Chapter 24
Simulating Spatial Health Inequalities

Dianna M. Smith

Abstract Simulation models are historically vital to epidemiology and have recently become more common in social sciences such as geography, where the aim is to understand underlying causes of population health inequalities. Traditionally the methods used to estimate health outcomes at the local level rely heavily on regression techniques and are restricted by data availability. One possible way forward is to take the best elements of current methods to model interactions between individuals at multiple scales, and extend such models to predict changes in health over time. Complex systems dynamics models and agent-based models in particular are one methodological improvement that would realise both of these goals.

24.1   Introduction

Little is known about population health at the local level. Although there is knowledge of trends towards changing health behaviours and outcomes in the national population as a result of regular surveys, the focus on health at such a coarse scale can mask local variation. For instance, the UK is known to have very high rates of child and adult obesity as a nation, but there are likely to be areas where rates are far above or below the mean values (Moon et al. 2007). Understanding how these rates might change and what might be driving the change among different populations in disparate areas are questions often asked by epidemiologists and health professionals. Social and spatial variation in health has remained a focus of health geography over the previous decades, particularly in wealthier nations as the gaps between the wealthy and poorer members of society has become clearer. Recent research addressed this issue in The Widening Gap, a book which clearly illustrated the

D.M. Smith (✉)
Department of Geography, Queen Mary, University of London, London, UK
e-mail: d.smith@qmul.ac.uk

evolving geographies of health inequalities in Britain through extensive data analysis and detailed maps of health outcomes over time and space (Shaw et al. 1999). This work and similar publications (Macintyre et al. 1993) have highlighted the importance of geography, and the potential influence of local environments, in understanding public health variation within a country.

The fascination with social and spatial variation in health has extended beyond the exploration of historical patterns to include present-day trends in non-contagious and infectious disease, health behaviours and the introduction of predictive models that aim to estimate how the spread of disease or health-related outcomes may adjust over time in relation to a changing population. The predictive aspect of many models is especially relevant in countries where health care is provided by the state rather than through individual health insurance schemes. If health professionals and policy makers have an idea of the current and future patterns in smoking, obesity, or cardiovascular disease, then they are better prepared to allocate resources to areas of greatest demand.

The utility of dynamic models is clear in the wake of recent infectious disease outbreaks across the world (H1N1, H5N1 flu) as discussed in a recent article appearing in Nature (Epstein 2009) and later in this book (Simoes 2012). Both authors show how agent-based models (ABMs) of the disease diffusion can be used to formulate policy response to current and future infectious disease outbreaks at the macro scale. This chapter will outline the development of public health models in epidemiology and the social sciences such as geography, and focus particularly on the microspatial, local-level element of any models. The current options available for static models, which estimate health characteristics of populations for one point in time, will first be outlined to give readers an overview of the various techniques and algorithms used by researchers and health organisations to model public health. The chapter concludes with a discussion of the advancement towards dynamic models, which consider population change and observed predictors of disease/behaviours to estimate future public health trends. ABMs, already suggested by epidemiologists to be the best way forward in modelling public health (Auchincloss and Diez Roux 2008), are one strong alternative to the traditional regression-based models.

24.2 Individual Level Models of Health Outcomes

Few geographical health-oriented models deal with individuals; most are prevalence models which look at aggregate local area population characteristics (from a population census) to identify the likelihood of various health outcomes occurring at the population level. Often these models are simply identifying the population-level attributes which are known to influence the disease or health outcome of interest, for instance, the risk of type 2 diabetes increases with age so is more prevalent among retirees than university students. Implicitly all of the models outlined in the following section will consider the relationship between geographic place and health, the intersection of context and composition. This place/person relationship has become a central interest in the study of spatial inequalities; if you remove people from one

environment and place them in another, will it impact their health? How can these associations be identified and quantified?

Epidemiologists and geographers have developed several modelling approaches which use some type of regression equation to derive probabilities of behaviours or disease in relation to the local population and/or environment. Most methods are a type of direct standardisation, where the predictors of a disease/outcome in a representative population (such as a national level health survey like the annual Health Survey for England [HSE] or the periodic New Zealand Health Survey [NZHS]) are statistically identified through regression analysis to estimate the likelihood of individuals with certain predictive traits to experience a health condition. The regression may take into account only the individual-level (compositional) characteristics, or it may be extended to include area-level (contextual) attributes. The probabilities created from the regression process can then be applied to the local population.

