Schmollers Jahrbuch 128 (2008), 109 – 120
Duncker & Humblot, Berlin

Neighbourhood Effects on Health: A Structural Equation Modelling Approach

By Mai Stafford, Amanda Sacker, Anne Ellaway, Steven Cummins, Dick Wiggins, and Sally Macintyre

Abstract

Many studies document associations between area deprivation and health but the explanatory pathways linking deprivation to health are not clear. Potential neighbourhood determinants of health include socio-relational characteristics, the built environment and neighbourhood amenities. Using obesity as an example, we theorised a model of the potential causal pathways linking neighbourhood characteristics, through diet and physical activity, to obesity. A structural equation modelling approach was used to test the model empirically using health data from national surveys in England and Scotland. The advantages and limitations of structural equation modelling are discussed and we contend that the approach provides a useful way of combining data from several sources to test theorised explanatory models linking the neighbourhood to health.

JEL Classifications: C3, I18

1. Background

Several recent studies have shown that morbidity and mortality varies across different kinds of neighbourhoods (Duncan et al., 1995; Wiggins et al., 1998). Neighbourhood deprivation is one exposure that has been strongly and consistently related to several health outcomes (Diez-Roux et al., 1997; Brooks-Gunn et al., 1998; Galea et al., 2007; Winkleby et al., 2007). Many of the more recent studies are based on multilevel data and analysis combining information on individual and area level determinants of health. These are able to test for contextual explanations (due to something about the neighbourhood itself) versus compositional explanations (due to the characteristics of the resident population) for neighbourhood inequalities. Reviews of these multilevel studies generally indicate that area deprivation is related to health, although the relationship is weaker and smaller in magnitude than the relationships between well-established *individual* socioeconomic factors and health (Pickett / Pearl, 2001; Riva et al., 2007).

Area deprivation is typically captured using summary indicators of multiple deprivation based on census characteristics (Townsend et al., 1988; Carstairs / Morris, 1991). Multiple deprivation is a marker for features of the local area that are plausibly causally related to health. However, neighbourhood effects largely remain a "black box of somewhat mystical influences on health" (Macintyre et al., 2002). These studies have contributed to a rediscovery of the importance of the (local) social environment for health (Krieger, 1994; Diez Roux, 1998) but they leave unanswered questions about precisely what it is about neighbourhoods that are important and where efforts to improve neighbourhoods should be focused. Recently, researchers have proposed theoretical frameworks (Macintyre et al., 2002; Bernard et al., 2007; Franzini et al., 2005) and empirical work has begun to explore neighbourhood socio-relational characteristics, urban design and services as potential explanatory factors linking neighbourhood deprivation to health outcomes. However, a number of limitations remain. Firstly, empirical research has tended to use only a small number of indicators to characterise neighbourhoods. Secondly, studies have sometimes used summary indices whereby different neighbourhood problems and / or resources are counted and combined into a single score. This limits the ability to distinguish the particular neighbourhood characteristics that are salient for a given health outcome. A more detailed understanding of the specific neighbourhood determinants is required before suitable interventions can be devised. Thirdly, an alternative approach has been to estimate the effect of one neighbourhood exposure whilst statistically controlling for others using multiple regression methods. This can be problematic when neighbourhood characteristics are highly correlated with each other, as is often the case.

2. Structural Equation Modelling Approach

Accordingly, we know very little about how the various contextual domains relate to each other and how they jointly influence health. A structural equation modelling approach is a useful tool to overcome some of these limitations and has several advantages. It goes beyond simple description of the association between a given contextual characteristic and health in that it allows the researcher to develop a theoretical pathway from a given characteristic, through other contextual characteristics and individual responses, to health status. The method forces the researcher to be explicit about their theoretical model. It can be used to test alternative pathways so that a researcher can discount those which do not accord with the data. This allows an iterative refinement towards a final working model. It may also help the researcher and reader to identify important variables which have so far been omitted from the model.

A structural equation modelling approach with latent variables is an efficient way of combining data from several sources. Structural equation modelling is based on the conception that the measured variables are indicators of an underlying unobserved (latent) construct. Rather than estimating the relationships between the measured variables, the relationships between the underlying constructs are estimated. The extent to which the observed indicators capture this underlying construct is assessed using confirmatory factor analysis in a measurement model (Lawley / Maxwell, 1971). Structural equation modeling improves upon simply combining variables into a summary index as it allows measurement error to be taken into account and for data from different sources to be weighted for importance as indicators of the underlying factor.

Developments in statistical software mean that path models incorporating categorical outcomes, continuous outcomes or a mixture of the two can be estimated. Of particular relevance for neighbourhood effects studies is the ability to model complex variation arising from the geographical clustering of the data. This is now possible using standard structural equation modelling software.

Structural equation modelling proceeds in stages. The first stage tests the measurement properties of the underlying latent variables in the model using confirmatory factor analysis. At the second stage, the researcher states how they think the constructs of interest are related to each other. The constructs may all be measured by latent variables, by observed variables or by a combination of the two. A path diagram can help here (Figure 1). Based on these theorized inter-relationships, the expected variances and covariances implied by the model can be calculated. These are then compared with the population variances and covariances estimated from the empirical data. The goodness of fit of the empirical data to the theorized model is then assessed. Several goodness of fit indices can be used (Bollen, 1989). Here we use the Chi-squared test of model fit and the Root Mean Square Error of Approximation (RMSEA). As a guideline, a Chi-squared statistic of the same order of magnitude as the degrees of freedom indicates a well-fitting model and an RMSEA < 0.05 indicates a model with a good approximation to the data.

3. An Example: Neighbourhood Effects on Obesity

Here we illustrate the use of structural equation modelling in studying the relationships between a range of neighbourhood characteristics and obesity. Physical activity level and dietary intake are two key determinants of obesity. There is growing interest in the importance of the local residential (and wider) context for physical activity and diet. Although individual factors play a part, the recent and rapid increase in obesity in many developed and developing countries (Hedley et al., 2004) may also be attributed to "obesogenic environments" (Swinburn et al., 1999; Hill & Peters, 1998; Poston & Foreyt, 1999). This includes the availability of energy-dense food, food marketing, increasing car use and car-friendly street layouts (Egger & Swinburn, 1997).

This work is based on research published elsewhere and the reader is referred there for a full discussion of the literature linking features of the neighbourhood environment to obesity (Stafford et al., 2007). To summarise, based on a review of the literature, we developed a theoretical causal model linking socio-relational characteristics, the built environment and local services / amenities to obesity (Figure 1). This provides the theoretical model for the analyses presented here. Such a model can be drawn for any given health outcome although the relevant neighbourhood determinants and their relative importance are likely to vary according to the outcome under study.

4. Study Areas and Study Participants

Individual level data for the study come from the Health Survey for England (HSE) and the Scottish Health Survey (SHS). These are on-going cross-sectional surveys which provide a sample that is representative of the general population of England and Scotland. A multi-stage selection procedure is used to select a random sample of postcode sectors (average population about 5000) and, within each postcode sector, a random sample of approximately 19 households (Erens / Primatesta, 1998; Shaw et al., 2000). Data from 1994 – 1999 were combined to maximize sample size and provide health data on 438 postcode sectors for the present study. Postcode sectors were used to define neighbourhood boundaries.

The HSE and SHS surveys provide information on obesity and important individual level determinants of health. Data on height and weight were measured by a trained nurse in the participant's home. Body mass index (BMI) captures overall obesity and was calculated as weight (kg) divided by the square of the height (metres). Data on physical activity and diet were unfortunately not available from the HSE and SHS. Instead, we used neighbourhood average sports participation as a proxy for individual level sports participation. Sex, age and social class based on occupation were obtained by face-to-face interview. Social class was classified into 6 groups according to the Registrar General's classification.

A total of 6848 participants had complete data on neighbourhood characteristics, BMI, age, sex and social class. Analyses based on cases with some missing data did not differ substantially from those with complete cases so, for brevity, only the latter are presented. The sample comprised adults aged 16 plus, 14 % of whom were 65 and over. There were similar proportions of women and men and 48 % were classed as manual workers. Mean BMI was 26.3 kg / m^2.

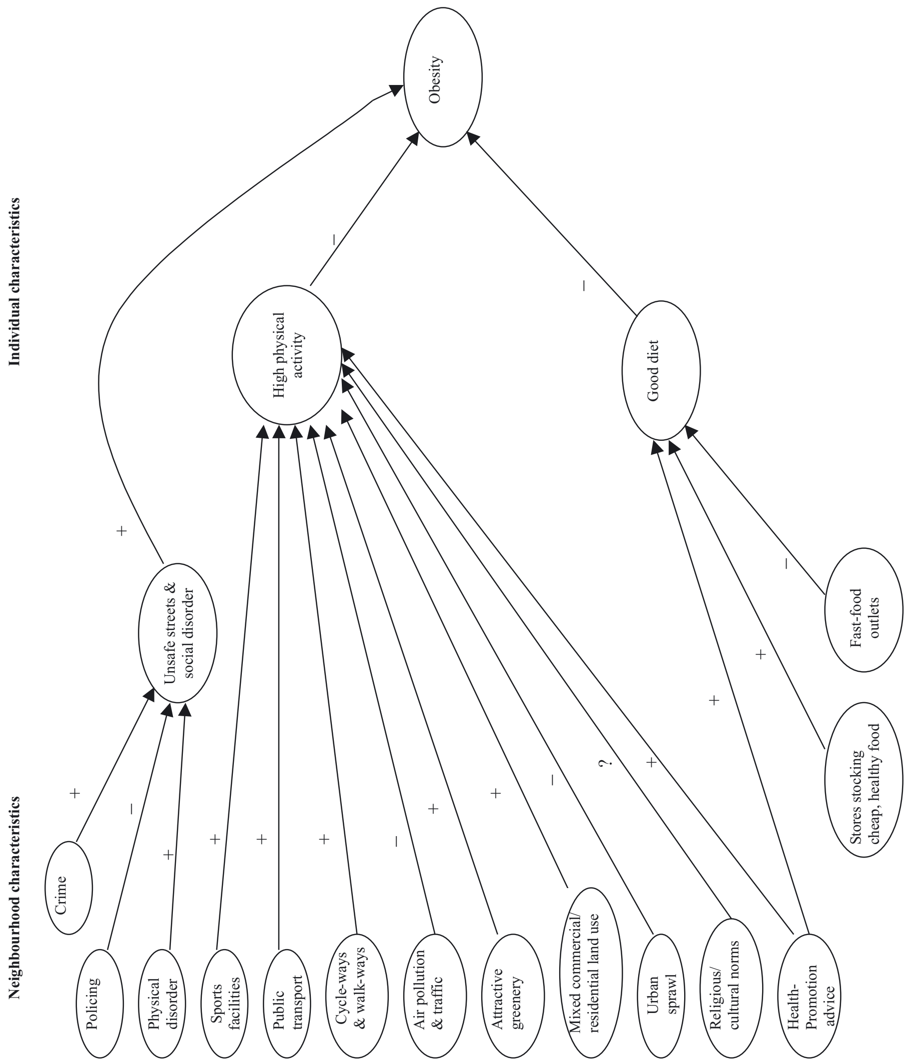


Figure 1: Theoretical model linking the neighbourhood environment to diet, physical activity and obesity

