Geospatial Health 7(2), 2013, pp. 161-168

Spatialising health research: what we know and where we are heading

Tse-Chuan Yang[1], Carla Shoff[1], Aggie J. Noah[2]

[1]Social Science Research Institute and Population Research Institute, The Pennsylvania State University, University Park, PA 16802, USA; [2]Department of Sociology and Population Research Institute, The Pennsylvania State University, University Park, PA 16802, USA

Abstract. Beyond individual-level factors, researchers have adopted a spatial perspective to explore potentially modifiable environmental determinants of health. A spatial perspective can be integrated into health research by incorporating spatial data into studies or analysing georeferenced data. Given the rapid changes in data collection methods and the complex dynamics between individuals and environment, we argue that geographical information system (GIS) functions have shortcomings with respect to analytical capability and are limited when it comes to visualizing the temporal component in spatio-temporal data. In addition, we maintain that relatively little effort has been made to handle spatial heterogeneity. To that end, health researchers should be persuaded to better justify the theoretical meaning underlying the spatial matrix in analysis, while spatial data collectors, GIS specialists, spatial analysis methodologists and the different breeds of users should be encouraged to work together making health research move forward through addressing these issues.

Keywords: spatial data, spatial analysis, health research, geographical information systems.

Introduction

In December 2010, the "Healthy People 2020" project was launched with the overall aim of eliminating health disparities and achieving health equity (US Department of Health and Human Services, 2012). Its approximately 1,200 objectives are organised into 42 important public health topic areas and four overarching goals, one of which proposes to create social and physical environments promoting a good health status for all. This particular goal reflects a scholarly momentum to move beyond individual and family in health research to investigate the complex interplay of factors operating at multiple levels as discussed by Macintyre et al. (2002). It is an approach that sees the individual as embedded in a larger context making health research an inherently spatialised science.

While the relationship between an individual's health and his/her environment is indeed dynamic, the goal of "Healthy People 2020" seems to be somewhat stationary. That is, the goal appears to focus on the health consequences without considering the dynamic processes and mechanisms in which result from the interactions between individual and the environment. Individual health is affected by the environment, while the environment is also continuously reshaped through various activities (Cromley and McLafferty, 2011). The rapid development in geographical information systems (GIS), the improved availability of spatial data and the advancement in spatial analysis have all contributed to the exploration of the role of the environment taking spatio-temporal aspects into account (Rushton, 2003; Elliott and Wartenberg, 2004). Specifically, GIS provides a platform connecting the individual with the environment allowing researchers to inspect the distribution of diseases and investigate potentially modifiable ecological explanations for disease clusters, which may clarify the aetiology of health-related events (Chen et al., 2008; Du et al., 2010). The spatial data that could objectively describe the physical environment are increasingly used to understand how the physical environment is associated with, for example, obesity, asthma and stress-related outcomes (Maantay, 2002; Berrigan and McKinnon, 2008; Tucker et al., 2009; Matthews and Yang, 2010). Clearly, a spatial perspective and spatial modelling should facilitate the examination of the local dynamics between population and environment (Fotheringham, 1997; Haining, 2003).

Studies summarising how spatial approaches have been used in epidemiology, public health and demography have mainly been focused on visualization,

Corresponding author:
Tse-Chuan Yang
Social Science Research Institute and Population Research Institute
The Pennsylvania State University
University Park, PA 16802, USA
Tel. +1 814 865 5553; Fax +1 814 863 8342
E-mail: tuy111@psu.edu

physical environment data and regression methods addressing spatial dependence (Chung et al., 2004; Elliott and Wartenberg, 2004; Reibel, 2007; Auchincloss et al., 2012). In contrast, we intend to complement previous research by emphasising: (i) the availability of data on the social environment; (ii) approaches that embed individuals into context; (iii) local, spatial aspects; and (iv) emerging issues related to spatialising health research (e.g. meanings behind spatial adjacency matrices and spatio-temporal dynamics between health and place). Our aim is to extract constructive future directions for collectors of spatial data, developers of spatial methodology and users of spatial analysis.