Indirect standardisation methods take the opposite approach, which is to look at the predictors of a heath outcome in a sample population such as a local health survey and apply it upwards to the national population. This is much less frequently used in spatial public health models due to the prohibitive cost associated with carrying out comprehensive local level health surveys which provide the health data for this technique.

The focus in this chapter is to consider models with finer spatial scale, and usually these models may be categorised into one of the following groups: epidemiological, synthetic population estimation (multilevel or spatial microsimulation) or Empirical Bayesian. The next section outlines these main types of static estimation models and gives examples of their application within the United Kingdom. There is an evolution from the earliest estimation models as computational power and data collection has improved, as will be shown later in this section, where the line between more traditional static models has begun to blur into the dynamic microsimulation models (Wu and Birkin 2012; Portz and Seyfried 2011) that can be seen as predecessors of ABMs.

24.2.1   Multilevel Models

One of the most inherently geographical approaches to creating local-level estimates of health outcomes or behaviours is the use of multilevel, or hierarchical models, to develop local prevalence estimates (Moon et al. 2007; Pearce et al. 2003; Twigg and Moon 2002). The structure of multilevel models is described in the name; people are 'nested' within multiple area levels, such as neighbourhoods, schools or work environments. Multilevel models have gained substantial popularity in the social sciences as they allow researchers to quantify the magnitude of the influence that place-based characteristics might have on population health. For example, how might neighbourhood deprivation affect mental well-being? (Fagg et al. 2006).

Prior to the implementation of a multilevel model, relevant predictors for the health outcome need to be identified. Each of the predictors need to be relevant to the

health outcome and present in both the survey dataset and the small-area population data (Twigg and Moon 2002). Logistic regression models are preferable in situations where the outcome is a dichotomous value (not diabetic/diabetic; non-smoker/smoker) and the predictors are either a continuous scale (such as age) or categorical (such as ethnic groups) (Gatrell 2002).

One example of the multilevel modelling framework is the creation of nationwide probabilities for smoking status based on data (age, sex, smoking status, home Output Area [OA]) from the Scottish Health Survey and the 1991 Census (Pearce et al. 2003). Each of the 13,784 respondents are grouped into 12 age-sex bands to calculate the probability of smokers in each band; the age-sex distribution is available from the 1991 Census. Using the known smokers/non-smokers and their area of residence alongside age and sex bands at the individual level, the authors were able to estimate additional data from the census about each OA's population characteristics, including 16 'person' variables (including % unemployed) and 9 'household' variables (e.g., % owner occupied households). There were also two variables at the next largest area (pseudo postcode sector), deprivation and an Office of National Statistics (ONS) Ward classification (Pearce et al. 2003).

After testing a series of multilevel models and identifying the significant variables influencing smoking at the individual and area level, the parameter estimates from the final multilevel model were used to calculate new probabilities for smoking in each of the age/sex groups, based on several new variables; these probabilities were then applied to all of the output areas (where all of the predictive variables were available) across Scotland. The results showed a wide range of smoking prevalence, but the predictor variables which proved most significant were consistent with previous studies (Pearce et al. 2003). The combination of small-area data with survey responses is very similar to the epidemiological modelling approach; however, the multilevel framework allows researchers to clearly identify significant predictors at more than one scale. In addition, the inclusion of cross-level interactions between predictor variables adds greater accuracy to the resulting estimates (Twigg and Moon 2002).

One limitation of the multilevel modelling framework for the creation of prevalence estimates is the need for data on predictor variables to be available at the geographic scale for the resulting estimates. As will be explained later in this section, spatial microsimulation techniques are not as limited by data to create estimates. Where the multilevel modelling approach assigns the parameter estimates to individuals matching a multifaceted profile (for instance, white males aged 30–39 years in social class AB), the microsimulation method assigns probabilites for behaviours iteratively to each of the four attributes in turn (ethnicity, sex, age and social class).

The results from this prevalence estimation approach can be tested for accuracy by comparing the outputs against known local-level surveys. Previous results have indicated that the method is quite robust when used for tobacco smoking estimation, although less reliable in accurately predicting alcohol consumption (Twigg and Moon 2002).

24.2.2   *Epidemiological Models*

The primary difference between epidemiological models and the alternative options for static estimation processes is the use of multiple datasets to generate probabilities. The challenge with this type of model is that the user is limited to only estimate outcomes on typically small-scale studies for derivation of reference rates. However, one particular application of this method has been used extensively by the National Health Service (NHS) as a way of firstly estimating the national (English) prevalence of type 2 diabetes, which is often undiagnosed, and to also create these estimates at a more local level (Forouhi et al. 2006). Because these models are dependent on relatively small local surveys, they may use data from sources that are far apart in time and place. In the case of the model of Forouhi et al. (2006) model, they used reference rates based on age, sex and ethnicity from six datasets ranging from 1986 to 2000. The authors created a set of time and place adjustments to correct for differences between the study populations and locations.