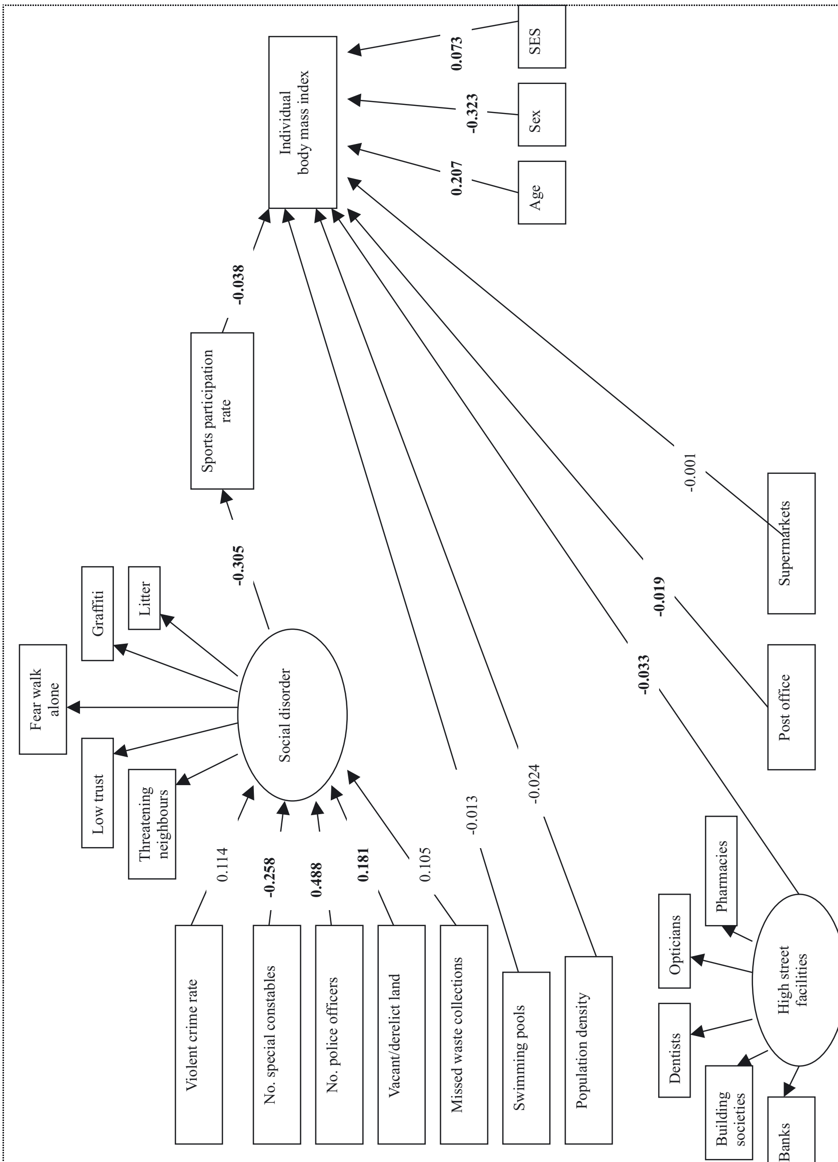
5. Neighbourhood Level Data

Information on local services and the built environment was obtained from central government departments, local authorities, voluntary and public sector agencies, commercial and industrial organizations. Data were collected between 2000 and 2001 at various spatial scales and were converted to postcode sector level for the present analysis. A full outline of the methods used and data collected can be found elsewhere (Cummins et al., 2005).

Information on socio-relational characteristics of the neighbourhood was not available from the sources listed above and so a tailor-made postal survey was administered. The postal survey was sent to a random sample of residents aged 16 and over living in study neighbourhoods. An average of 28 respondents in each of the 438 neighbourhoods provided data on neighbourhood disorder (response rate 42 %). Within each neighbourhood, responses to the questionnaire were aggregated by taking the mean score for all respondents. More detail on the design and validation of the survey is given elsewhere (Stafford et al., 2003).

Variables capturing the following neighbourhood domains were included in this study: Social disorder; Local high street services; Crime; Policing; Physical dereliction; Leisure centres; Supermarkets; Fast-food outlets; Urban sprawl. Latent variables captured the first two of these domains. Social disorder was captured by five items from the postal survey ("Neighbours are threatening"; "Most people in this area can't be trusted"; "People would be afraid to walk alone in this area after dark"; "Vandalism and graffiti are a big problem in this area"; "This area is always full of litter and rubbish"). Local high street services was captured using five items (Number of banks; Number of building societies; Number of dentists; Number of opticians; Number of pharmacies).

Data from the various sources were combined into a single dataset for analysis leading to the formulation of the model to be tested empirically (Figure 2). A number of differences between this model and the theoretical model laid out in Figure 1 should be noted. Firstly, not all of the neighbourhood characteristics could be captured by the available data. Information on neighbourhood provision of public transport, cycle-ways and walk-ways, attractive greenery, religious and cultural norms and health promotion advice could not be found within the constraints of the project. (Physical visits to all the study sites by a trained observer were not feasible.) Secondly, the actual measures capturing some of the neighbourhood characteristics were far from ideal. In particular, sports facilities were measured using number of local authority swimming pools (so that other types of public and private sports centres were not included) and fast-food outlets were measured using MacDonald's restaurants (thus omitting all other fast-food chains and independents). Note also that the counting of numbers of amenities and services says nothing



Numbers on each arrow are standardised path coefficients. **Bold** indicates statistical significance at the 5 % level.

Figure 2: Structural model linking neighbourhood characteristics to body mass index

about their quality or appropriateness for the needs of the local community. Even where data were a reasonable approximation to the theorized characteristic, questions over its accuracy across all the study sites remain since the data were not originally collected for research purposes. Finally, the spatial unit of data collection was usually not the postcode sector level and so further manipulation had to be undertaken to convert the data to this unit. For example, crime rate was measured at the local authority level, which has an average population of about 125,000. Here we assigned the same crime rate to each postcode sector within a local authority. This introduces additional error into the measurement of crime. Since this error is not systematic, it likely biases the estimate of the association between crime rate and obesity towards the null hypothesis of no association.

Robust maximum likelihood estimation methods were used to handle the clustering of participants within postcode sectors and the non-normality of several of the variables. Standardised factor loadings and path coefficients are presented throughout. Mplus software was used for all the modelling (Muthen / Muthen, 2005).

6. Results: Assessing the Full Structural Equation Model

The full structural model capturing relationships between the various neighbourhood characteristics and BMI is presented in Figure 2. The data fitted the theorized model well (Chi-squared test of model fit 360.7 on 171 degrees of freedom, RMSEA 0.013). Starting with the most proximate contributors to obesity, resident's BMI was positively associated with age, negatively associated with socioeconomic position and was higher for women. BMI was negatively associated with average sports participation rate. Average sports participation was used here as a proxy for individual sports participation: people living in high participation neighbourhoods had lower BMI, as expected. High street facilities and proximity to a post office were negatively associated with BMI. The associations between the number of swimming pools, number of local supermarkets and population density and BMI did not achieve statistical significance but were in the expected direction. (A model including the presence of McDonald's restaurants did not converge so this variable was dropped from the analysis. Simple correlation analysis indicated very low correlation between BMI and McDonald's restaurants so its omission from the final model is unlikely to have a substantial impact on other estimates.) There was no direct association between neighbourhood disorder and BMI. However, there was an indirect association. BMI was linked to disorder through average sports participation rate (indirect path coefficient 0.013 $p < 0.05$). In other words, neighbourhood disorder influences BMI because it is associated with lower participation in sporting activities.

Moving along the explanatory pathway, higher levels of neighbourhood disorder were seen in places with higher crime rates. Neighbourhoods with greater numbers of special constables had lower levels of social disorder. Number of regular police officers was also associated with disorder although the direction of the association was counterintuitive. A positive association between signs of physical dereliction (viz vacant and derelict land) and neighbourhood disorder was found.

7. Strengths and Limitations of the Study

A primary limitation is the mismatch between the idealized and available dataset, as discussed earlier. Despite these data limitations, we were able to gather a large amount of data relating to the neighbourhood from several sources. We used a latent variable approach to combine data from these disparate sources where they were thought to capture one underlying factor. This improves upon simply combining them into a summary index as it allows measurement error to be taken into account and for data from different sources to be weighted for importance as indicators of the underlying factor.