GIS and spatial data

GIS is often conceptualised as a toolbox that enables users to capture, manage, analyse and display spatially referenced data (Burrough, 1986; Star and Estes, 1990; ESRI, 2012). The rapid development of GIS technology and its convergence with other geospatial technologies have resulted in the emergence of GIS as a science (Goodchild, 1992; Wilson and Fotheringham, 2007). The various types of spatial information emanating from, for example, global positioning systems (GPS), Earth-observing satellites, cartography, censuses, and surveys including administrative and statutory data (Walford, 2001) that are increasingly available to health researchers, have broadened the GIS scope allowing researchers to focus on questions that could not be answered previously (Reibel and Bufalino, 2005; VanWey et al., 2005; Reibel, 2007).

Geospatial data are increasingly utilised to visualise the relationships between health outcomes and other potential predictors. For example, the World Health Organization (WHO) instituted the Public Health Mapping and GIS program in 1993 (WHO, 2012), and has applied GIS and visualisation of epidemiological data to reveal public health trends and interrelationships. These methods include basic choropleth maps and density estimations, the main strength of which is to display and facilitate analysis of the spatial variability of health outcomes (Rushton, 2003). One main example of visualisation methods is hotspot detection (also known as local cluster detection), a type of analysis that identifies regional anomalies and their spatial patterns as a means for disease surveillance (Auchincloss et al., 2012). These visualisation methods focus predominately on exploratory spatial data analysis of health outcomes without fully accounting for other confounders.

Health researchers have recently started to pay attention to social environment factors and also to the conceptualisation of why place matters for the health of individuals and communities (Entwisle, 2007). This means that research has started to move beyond the simplistic visualisation of health outcomes and their associations with various physical environments to investigate more comprehensive effects of the immediate environment on individuals in order to explain health behaviour and outcomes in a local context (Macintyre et al., 2002; Cummins et al., 2007; Entwisle, 2007). Increasingly, this type of research utilises residential characteristics to examine how social and physical built characteristics are associated with different health outcomes. Characteristics of the social environment include socioeconomic factors such as employment status as well as embedded social relationships, e.g. such as social ties, social capital and social efficacy (Sampson and Groves, 1989; Sampson et al., 1999).

Incorporation of data from built environment (i.e. human-made physical environments that are utilised on a day-to-day basis, e.g. public buildings, parks, etc. (Renalds et al., 2010)), the social environment, and innovative GIS technologies have started to enable researchers to assess environmental risk factors in health research. For example, GIS techniques can accurately create various exposure estimates, including residential proximity to pollution sources (major streets, toxic release inventory sites, etc.) and proximity to various industrial hazards (Clougherty et al., 2007; Ryan et al., 2007; Salam et al., 2008). This type of risk data provides invaluable information for investigating health disparities and environmental injustice across different demographic groups (Evans and Kantrowitz, 2002). Any data with geographical references (e.g. addresses, tract, county, etc.) can be joined with boundary shapefiles and be displayed and analyzed using spatial analytic methods (see annex for detailed discussions). In the United States of America (USA), this type of data is readily available from the Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) system (Watcher, 2005, among others).

Spatial analysis

The concept of homogeneity drives most of the recent developments in spatial analysis. Hierarchical modelling addresses the homogeneity in environmental exposure among people in the same area, and spatial econometrics tackles the homogeneity in spatial relationships among observations. However, homogeneity is only one of the two features of spatial data; the other, heterogeneity, is now receiving more attention (Fotheringham, 2009). More specifically, individuals residing in the same area are always assumed to have equal exposure to risk factors in current hierarchical modelling framework, a relatively crude supposition that ignores the possibility of heterogeneity. It has been suggested that using spatial proximity to capture heterogeneity would generate more accurate exposure measures to investigate the relationships between physical environment and health (McMaster et al., 1997; Maantay, 2002, 2007). In addition, another assumption in hierarchical modelling is that the higher-level analytic units are mutually independent (Raudenbush and Bryk, 2002). This assumption not only ignores both spatial homogeneity and spatial heterogeneity, but also neglects that individuals may be affected by areas outside their residence.