Once the reference rates are created from the epidemiological datasets, they can be applied to crosstabulated 2001 Census data (age-sex-ethnicity) at the smallest area level where such crosstabulations are available. The benefit of this modelling approach is that all the data are based on a variety of real-world datasets. However, users are constrained by the need for crosstabulated census data to build up the estimates. In the case of diabetes, the lack of data flexibility meant that socioeconomic status was not used as a predictor in the model, although this variable is known to influence diabetes incidence (Connolly et al. 2000; Evans et al. 2000). Unlike the multilevel modelling framework, there is no scope for adding area-level predictors such as land use mix.

A different type of synthetic population estimation similar to the epidemiological models described above, but with greater flexibility in how the predictor variables are included in the model, is through the incorporation of Bayesian estimates. As with the epidemiological method, the models are designed to be used at a scale where crosstabulations of the necessary attributes are available. This method has been used to estimate coronary heart disease (CHD) and diabetes in England (Congdon 2006, 2008).

The diabetes estimates created in this way are similar to the epidemiological model described above, but the initial data come from the 1999 and 2003 HSE to calculate age by sex by ethnic group specific prevalence rates for both type 1 and 2 diabetes. The estimated rates are then applied to the 2001 Census wards, where the age-sex-ethnic group population distributions are known. The Bayesian methods employed by Congdon include a 1999 diabetes risk factor to create accurate predictions of diabetes prevalence and confirm the probabilities for diabetes created from the regression of 2003 data. There is significant overlap in the modelling techniques between Congdon's model and those implemented using a multilevel approach or epidemiological method.

24.2.3   Microsimulation

Spatial microsimulation techniques offer the ability to link non-spatial datasets, such as national health surveys, with spatial data such as sociodemographic attributes from the population census. Unlike the other approaches, the microsimulation model is not dependent on having cross-tabulated data at each area level where the estimates are being created. Instead, the purpose of spatial microsimulation is to iteratively replicate known characteristics of the population which predict the health outcome of interest reliably at the local level. There are several different computational algorithms for spatial microsimulation, which are outlined elsewhere in this book (Birkin and Wu 2012). Deterministic reweighting has been used in a suite of models for health behaviours and outcomes including smoking, diabetes and obesity (Tomintz et al. 2008; Smith 2007). Other options include stochastic methods such as simulated annealing and combinatorial optimization. With deterministic reweighting, a probability for each person who responded to the non-spatial survey to live in each local area is calculated, based on a reweighting algorithm that takes each of the predictive variables in turn (Smith et al. 2009; Ballas et al. 2006). The sums of all the probabilities generated for each area will add up to the census-based population total. These probabilities can be used to generate prevalence estimates as they will give an indication of the proportion of the population affected by the health outcome/behaviour.

   Where microsimulation differs strongly from the alternative methods outlined above is that multiple outcomes or behaviours may be estimated for a local population at one time rather than creating a series of outcome-specific models which have to be re-run for every desired characteristic. For example, if the prevalence rates of adult obesity and type 2 diabetes were created using a multilevel modelling framework, this would require two separate modelling runs for each health condition rather than only one with spatial microsimulation. However, the lack of specificity in the synthetic population creation from microsimulation may mean that resulting estimates are not as accurate as alternative methods because different health conditions may be best predicted by very different sociodemographic characteristics. The predictors of smoking behaviour and high levels of physical activity are quite different, so it is unlikely that one model might provide the most accurate estimation of both outcomes. If the conditions are predicted by similar characteristics, such as with obesity and diabetes, then the use of one model is appropriate.

   As with the other static prevalence models, validation of prevalence estimates is difficult due to the lack of real-world data. Options to test the reliability of the model predictions can include comparing the model estimates against a related outcome with known prevalence at the same scale, or aggregating the estimates up to a geography where the prevalence is known (Tomintz et al. 2008; Congdon 2008). All of the models are only estimating health based on observed relationships between the modelled health outcome and the local populations' sociodemographic profile that is associated with that outcome (Moon et al. 2007).