A structural equation modeling approach was then used to test theorized pathways linking the latent variables, directly measured variables and obesity. We cannot make claims about causality based on observational, cross-sectional data. However, we contend that this approach takes us a step forward in understanding the complex interplay between different facets of the neighbourhood and health.

Nevertheless, some limitations with structural equation modeling must be acknowledged. Structural equation modeling compares the empirical data with a given theorised model but it is possible that two or more different formulations of a model can be supported by the data. In other words, the solution is not necessarily a unique one. The choice of model specification should be guided as much by theory, plausibility and expert knowledge of the subject as by statistical considerations. Secondly, structural equation modeling usually proceeds in an iterative way whereby the researcher sets out their initial model, tests this against the data, refines the model and tests against the data again. Most structural equation modeling software packages produce modification indices which guide the researcher in how to modify the model to ensure a better fit to the data. In principle, the refinements should be tested on a different dataset. In practice, the researcher often uses the same set of data to test the first and subsequent formulations of the model, thus generating a potentially spuriously well-fitting model.

In common with most other studies in the field, other limitations include the use of administrative boundaries to approximate neighbourhoods, the lack of longitudinal data and the endogeneity problem. The latter problem of self-selection of obese people into certain types of neighbourhood could invoke a spurious relationship between neighbourhood characteristics and BMI. However, we have adjusted for personal socioeconomic position and this is a key determinant of a person's ability to choose where to live. In our view, it seems implausible that a large section (the overweight and obese section) of the sample would be choosing to live in places characterised by social disorder, less green space and fewer supermarkets / other places to buy healthy food. Unfortunately, longitudinal data that would allow us to investigate the extent of this problem and to assess the temporality of the association between neighbourhood environment and health and the obesity levels of movers and non-movers were not available. Future studies should aim to incorporate longitudinal data to discount the possibility that the observed relationship between neighbourhood environment and BMI is due to obese people having recently moved to obesogenic neighbourhoods.

8. Concluding Remarks

In summary, the study highlights the importance of the neighbourhood environment, including socio-relational characteristics, the built environment and amenities, for health. Efforts to tackle obesity should incorporate strategies which are cognizant of its wider, contextual determinants. Structural equation modeling is a useful approach for studying neighbourhood effects. Whilst it cannot be used to attribute causality, we contend that it is one step on the way to understanding the causal, multilevel processes between neighbourhood environments and individual's health.

References

Bernard, P. / *Charafeddine,* R. / Frohlich, K. L. / *Daniel,* M. / *Kestens,* Y. / *Potvin,* L. (2007): Health inequalities and place: a theoretical conception of neighbourhood, Social Science & Medicine 65 (9), 1839 – 52.

Bollen, K. A. (1989): Structural equations with latent variables, New York.

Brooks-Gunn, J. / *McCormick,* M. C. / *Klebanov,* P. K. / *McCarton,* C. (1998): Health care use of 3-year-old low birth weight premature children: effects of family and neighborhood poverty, Journal of Pediatrics 132 (6), 971 – 5.

Carstairs, V. / *Morris*, R. (1991): Deprivation and Health in Scotland, Aberdeen.

Cummins, S. / *Macintyre,* S. / *Davidson,* S. / *Ellaway,* A. (2005): Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources, Health & Place 11 (3), 249 – 60.

Diez-Roux, A. V. (1998): Bringing context back into epidemiology: variables and fallacies in multilevel analysis, American Journal of Public Health 88, 216 – 22.

Diez-Roux, A. V. / *Nieto,* F. J. / *Muntaner,* C. / *Tyroler,* H. A. / *Comstock,* G. W. / *Shahar,* E. / *Cooper,* L. S. / *Watson,* R. L. / *Szklo,* M. (1997): Neighborhood environments and coronary heart disease: a multilevel analysis, American Journal of Epidemiology 146 (1), 48 – 63.

Duncan, C. / *Jones,* K. / *Moon,* G. (1995): Psychiatric morbidity: a multilevel approach to regional variations in the UK, Journal of Epidemiology & Community Health 49 (3), 290 – 5.

Egger, G. / *Swinburn,* B. (1997): An "ecological" approach to the obesity pandemic, BMJ, 315 (7106), 477 – 80.

Erens, B. / *Primatesta,* P. (1998): Health Survey for England: Cardiovascular Disease 1998, London.

Franzini, L. / *Caughy,* M. / *Spears,* W. / *Esquer,* M. E. F. (2005): Neighborhood economic conditions, social processes, and self-rated health in low-income neighborhoods in Texas: a multilevel latent variables model, Social Science & Medicine 61 (6), 1135 – 50.

Galea, S. / *Ahern,* J. / *Nandi,* A. / *Tracy,* M. / *Beard,* J. / *Vlahov,* D. (2007): Urban neighborhood poverty and the incidence of depression in a population-based cohort study, Ann Epidemiol 17 (3), 171 – 9.

Hedley, A. A. / *Ogden,* C. L. / *Johnson,* C. L. / *Carroll,* M. D. / *Curtin,* L. R. / *Flegal,* K. M. (2004): Prevalence of overweight and obesity among US children, adolescents, and adults, 1999 – 2002, Journal of the American Medical Association 291 (23), 2847 – 50.

Hill, J. O. / *Peters,* J. C. (1998): Environmental contributions to the obesity epidemic, Science 280 (5368), 1371 – 74.

Krieger, N. (1994): Epidemiology and the web of causation: has anyone seen the spider? Social Science & Medicine 39, 887 – 903.

Lawley, D. N. / *Maxwell,* A. E. (1971): Factor analysis as a statistical method – 2nd edition, London.

Macintyre, S. / *Ellaway,* A. / *Cummins,* S. (2002): Place effects on health: how can we conceptualise, operationalise and measure them? Social Science & Medicine 55 (1), 125 – 39.

Muthen, L. K. / *Muthen,* B. O. (2005): MPlus user's guide – 3rd edition. Los Angeles.

Pickett, K. E. / *Pearl,* M. (2001): Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review, Journal of Epidemiology & Community Health 55 (2), 111 – 22.

Poston, W. S. 2nd / *Foreyt,* J. P. (1999): Obesity is an environmental issue, Atherosclerosis, 146 (2), 201 – 9.

Riva, M. / *Gauvin,* L. / *Barnett,* T. (2007): Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998, Journal of Epidemiology and Community Health, 61 (10), 853 – 61.

Shaw, A. / *McMunn,* A. / *Field,* J. (2000): The Scottish Health Survey 1998. Volume 1. A survey carried out on behalf of The Scottish Executive Department of Health, UK.

Stafford, M. / *Bartley,* M. / *Wilkinson,* R. / *Boreham,* R. / *Thomas,* R. / *Sacker,* A. / *Marmot,* M. (2003): Measuring the social environment: social cohesion and material deprivation in English and Scottish neighbourhoods, Environment & Planning A 35 (8), 1459 – 75.

Stafford, M. / *Cummins,* S. / *Ellaway,* A. / *Sacker,* A. / *Wiggins,* D. / *Macintyre,* S. (2007): Pathways to obesity: identifying local, modifiable determinants of physical activity and diet, Social Science & Medicine 65 (9), 1882 – 97.

Swinburn, B. / *Egger,* G. / *Raza,* F. (1999): Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity, Preventative Medicine 29, 563 – 70.

Townsend, P. / *Phillimore,* P. / *Beattie,* A. (1988): Health and Deprivation: Inequality and the North, London.

Wiggins, R. D. / *Bartley,* M. / *Gleave,* S. / *Joshi,* H. / *Lynch,* K. (1998): Limiting long-term illness: a question of where you live or who you are? A multilevel analysis of the 1971 – 1991 ONS longitudinal study, Risk Decision & Policy 3, 181 – 98.

Winkleby, M. / *Sundquist,* K. / *Cubbin,* C. (2007): Inequities in CHD incidence and case fatality by neighborhood deprivation, American Journal of Preventative Medicine 32 (2), 97 – 106.