Although widely used in health studies, the spatial econometrics approach has several inevitable methodological shortcomings. First, while ecological studies may be useful for the examination of the structural and contextual factors for disease development and health behaviour (Schwartz, 1994), the findings are subject to the ecological fallacy, which refers to the error of deriving conclusions about individuals based on the results of aggregate analyses or *vice versa* (Piantadosi et al., 1988). Second, though the spatial matrix considers the potential impact of the values of the dependent variable (or omitted variables) in neighbouring areas, it largely ignores the exogenous interaction effect, which refers to the fact that the value of the dependent variable in a unit depends on the values of the independent variables in neighbouring units (Elhorst, 2010). Finally, as research using spatial econometrics is confined to data aggregated to a certain geographic unit, the modifiable areal unit problem (MAUP) should be applicable (Openshaw, 1984; Raghavan et al., 2013).

Several empirical studies have shown that the relationships between health outcomes and predictors may vary intrinsically across space (Chen et al., 2010; Shoff et al., 2012; Yang and Matthews, 2012) and that spatial homogeneity may lead researchers to overlook the locally spatial process and misinterpret their modelling results (Fotheringham, 1997). The increasing emphasis on heterogeneity echoes the idea that individuals interact with the environment differently and the same stimulus may not necessarily lead to the same outcome (Fotheringham, 2009; Yang and Matthews, 2012). Though some efforts have been made to develop novel spatial analysis methods that concentrate on heterogeneity (Reich et al., 2011; Chen et al., 2012), or incorporate both homogeneity and heterogeneity into a single framework (Paez et al., 2002), the development in this area is still in its infancy. Exploring heterogeneity within and across space would facilitate the understanding of the place-specific dynamics between health and place.

A frequently asked question concerns the meaning behind the spatial matrix, but few answers have been provided (Leenders, 2002). In the literature, the first order Queen contiguity matrix is commonly used (Yang et al., 2009, 2011; Sparks and Sparks, 2010; Sparks et al., 2013), but the choice for this spatial matrix has never been well justified. While it could become conventional practice to examine whether the analytic results are robust by using different types of spatial matrices (Ertur and Koch, 2007), doing so would still keep the underlying meaning of a spatial matrix unanswered. We suggest that the selection of the spatial matrix should be connected to, or at least driven by, appropriate theories. For instance, the diffusion theory (Rogers, 2003) explains that the geographically closest neighbours matter in population health because new preventive health care, or the idea of early detection, would travel to nearby areas faster than those farther away. Similarly, following a recent study (Meyers et al., 2005), defining a spatial matrix with migration flows between places may facilitate the understanding of the spread of infectious diseases. Note that while hierarchical modelling assumes independence among spatial units, this assumption may be removed by defining spatial relationships as spatial econometricians do. Nonetheless, it may take extra efforts to develop the methodological framework for this type of analysis.

Finally, the MAUP is a limitation shared by hierarchical modelling and spatial econometrics. There is no agreement on how to solve the MAUP, but it has been suggested that the most ideal solution would be to obtain individual location data and implement analyses with various spatial scales (Weeks, 2004). Nonetheless, due to restrictive access policies and privacy concerns, this approach is not common (Wartenberg and Thompson, 2010). While several techniques have been proposed that could help answer questions related to location (Armstrong et al., 1999; Zimmerman et al., 2008), the potential adverse effects with respect to the analytic results need to be tested. Explicitly, more efforts should be made to develop methods that simultaneously preserve privacy and maintain both spatial and statistical relationships among observations (Boulos et al., 2009). Should this method be available, health researchers would be empowered to overcome the MAUP by dissecting the impact of environmental factors on health and investigating them at different spatial scales, which would further clarify how place influences human health.

