Place effects on health: how can we conceptualise, operationalise and measure them?

Sally Macintyre*, Anne Ellaway, Steven Cummins

MRC Social and Public Health Sciences Unit, 4 Lilybank Gardens, Glasgow, Scotland G12 8RZ, UK

Abstract

In this paper we highlight what we consider to be a lack of adequate conceptualisation, operationalisation and measurement of “place effects”. We briefly review recent historical trends in the study of the effects of place on health in industrial countries, and argue that “place effects” often appear to have the status of a residual category, an unspecified black box of somewhat mystical influences on health which remain after investigators have controlled for a range of individual and place characteristics. We note that the distinction between “composition” and “context” may be more apparent than real, and that features of both material infrastructure and collective social functioning may influence health. We suggest using a framework of universal human needs as a basis for thinking about how places may influence health, and recommend the testing of hypotheses about specific chains of causation that might link place of residence with health outcomes.

Keywords: Place; Area effects; Compositional explanations; Contextual explanations; Collective social functioning; Universal human needs

Introduction

In this paper we highlight what we consider to be a lack of adequate conceptualisation, operationalisation and measurement of “place effects”. We briefly review recent historical trends in the study of the effects of place on health in industrial countries, and argue that “place effects” often appear to have the status of a residual category, an unspecified black box of somewhat mystical influences on health which remain after investigators have controlled for a range of individual and place characteristics. We note that the distinction between “composition” and “context” may be more apparent than real, and that features of both material infrastructure and collective social functioning may influence health. We suggest using a framework of universal human needs as a basis for thinking about how places may influence health, and recommend the testing of hypotheses about specific chains of causation that might link place of residence with health outcomes.

Recent history of ecological analysis

During a period roughly dating from the end of the Second World War to the early 1990s, neither epidemiology, medical geography, nor medical sociology tended directly to study the impact of the local social or physical environment on human health (Macintyre & Ellaway, 2000a), despite the abundance of community studies exploring life in particular localities (Gans, 1962; Suttles, 1972; Young & Willmott, 1957). Below we summarise what we believe to be the main reasons for this absence of attention to the influence of the local environment on human health.

Firstly, there has been a wariness about the use of ecological data (Macintyre & Ellaway, 2000a; Pearce, 2000; Schwartz, 1994), following persuasive critiques of the ecological fallacy. The “ecological fallacy” involves inferring individual level relationships from relationships...
observed at the aggregate level. (In an influential sociological paper, Robinson pointed out that ecological and individual correlations between the same variables can differ markedly, and can even have different signs, illustrating this with an individual level correlation between foreign birth and illiteracy of 0.118, the same data producing a correlation of −0.619 when aggregated to the level of states in the USA (Robinson, 1950)).

Secondly, there were methodological developments in statistics, computing and survey methods which vastly improved researchers’ ability to analyse and use data on individuals, whether gathered by routine statistics such as censuses or by specially designed health surveys. The increased capacity to manipulate large datasets, whether collected for the purposes of health research or not, provided the opportunity to analyse individual predictors of health, and their interactions, in complex multivariate analysis. In the UK, this also provided the opportunity to develop a number of small area indices, either of deprivation/affluence (Carstairs & Morris, 1991; Townsend, Phillimore, & Beattie, 1988), or of other more general characteristics such as the economic or industrial base of the locality (Wallace, Charlton, & Denham, 1995). Much research became driven by opportunities provided by data and technique.

For example, the Longitudinal Study in England and Wales (a follow-up of a 1% sample of the 1971 census linked to successive censuses and to vital statistics such as birth, death, and marriage registrations) has proved extremely useful in studying predictors of mortality (Drever & Whitehead, 1997; Fox & Goldblatt, 1982), but has been constrained by the questions asked, for various government purposes, at each census. Since the UK census does not have a question on income, but does have questions about housing tenure and access to private transport, much analysis of the relationship between deprivation and mortality (whether at an individual or aggregate level) has used tenure and car access as major indicators of deprivation (Macintyre, Hiscock, Kearns, & Ellaway, 2000). It was not that there was any a priori theorising about the role of housing tenure or car access in influencing health, rather that researchers used what data were available on an extremely large and representative, and therefore powerful, population sample.

Thirdly, from the 1950s onwards methodological, conceptual and political individualism was dominant in many industrialised countries (McKinlay, 1995; Mechanic, 1993). This individualism emerged partly from analyses of the epidemiological transition, which emphasised the role in chronic disease of individual lifestyle choices (particularly, the “big four” of smoking, drinking, diet, and exercise), rather than the structural and environmental conditions which had been understood to shape patterns of infectious disease or diseases of extreme want (Research Unit in Health and Behavioural Change, 1989). This individualism was reinforced in the 1980s by the political resurgence of neo-liberalism in North America (Reaganism), Britain (Thatcherism) and other developed countries (e.g. Australia). This was famously summed up by Margaret Thatcher in her remark that “there is no such thing as society, there are only individuals”.

Fourthly, there have been trends within geography, which have emphasised, on the one hand, qualitative, post-modernist, social constructionist and cultural approaches towards health (Gesler, 1992). This has included work on disability (Dyck, 1995; Glesson, 1999; Hall, 2000; Imrie, 1996) caring (Milligan, 1999, 2000) and mental health (Davidson, 2000; Parr, 1997; Pinfold, 2000; Wolch & Philo, 2000). On the other hand, there has been a focus on highly technical analyses of spatial patterns of disease incidence using geographical information systems (GIS), or on the role of specific environmental pathogens (Gattrell, Bailey, Diggle, & Rowlingson, 1996; Langford, 1994; