geographic clusters in chronic disease, its predictors, risk factors and covariates is critical in order to identify populations for targeted interventions, site screening facilities in order to optimize service to the community, and to generate hypotheses regarding potential causal factors (Rushton, Peleg et al. 2004; Meliker, Goovaerts et al. 2009; Richards, Berkowitz et al. 2010).

2. Overview of the contents of boundary analysis special issue

These examples perhaps convey some understanding as to why geographic boundary analysis might be of use in epidemiology and public health. The remainder of this special issue provides perspectives on applications of boundary analysis in the health sciences, methodology papers that describe novel techniques and application papers that use boundary analysis to advance our understanding of specific disease outcomes.

The first paper by Copeland on boundary analysis in public health investigations provides a perspective from the Director of the Michigan Cancer Surveillance Program. Copeland describes the role of public health and how boundary analysis can provide tools for public health investigations. The value of epidemiology in public health is that it ties discovery regarding disease causality to putting this knowledge into action, with the objective of implementing prevention measures on a population scale. As such, the scientific study of disease origin, pathology and transmission patterns is crucial, as is information regarding the timing and placement of disease-related events in geographic space and time. One aspect is the identification of disease clusters, and here Copeland finds four salient features that can be better informed via boundary analysis: the value of individual level data; the importance of scale in an analysis; the need for geographic and analytical precision and avoiding false findings. A second aspect is associating cancer outcomes with putative causes, risk factors and covariates. Copeland sees particular value in the ability of boundary analysis to relax constraints imposed by arbitrary administrative boundaries such as the margins of census units (also called “fiat” boundaries), an approach demonstrated later in this special issue in Grady’s analysis of racial segregation and underweight birth. The ability to precisely identify the actual boundaries of clusters of health events, independent of fiat boundaries, may provide superior information for assessing health event data, analyzing covariates, planning interventions and monitoring intervention outcomes.

The second perspective paper by Jacquez states three reasons for using boundary analysis. First, boundaries are often of direct scientific interest since they are the edges of neighborhoods characterized by socio-economic status, employment and deprivation; zones of population mixing in population genetics; places where environmental exposures are changing and so on. Second, boundary analysis supports the definition of sampling protocols, 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. Knowledge of geographic boundaries helps assure a sampling protocol that draws observations from an appropriate range of values. Third, boundary analysis can relax 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. Jacquez then illustrates an analytic framework for boundary analysis in a study of leukemia in central New York.

The first methods paper by Goovaerts addresses the problem of how to account for multiple testing and spatial autocorrelation in boundary analysis. Boundary analysis employs a host of different approaches, including edge detection, gradient operators, distribution-based methods, distribution-free approaches, Bayesian methods as well as geostatistical techniques. The problem of how to appropriately model spatial autocorrelation in the data pervades all approaches, as does appropriate handling of the multiple tests that can arise when repeatedly evaluating boundaries at multiple locations and/or across spatial scales. Goovaerts considers four alternative approaches in boundary analysis. The first seeks to quantify changes in rates across boundaries using both differences in rates as well as rate ratios. The second offers an approach to detecting a minimum magnitude in the boundary (e.g. slope gradient) using neutral hypotheses of non-uniform risk across the study area. The third accounts for observed differences in population sizes as well as spatial patterns in the variable of interest within the randomization method itself, affording the ability to subsume these within the null hypothesis of the test. The fourth provides the ability to account for multiple testing when evaluating the statistical significance of individual boundaries themselves using p-values. Together, these techniques address several fundamental methodological problems of boundary analysis, and their enhanced ability to detect “true” boundaries and reduce false positives is demonstrated using simulation studies.

NIH-PA Author Manuscript

The next contribution by Maxwell details how to identify land cover boundaries from satellite and remotely sensed imagery using object-based image analysis (OBIA). Maxwell argues that satellite imagery, aerial photography and related technologies constitute an important resource in understanding spatio-temporal relationships between environment and health. Satellite imagery is now available for almost all locations around the world at spatial resolutions down to fractions of a meter, at fine spectral resolutions that range beyond the visible spectrum, and with temporal resolution (how frequently the same location is imaged by the satellite) on the order of 24 hours. However, the identification of boundaries of land cover features (e.g. trees, buildings, and crop fields) has traditionally used a very time consuming process of hand digitizing. Boundary detection algorithms are increasingly being applied using (OBIA) technology to automate the process. Maxwell presents an overview of the OBIA approach and a demonstration using a time series of high resolution aerial photography (1 m) and medium resolution Landsat imagery (30 m) to assess pesticide spray drift exposure in California.

NIH-PA Author Manuscript

Grady’s contribution explores how improved neighborhood boundary definitions enhance the evaluation of relationships between racial residential segregation and low birth weight. The current state of practice typically uses census tract boundaries to estimate individual exposures, but census tracts do not necessarily capture aspects of neighborhood environments that reflect racial segregation. Grady therefore applied an automated zone matching (AZM) methodology to optimize neighborhood boundary definitions and to relax the “fiat boundaries” constraint imposed by the census geography. She then assessed the impact of racial isolation and racial clusters on intrauterine growth retardation (IUGR) and preterm birth using multilevel models and the AZM boundaries and, and contrasted these results with those obtained using census boundaries. The predicted means of IUGR incidence across zones and census tracts differed, perhaps reflecting the increased homogeneity of local neighborhoods defined using the AZM methodology. Grady suggests additional research is needed to optimize neighborhood boundary definitions and to assess the sensitivity of segregation measures to changes in scale.