Suggestions for future health research

GIS and spatial in explicit data have several disadvantages, one of which is that the results are mainly descriptive. These GIS results are exploratory in nature, because they can imply that there are relationships between variables across space, but they lack explanatory power to address the spatial relationship of the data (Chung et al., 2004). A second disadvantage of GIS software is that it often lacks analytical capability, forcing researchers to turn to other software for that. There is also a need for an infrastructure by which health-related data, built environment and social environment data can easily be obtained. When studying health outcomes, it is not only important to study the environment in which the individual currently lives, but it may also be important to take into account prior environments (Elliot and Wartenberg, 2004; Han et al., 2013). In order to fully understand the relationship between place and health, the ability to link individuals spatially and temporally is critical. Currently, there is a lack of a unified system for obtaining various types of spatial data. GIS and spatial analysis can be incorporated into health research more easily if researchers can access the data they need.

The connection between health data and place is mostly predetermined and it is assumed that the place of occurrence is the only environment that matters. The temporal dimension is absent in this assumption. Arguably, the physical and built environments (e.g. street and land use) change relatively slowly over time. However, individuals are not fixed to the environment in which they currently reside and the change in the exposure to both social and built environment must be assumed to have an impact on health (Leventhal and Brooks-Gunn, 2003; Yabiku et al., 2009; Leventhal and Dupere, 2011). Another reason why the temporal dimension warrants further investigation from health researchers is the increasing availability of social environment data. Since 2010, the US Census Bureau releases the American Community Survey (ACS) five-year estimates, a rolling survey that provides the most current information (on an annual basis) on population and communities, such as demography, economy, housing and other important social variables. Much health information is maintained in a similar fashion, e.g. the National Center for Health Statistics provides mortality data that can be summarised into county-level every year (NCHS, 2010). Before the ACS five-year estimate data became available, many health studies were forced to use decennial data around the census years. A better temporal dimension may become more readily available in the future from GPS devices, cellular phones, and social network web sites (Auchincloss et al., 2012).

Future health research may experience a growth in spatio-temporal data with the addition of new sources of temporal data, and with it comes the demand for spatio-temporal analytic tools. While hierarchical modelling is capable of providing growth curve analysis (Raudenbush and Bryk, 2002), spatio-temporal dynamic modelling remains underdeveloped in spatial econometrics (Banerjee et al., 2004; Gelfand et al., 2005). Health researchers are not yet equipped with appropriate tools permitting comprehensive investigations on how human health is shaped by space over time. However, the proliferation of spatio-temporal data calls for the development of analytic tools for this challenge.

The incorporation of a spatial perspective into health studies helps researchers to identify the determinants of health that are beyond individual genetic, biological and behavioural factors. The process of spatialising health research is related to the use of GIS and spatial data, as well as the application of spatial analytical techniques. The increasing availability of data related to both social environment and built environment (e.g. streets, parks, etc.) has allowed questions that could not be answered previously. Meanwhile, the rapid development of GIS has made it possible to manage all forms of data in one single platform (Goodchild and Janelle, 2010) as well as facilitating exploratory, spatial data analysis (Cromley and McLafferty, 2011). More importantly, with the development of software programmes, implementing complex spatial analytical models, capable of untangling the associations between health and environmental predictors, has become a norm, yielding reliable results as they account for spatial homogeneity, one source of bias. These changes have contributed to the growth of health studies adopting a spatial perspective (Chung et al., 2004; Auchincloss et al., 2012).

Although the development of spatialising health research in the past decades has been successful and laudable (Koch, 2009), there are several weaknesses that should be addressed in the near future in order to continue the current momentum. First, most spatial data are temporarily invariant and collected on the basis of arbitrary administrative boundaries. To better understand disease aetiology and the dynamics between individuals and environments, it would be crucial to make spatio-temporal data available. Second, the temporal dimension has not been well integrated into GIS. While some transportation geographers have begun to expand the spatio-temporal visualisation capacity (Chen et al., 2011), its development remains in its infancy. Third, following the previous two flaws, existing spatial statistical models are focused on the analysis of cross-sectional data and the analytic results may have limited implications in causality. Fourth, the progress in advanced spatial analysis methods are driven by the concept of spatial homogeneity, and little attention has been paid to spatial heterogeneity until recently (Fotheringham et al., 2002; Chen et al., 2012). Spatial heterogeneity has been suggested to provide nuanced insights in making place-specific policies and health research should begin to embrace this concept (Yang and Matthews, 2012). Fifth, though the privacy issues in GIS health research have drawn geographers' attention (Armstrong et al., 1999), no useful tool have been put forward to maintain privacy and data utility simultaneously. Future investigations into how to preserve confidentiality and statistical relationships among observation are warranted. Finally, the spatial weight matrix used in spatial analysis should be better justified or connected to health and/or epidemiological theories; in so doing, the spatial analysis results would have a broader impact on testing and reframing health theories.

The weaknesses identified above are concentrated on spatial data and the methods used to analyse them. Clearly, health research is not the only discipline that faces these challenges, and geography or geosciences should not be the only discipline that attempts to address them. Since spatial data collectors, spatial analysis methodologists, spatial analysis users and spatial tool developers come from a range of disciplines (e.g. economics, statistics, demography and computer science), seeking solutions to the challenges identified here requires multidisciplinary collaboration. For example, methodologists (e.g. statisticians) could help those depending on spatial analysis (e.g. health researchers) to resolve analytic or empirical problems. In addition, an ongoing dialogue between data collectors (e.g. federal agencies) and users from all disciplines should improve the quality and availability of health data and spatial analysis.

Acknowledgements

We would like to acknowledge the help from the Geographic Information Analysis (GIA) Core of the Population Research Institute (PRI) and the Social Science Research Institute, Penn State. PRI has received core support from the U.S. National Institute of Child Health and Human Development (R24 HD041025-06). We would also like to acknowledge the help from Family Demography Training (T-32HD007514) from the National Institute of Child Health and Human Development.

References

Armstrong MP, Rushton G, Zimmerman DL, 1999. Geographically masking health data to preserve confidentiality. Stat Med 18, 497-525.

Auchincloss AH, Gebreab SY, Mair C, Diez Roux AV, 2012. A review of spatial methods in epidemiology, 2000-2010. Annu Rev Public Health 33, 107-122.

Banerjee S, Carlin BP, Gelfand AE, 2004. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, 452 pp.

Berrigan D, McKinno RA, 2008. Built environment and health. Prev Med 47, 239-240.

Boulos MNK, Curtis AJ, AbdelMalik P, 2009. Musings on privacy issues in health research involving disaggregate geographic data about individuals. Int J Health Geogr 8, 46-53.

Burrough PA, 1986. Principles of geographical information systems for land resources assessment. Clarendon Press, 193 pp.

Chen J, Roth R, Naito A, Lengerich E, MacEachren A, 2008. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. Int J Health Geogr 7, 57-74.

Chen J, Shaw SL, Yu H, Lu F, Chai Y, Jia Q, 2011. Exploratory data analysis of activity diary data: a space-time GIS approach. J Transp Geogr 19, 394-404.

Chen YJ, Deng WS, Yang TC, Matthews SA, 2012. Geographically weighted quantile regression (GWQR): an application to US mortality data. Geogr Anal 44, 134-150.

Chen YJ, Wu PC, Yang TC, Su HJ, 2010. Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan. Sci Total Environ 408, 2042-2049.

Chung K, Yang DH, Bell R, 2004. Health and GIS: toward spatial statistical analyses. J Med Syst 28, 349-360.

Clougherty JE, Levy JI, Kubzansky LD, Ryan PB, Suglia SF, Canner MJ, Wright RJ, 2007. Synergistic effects of traffic-related air pollution and exposure to violence on urban asthma etiology. Environ Health Persp 115, 1140-1146.

Cromley EK, McLafferty SL, 2011. GIS and public health. The Guilford Press, 503 pp.

Cummins S, Curtis S, Diez-Roux AV, Macintyre S, 2007.

Understanding and representing "place" in health research: a relational approach. Soc Sci Med 65, 1825-1838.

Du P, Lemkin A, Kluhsman B, Chen J, Roth RE, MacEachren A, Meyers C, Zurlo JJ, Lengerich EJ, 2010. The roles of social domains, behavioral risk, health care resources, and *Chlamydia* in spatial clusters of US cervical cancer mortality: not all the clusters are the same. Cancer Cause Control 21, 1669-1683.

Elhorst JP, 2010. Applied spatial econometrics: raising the bar. Spat Econ Anal 5, 9-28.

Elliott P, Wartenberg D, 2004. Spatial epidemiology: current approaches and future challenges. Environ Health Persp 112, 998-1006.

Entwisle B, 2007. Putting people into place. Demography 44, 687-703.

Ertur C, Koch W, 2007. Growth, technological interdependence and spatial externalities: theory and evidence. J Appl Econ 22, 1033-1062.

ESRI, 2012. Redlands, CA. Available at: http://www.esri.com/ (accessed on August 2012).

Evans GW, Kantrowitz E, 2002. Socioeconomic status and health: the potential role of environmental risk exposure. Annu Rev Public Health 23, 303-331.

Fotheringham AS, 1997. Trends in quantitative methods I: stressing the local. Prog Hum Geog 21, 88-96.

Fotheringham AS, 2009. "The problem of spatial autocorrelation" and local spatial statistics. Geogr Anal 41, 398-403.

Fotheringham AS, Brunsdon C, Charlton ME, 2002. Geographically weighted regression: the analysis of spatially varying relationships. Wiley, 282 pp.

Gelfand AE, Banerjee S, Gamerman D, 2005. Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics 16, 465-479.

Goodchild MF, 1992. Geographical information science. Int J Geogr Inf Syst 6, 31-45.

Goodchild MF, Janelle DG, 2010. Toward critical spatial thinking in the social sciences and humanities. GeoJournal 75, 3-13.

Haining RP, 2003. Spatial data analysis: theory and practice. Cambridge University Press, 432 pp.

Han D, Matthew R, Bonner MR, Jing Nie J, Freudenheim JL, 2013. Assessing bias associated with geocoding of historical residence in epidemiology research. Geospat Health 7, 369-374.

Koch T, 2009. Social epidemiology as medical geography: back to the future. GeoJournal 74, 99-106.

Leenders RTAJ, 2002. Modeling social influence through network autocorrelation: constructing the weight matrix. Soc Networks 24, 21-47.

Leventhal T, Brooks-Gunn J, 2003. Moving to opportunity: an experimental study of neighborhood effects on mental health. Am J Public Health 93, 1576-1582.

Leventhal T, Dupéré V, 2011. Moving to opportunity: does long-term exposure to low-poverty neighborhoods make a difference for adolescents? Soc Sci Med, 73, 737-743.

Maantay J, 2002. Mapping environmental injustices: pitfalls and potential of geographic information systems in assessing environmental health and equity. Environ Health Persp 110, 161-171.

Maantay J, 2007. Asthma and air pollution in the Bronx: methodological and data considerations in using GIS for environmental justice and health research. Health Place 13, 32-56.

Macintyre S, Ellaway A, Cummins S, 2002. Place effects on health: how can we conceptualise, operationalise and measure them? Soc Sci Med 55, 125-139.

Matthews SA, Yang TC, 2010. Exploring the role of the built and social neighborhood environment in moderating stress and health. Ann Behav Med 39, 170-183.

McMaster RB, Leitner H, Sheppard E, 1997. GIS-based environmental equity and risk assessment: methodological problems and prospects. Cartogr Geogr Inform Sci 24, 172-189.

Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC, 2005. Network theory and SARS: predicting outbreak diversity. J Theor Biol 232, 71-81.

NCHS, 2010. Compressed mortality file, 1999-2007 (machine readable data file and documentation, CD-ROM series 20, No.2M). National Center for Health Statistics.

Openshaw S, 1984. Ecological fallacies and the analysis of areal census data. Environ Plann A 16, 17-31.

Páez A, Uchida T, Miyamoto K, 2002. A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. Environ Plann A 34, 733-754.

Piantadosi S, Byar DP, Green SB, 1988. The ecological fallacy. Am J Epidemiol 127, 893-904.

Raghavan RK, Brenner KM, Harrington Jr. JA, Higgins JJ, Harkin KR, 2013. Spatial scale effects in environmental risk-factor modelling for diseases. Geospat Health 7, 169-182.

Raudenbush SW, Bryk AS, 2002. Hierarchical linear models: applications and data analysis methods. Sage Publications Inc, 485 pp.

Reibel M, 2007. Geographic information systems and spatial data processing in demography: a review. Popul Res Policy Rev 26, 601-618.

Reibel M, Bufalino ME, 2005. A test of street weighted areal interpolation using geographic information systems. Environ Plann A 37, 127-139.

Reich BJ, Fuentes M, Dunson DB, 2011. Bayesian spatial quantile regression. J Am Stat Assoc 106, 6-20.

Renalds A, Smith T, Hale P, 2010. A systematic review of built environment and health. Fam Community Health 33, 68-78.

Rogers E, 2003. Diffusion of innovations. Free Press, 576 pp.

Rushton G, 2003. Public health, GIS, and spatial analytic tools. Annu Rev Publ Health 24, 43-56.

Ryan PH, LeMasters GK, Biswas P, Levin L, Hu S, Lindsey M, Bernstein DI, Lockey J, Villareal M, Hershey GKK, Grinshpun

SA, 2007. A comparison of proximity and land use regression traffic exposure models and wheezing in infants. Environ Health Persp 115, 278-284.

Salam MT, Islam T, Gilliland FD, 2008. Recent evidence for adverse effects of residential proximity to traffic sources on asthma. Curr Opin Pulm Med 14, 3-8.

Sampson RJ, Groves WB, 1989. Community structure and crime: testing social-disorganization theory. Am J Sociol 94, 744-802.

Sampson RJ, Morenoff J, Earls F, 1999. Beyond social capital: spatial dynamics of collective efficacy for children. Am Sociol Rev 64, 633-660.

Schwartz S, 1994. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. Am J Public Health 84, 819-824.

Shoff C, Yang TC, Matthews SA, 2012. What has geography got to do with it? Using GWR to explore place-specific associations with prenatal care utilization. GeoJournal 77, 331-341.

Sparks PJ, Sparks CS, 2010. An application of spatially autoregressive models to the study of US county mortality rates. Popul Space Place 16, 465-481.

Sparks PJ, Sparks CS, Campbell JJA, 2013. An application of Bayesian spatial statistical methods to the study of racial and poverty segregation and infant mortality rates in the US. GeoJournal 78, 389-405.

Star J, Estes JE, 1990. Geographic information systems: an introduction. Prentice Hall, 303 pp.

Tucker P, Irwin JD, Gilliland J, He M, Larsen K, Hess P, 2009. Environmental influences on physical activity levels in youth. Health Place 15, 357-363.

US Department of Health and Human Services, 2012. Office of Disease Prevention and Health Promotion. Healthy People 2020. Washington, DC. Available at: http://www.healthypeople.gov (accessed on August 2012).

VanWey LK, Rindfuss RR, Gutmann MP, Entwisle B, Balk D, 2005. Confidentiality and spatially explicit data: concerns and challenges. Proc Natl Acad Sci USA 102, 15337-15342.

Walford N, 2001. Geographical data: characteristics and sources. Wiley Press, 274 pp.

Wartenberg D, Thompson WD, 2010. Privacy versus public health: the impact of current confidentiality rules. Am J Public Health 100, 407-412.

Watcher KW, 2005. Spatial demography. Proc Natl Acad Sci USA 102, 15299-15300.

Weeks JR, 2004. The role of spatial analysis in demographic research. In: Spatially integrated social science. Goodchild MF, Janelle DG (eds). Oxford University Press, 381-399 pp.

WHO, 2012. Public health mapping and GIS. Available at: http://www.who.int/health_mapping/en/ (accessed on August 2012).

Wilson JP, Fotheringham AS, 2007. The handbook of geographic information science. Blackwell Publishing, 634 pp.

Yabiku ST, Glick JE, Wentz EA, Haas SA, Zhu L, 2009. Migration, health, and environment in the desert southwest. Popul Environ 30, 131-158.

Yang TC, Jensen L, Haran M, 2011. Social capital and human mortality: explaining the rural paradox with county-level mortality data. Rural Sociol 76, 347-374.

Yang TC, Matthews SA, 2012. Understanding the non-stationary associations between distrust of the health care system, health conditions, and self-rated health in the elderly: a geographically weighted regression approach. Health Place 18, 576-585.

Yang TC, Teng HW, Haran M, 2009. The impacts of social capital on infant mortality in the US: a spatial investigation. Appl Spat Anal Policy 2, 211-227.

Zimmerman DL, Armstrong MP, Rushton G, 2008. Alternative techniques for masking geographic detail to protect privacy. In: geocoding health data: the use of geographic codes in cancer prevention and control, research, and practice. Rushton G, Armstrong MP, Gittler J, Greene BR, Pavlik CE, West MM, Zimmerman DL (eds.). Taylor and Francis, 127-138 pp.

Annex

When data are geographically referenced, they can be managed with geographical information systems (GIS) and be displayed and analyzed using spatial analytic methods. Despite the various advantages of utilising data with geographical references, there are important methodological caveats that researchers should consider. First, the underlying assumptions of non-spatial statistical procedures become problematic, e.g. the standard ordinary least squares (OLS) regression framework assumes that the parameter estimates obtained from the regression model are constant over space, i.e. a one unit change in an explanatory variable would provoke the same change in a dependent variable in all data points regardless of location. The problem with this assumption is that the relationships being tested may vary spatially for various reasons (Fotheringham et al., 2002).

A second research challenge is that the statistical procedures should explicitly correct for two common problems that commonly exist in georeferenced data, namely spatial dependence and spatial heterogeneity. Spatial dependence is the existence of a functional relationship between what happens at one point in space and what happens elsewhere. If similar values of a variable cluster in space, then positive spatial autocorrelation exists. On the other hand, if locations are surrounded by neighbours with dissimilar values, then negative spatial autocorrelation exists. Spatial heterogeneity refers to as variation in relationships over space (LeSage, 1999) and it exists when the mean, variance and covariance structures drift over a mapped process. When spatial heterogeneity is present, regionally-specific circumstances influence structural relationships (O'Loughlin et al., 1994). Both spatial dependence and spatial heterogeneity in georeferenced data are problematic as they violate the independence assumption in a standard OLS regression specification, and result in a biased, if not invalid, statistical inference from the model. Ignoring the spatial structure underlying georeferenced data can result in the estimated regression coefficients being biased, inconsistent or inefficient (Voss and Curtis, 2011).

In the past, health scientists have tried to address the problems with spatial dependence and spatial heterogeneity in georeferenced data by including dummy variables in their models in an attempt to capture differences in behaviour across geographical areas. However, this simplistic approach misses some of the central issues inherent in spatial data (Darmofal, 2009). The dummy variable approach informs whether or not there is a statistically significant difference in the behaviour in various areas in comparison to a reference category area, but it cannot tell the underlying processes that result in the spatial dependence (Darmofal, 2009). Recent development in exploratory spatial data analysis techniques, such as the local indicator of spatial autocorrelation, has better allowed health researchers to explore spatial patterns in georeferenced data. A similar approach is to conduct spatial clustering analysis that identifies the areas with abnormally high and/or low events of interest (e.g. diseases). Testing for spatial autocorrelation helps the researcher to assess the potential spatial effects (i.e. spatial dependence or spatial heterogeneity) in the georeferenced data. This assessment will guide researchers to better specify the data-generating models (Voss and Curtis, 2011).

References

Darmofal D, 2009. The social sciences and spatial analysis. In: Darmofal D (ed.). Spatial analysis for the social sciences. Cambridge University Press, 3-12 pp.

Fotheringham AS, Brunsdon C, Charlton ME, 2002. Geographically weighted regression: the analysis of spatially varying relationships. Wiley, 282 p.

LeSage JP, 1999. Spatial econometrics. Citeseer, 284 p.

O'Loughlin J, Flint C, Anselin L, 1994. The geography of the nazi vote: context, confession, and class in the Reichstag election of 1930. Ann Assoc Am Geogr 84, 351-380.

Voss PR, Curtis KJ, 2011. Spatial regression modeling, NICHD supported advanced spatial analysis training program for population scientists, "spatial regression modeling" Penn State University.