   Static models, like the dynamic models described in the next section, are limited by the available data that can be included in them. One of the biggest challenges

with any type of prevalence estimation is the use of older data for the baseline population (to include attributes such as age, sex, ethnicity). The UK census of population takes place every 10 years but the tables with population characteristics are not immediately available for researchers, so the models are never based on real-time population characteristics. Depending on the country, the larger/national health surveys, which can be used to create the estimates, may not be collected every year; the Health Survey for England is annual but the Scottish Health Survey has only recently been conducted each year (2008 through 2011).

24.2.4   Dynamic Models

Dynamic models attempt to create health outcomes not only for one point in time but also for future populations, by taking into account potential population changes such as an aging population. The utility of predictive models for future planning is particularly important for countries where health care is funded by the government and future budgets must be allocated in advance. Dynamic models can take the form of the regression analysis described above (dynamic microsimulation) or may be based on more intricate relationships, like the complex systems dynamics models which consider individual and environmental level interactions. The systems dynamics models are iterative in nature, building on the baseline data and creating new data as the populations evolve and interact; one specific example of this type of model is an ABM.

24.2.5   Dynamic Microsimulation Modelling

Dynamic microsimulation modelling is described in detail elsewhere in this book (Birkin and Wu 2012). Briefly, this method is an advance beyond the simple static models outlined earlier, often including a stochastic element to the population generation process. Similarly to the static models, health outcomes are estimated based on previous observed associations with demographic characteristics in a type of regression analysis. However, with the dynamic models, the baseline populations are allowed to change in line with expected demographic evolution within an area. For example, aging populations or migration of different ethnic groups between areas will affect the model's estimated outcomes, as will possible changes to government policy related to the behaviour, such as tobacco taxation and smoking policies.

Dynamic microsimulation models have already been created to estimate the future prevalence of obesity (Kopelman et al. 2007). However, the models are still constrained by linear relationships defined by regression analysis. Using obesity prevalence as an example, these models may fail to accurately represent how real people would react to a variety of influences such as less expensive food, better access to fitness facilities, or increased education about the risks associated with obesity.

24.2.6 Complex Systems Dynamics Models

Newer methods in disease estimation approach the health outcome as a complex system, with the aim of including as many potential influences as possible. Gatrell has recently acknowledged the difficulty of accurately modelling health outcomes using predictive models. However, the means of dealing with complexity is not strongly developed in public health applications (Gatrell 2005). Many of the issues raised by Gatrell are intuitive: the inability of the models to account for interactions between variables (beyond the simplistic methods in multilevel models); the simplified, linear nature of the models that are unable to account for non-linear relationships (which, arguably, are widespread in health research); the inherently complex nature of relationships between people and place; the idea of epidemiology as a 'web' of inter-connected mechanisms, which uniquely combine in individual lives (Gatrell 2005).

A recent issue of the American Journal of Public Health was devoted to exploring potential approaches to modelling complex systems, with several authors who echo Gatrell's call for improved models. One of the models, developed in the United States to estimate the impact of various governmental policy on diabetes prevalence, is created using systems dynamics (Jones et al. 2006). This is perhaps the closest that researchers have come to acknowledging the true complexity of public health. However, the model is currently only feasible at the national scale. This particular model, created by health planners at the Centers for Disease Control in the United States, was designed specifically to understand population dynamics related to diabetes. The intention was to inform public health strategy by predicting the future prevalence of diabetes through 2050. The model incorporated factors such as death rates, health insurance, diabetes diagnosis and medication. This model, along with others currently under development at the CDC, promises to improve health planning by better predicting the effects of interventions on public health (Jones et al. 2006).

There appears to be a trade-off in terms of the level of complexity allowed in a model and the unit of geographical analysis for which it can estimate disease prevalence. As complexity studies continue to gain momentum (and computational powers increase), this 'choice' may be resolved, leading to more robust models which can accurately depict current and future health trends at a finer spatial scale.

Complex systems dynamic models are a general category of advanced simulation models that includes ABMs (see Crooks and Heppenstall 2012). The benefit of this family of models is their ability to incorporate multiple scales of influence (like a multilevel model) as well as considering the changing relationships between influences on agents' health within the model. The inherent complexity in person-environment interactions is best modelled using this type of approach because the agents (people) in the model are allowed to react to changes in causal factors for disease from the local environment or each other. The environment may not be such an obvious causal factor in non-communicable disease as it is for illness such as malaria or Dengue fever, but much of the recent work that aims to investigate the

increasing trends in obesity suggests that a person's local environment plays a key role (Egger and Swinburn 1997).

An ABM has the unique ability to combine multiple scales/types of influence as well as interactions and feedback loops to ideally replicate interactions that cannot be represented in regression-based models (Auchincloss and Diez Roux 2008). The dynamic nature of ABMs is a great asset to health planning; people vary over time and are influenced by any number of factors at different ages, and this method is the best way to address such complexity. 'Agents' in the models do not necessarily have to be individuals but this is the most common configuration. Attributes and behavioural rules are assigned to the agents based on available data (commercial data, qualitative studies) to begin the simulation, and there is the option to add a random element to the evolving interactions that will dictate how the agents may respond to different situations. The model is then run numerous times to generate a variety of outcomes (Auchincloss and Diez Roux 2008). The model is usually created in a computer programming language like Java, but there are several ready-made programmes such as Recursive Porous Agent Simulation Toolkit (REPAST) that may be adapted by individual users who have less programming experience.

Obesity is a good example of how ABMs can move the epidemiological research forward (Galea et al. 2009). With the wealth of research devoted to studying obesity-promoting (obesogenic) influences at the personal and area level, the complexity of obesity aetiology is well documented (Kopelman et al. 2007). A recent example of agent-based modelling of BMI with regards to local stores and varied strength of an individual's social networks gave one illustration of a possible policy scenario (Galea et al. 2009). In this simplistic model, created in a ready-made ABM framework, the results suggested that people with weaker social network ties had a greater decrease in BMI. However, they were also more likely to have an increase from baseline BMI after the food stores had returned to normal.

A more ABM to predict the evolution of BMI at local levels would likely incorporate much more data. A good basis for a comprehensive ABM for obesity would be to include the obesogenic environment framework outlined by Egger and Swinburn (1997). Their ecological model of obesity breaks the 'environment' into four distinct types (physical, economic, political and sociocultural) and further subdivides these types into the micro (i.e., neighbourhoods, schools, homes) and macro (transport, health regulatory system). Then the individual factors could be introduced in the model (age, sex, ethnicity, social class, marital status, educational attainment, etc.). All of these individual attributes and their relative importance in predicting obesity may be identified from the same types of surveys used to inform the regression-based models. It would be best to isolate aspects of the different influences to understand the relative importance of certain parameters on different people. For instance, women may be less likely to use parks for physical activity than men, or men may make less healthy choices with regards to available food.

ABMs, as one of the complex systems dynamics models, are clearly a big step forward for epidemiological research. However, as with all methods, there are limitations to be considered. The rules that govern agent behaviour are often influenced

by the assumptions of the researchers creating the model, or may be overly simplistic. The parameters that are included in the model may not be based on large samples of observed data, particularly with regards to interactions (Galea et al. 2009).

24.3 Conclusion

Increasing computational power has changed the available methods and allowed for the evolution of complex models to more accurately capture the behaviours that contribute to health outcomes. While early prevalence models were restricted in power to a static population, the new developments in systems dynamic models and agent-based modelling have led to more flexible and powerful choices for social scientists and policy analysis.

As discussed elsewhere in this book, advanced computational methods are valuable in predicting the spread of infectious disease and have historically been used by many governments and health organisations to this end. The increasing ability to capture population health dynamics for non-communicable disease may have a significant role in protecting public health in the future as limited funds and resources may be allocated to areas of greatest need. Alternatively, the models will enable users to test the efficacy of various policies to reduce the prevalence of tobacco use, binge drinking or obesity among heterogeneous populations in disparate areas.

Although there are clear challenges to the use of a systems dynamic or agent-based approach to the simulation of population-level spatial health outcomes, the advancement beyond regression based models is a significant addition to the toolbox available for public health and social science. With careful consideration for the data included in the models, including rules of behaviour for the agents, ABMs provide a great improvement from previous methods that took little or no account of individual variation and interactions (Galea et al. 2009). Researchers are encouraged to be aware of limitations to this method. As with any new approach, the outputs must be interpreted with an understanding of the underlying processes that are used to generate them. However, the shift towards complex systems dynamics modelling is a move towards true individual-based modelling in non-infectious epidemiology.

References

Auchincloss, A. H., & Diez Roux, A. V. (2008). A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health. *American Journal of Epidemiology, 168*(1), 1–8.

Ballas, D., Clarke, G., Dorling, D., Rigby, J., & Wheeler, B. (2006). Using geographical information systems and spatial microsimulation for the analysis of health inequalities. *Health Informatics Journal, 12*, 65–79.

Birkin, M., & Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 51–68). Dordrecht: Springer.

Congdon, P. (2006). Estimating diabetes prevalence by small area in England. *Journal of Public Health, 28*, 71–81.

Congdon, P. (2008). Estimating CHD prevalence by small area: Integrating information from health surveys and area mortality. *Health & Place, 14*, 59–75.

Connolly, V., Unwin, N., Sherriff, P., Bilous, R., & Kelly, W. (2000). Diabetes prevalence and socioeconomic status: A population based study showing increased prevalence of Type 2 diabetes mellitus in deprived areas. *Journal of Epidemiology and Community Health, 54*, 173–177.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Egger, G., & Swinburn, B. (1997). An 'ecological' approach to the obesity pandemic. *British Medical Journal, 315*, 477–480.

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature, 460*, 687.

Evans, J. M. M., Newton, R. W., Ruta, D. A., McDonal, T. M., & Morris, A. D. (2000). Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus. *Diabetic Medicine, 17*, 478–480.

Fagg, J., Curtis, S., Stansfeld, S., & Congdon, P. (2006). Psychological distress among adolescents, and its relationship to individual, family and area characteristics in East London. *Social Science & Medicine, 63*, 636–648.

Forouhi, N. G., Merrick, D., Goyder, E., Ferguson, B. A., Abbas, J., Lachowycz, K., & Wild, S. H. (2006). Diabetes prevalence in England, 2001 – estimates from an epidemiological model. *Diabetic Medicine, 23*, 189–197.

Galea, S., Riddle, M., & Kaplan, G. A. (2009). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology*. doi:10.1093/ije/dyp296.

Gatrell, A. C. (2002). *Geographies of health: An introduction*. Oxford: Blackwell.

Gatrell, A. C. (2005). Complexity theory and geographies of health: a critical assessment. *Social Science & Medicine, 60*, 2661–2671. Burning issues – Selected papers from the 10th international symposium in medical geography, Manchester, 2003.

Jones, A. P., Homer, J. B., Murphy, D. L., Essien, J. D. K., Milstein, B., & Seville, D. A. (2006). Understanding diabetes population dynamics through simulation modeling and experimentation. *American Journal of Public Health, 96*, 488–495.

Kopelman, P., Jebb, S. A., & Butland, B. (2007). Executive summary: FORESIGHT tackling obesities: Future choices project. *Obesity Reviews, 8*, vi–ix.

Macintyre, S., Maciver, S., & Sooman, A. (1993). Area, class and health: Should we be focusing on places or people? *Journal of Social Policy, 22*, 213–234.

Moon, G., Quarendon, G., Barnard, S., Twigg, L., & Blyth, B. (2007). Fat nation: Deciphering the distinctive geographies of obesity in England. *Social Science & Medicine, 65*, 25–31.

Pearce, J., Boyle, P., & Flowerdew, R. (2003). Predicting smoking behaviour in census output areas across Scotland. *Health & Place, 9*, 139–149.

Portz, A., & Seyfried, A. (2011). Analyzing Stop-and-Go Waves by Experiment and Modeling in R.D. Peacock, E.D. Kuligowski, J.D. Averill, (eds) Pedestrian and Evacuation Dynamics, Springer, pp 577–586

Shaw, M., Dorling, D., Gordon, D., & Davey Smith, G. (1999). *The widening gap: Health inequalities and policy in Britain* (Studies in poverty, inequality and social exclusion). Bristol: Policy Press.

Simoes, J. (2012). *An agent-based approach to spatial epidemics through GIS*. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 591–610). Dordrecht: Springer.

Smith, D. M. (2007). *Potential health implications of food access*. Unpublished PhD thesis, School of Geography, University of Leeds, Leeds.

Smith, D. M., Clarke, G. P., & Harland, K. (2009). Improving the synthetic data generation process in spatial microsimulation models. *Environment and Planning A, 41*, 1251–1268.

Tomintz, M. N., Clarke, G. P., & Rigby, J. E. (2008). The geography of smoking in Leeds: Estimating individual smoking rates and the implications for the location of stop smoking services. *Area, 40*, 341–353.

Twigg, L., & Moon, G. (2002). Predicting small area health-related behaviour: A comparison of multilevel synthetic estimation and local survey data. *Social Science & Medicine, 54*, 931–937.

Wu, B. M., & Birkin, M. H. (2012). Agent-based extensions to a spatial microsimulation model of demographic change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 347–360). Dordrecht: Springer.