This special issue concludes with a contribution by Grillet and colleagues that assesses the space-time dynamics of boundaries in the incidence of malaria in north-eastern Venezuela. Mosquito-borne diseases in particular might be expected to have irregular boundaries of local outbreaks since the environmental factors that govern vector abundance are often geographically heterogeneous and vary through time. Grillet et al use the State Transition Index (STI) to assess the spatial locations of malaria incidence boundaries and their time dynamic. They applied the STI approach to *Plasmodium vivax* infection spread, and evaluated the role of population size on disease persistence. As expected, boundaries in malaria were spatially variable, with waves of infection attributed to asynchrony in the infection dynamics between host populations. Grillet and colleagues suggest the spatial spread of infection diffusion from large, populated villages where infection is near-endemic to smaller, less populated localities with irregular infection outbreaks may explain this pattern. The public health implication is that malaria surveillance and control in larger communities is required in order to prevent the spread of infection to smaller communities.

3. Conclusions

These contributions have been selected to give a flavour and appreciation for the emerging role of boundary analysis in spatio-temporal epidemiology and public health. The coverage is nowhere near exhaustive. The two perspective papers offer assessments from a public health surveillance professional and from a medical geographer in private industry, but the views of academic researchers and educators, and from government agencies such as the National Institutes of Health are absent. The methodology papers address several important problems in boundary analysis using randomization, geostatistical, image analysis and related techniques, but virtually no coverage is given to Bayesian boundary analysis, and the problems addressed are narrowly focused. On the applied side, we've attempted to provide broad coverage and believe we have had some success with applied studies of cancer, low birth weight, assessment of pesticide spray drift, and of malaria incidence. Notice, however, these are primarily in the epidemiological realm and additional coverage on applied public health practice, for example on how to better target interventions, would have been most useful. Access to the literature on methodological detail, application areas, and public health utility is provided in several of the articles in this special issue.

References

Fang Z, Kulldorff M, et al. Brain cancer mortality in the United States, 1986 to 1995: a geographic analysis. Neuro-oncol 2004;6(3):179–87. [PubMed: 15279710]

Flegal KM, CM, Ogden C, Johnson CL. Prevalence and trends in obesity among US adults. Journal of the American Medical Association 2002;288:1723–7.

Frank LD, Andresen MA, et al. Obesity relationships with community design, physical activity, and time spent in cars. American Journal of Preventive Medicine 2004;27(2):87–96. [PubMed: 15261894]

Greiling, D.; Jacquez, G., et al. Space-Time Visualization and Analysis in the Cancer Atlas Viewer. In: Fischer, MM.; Getis, A., editors. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Berlin Heidelberg: Springer-Verlag; 2010. p. 737-752.

Jemal A, Kulldorff M, et al. A geographic analysis of prostate cancer mortality in the United States, 1970–89. Int J Cancer 2002;101(2):168–74. [PubMed: 12209994]

Koopman J, Chick S, et al. Analyzing sensitivity to model form assumptions of infection transmission system models. Ann Epidemiol 2000;10(7):472. [PubMed: 11018405]

Kuh, D.; Ben-Shlomo, Y. A Life Course Approach to Chronic Disease Epidemiology: Tracing the Origins of Ill-health from Early to Later Life. Oxford: Oxford University Press; 1997.

Meliker JR, Goovaerts P, et al. Breast and prostate cancer survival in Michigan: can geographic analyses assist in understanding racial disparities? CANCER. 2009 (Print Issue Date May 15, 2009). [PubMed: 19365825]

Mokdad AH, SM, Dietz WH, Bowman BA, Marks JS, Koplan JP. The spread of the obesity epidemic in the United States, 1991–1998. Journal of the American Medical Association 1999;282:1511–1519.

Pickle LW. A history and critique of U.S. mortality atlases. Spatial and Spatio-temporal Epidemiology 2009;1(1):3–17.

Richards TB, Berkowitz Z, et al. Choropleth Map Design for Cancer Incidence, Part 1. Preventing Chronic Disease 2010;7(1)

Rushton G, Peleg I, et al. Analyzing Geographic Patterns of Disease Incidence: Rates of Late-Stage Colorectal Cancer in Iowa. Journal of Medical Systems 2004;28:223–236. [PubMed: 15446614]

Scribner R, Ackleh AS, et al. A Systems Approach to College Drinking: Development of a Deterministic Model for Testing Alcohol Control Policies. Journal of Studies on Alcohol and Drugs. 2009 In Review. [PubMed: 19737506]

Subramanian, SV. Multilevel modeling. Handbook of Applied Spatial Analysis: Software Tools. In: Fischer, MM.; Getis, A., editors. Methods and Applications. Springer; 2010.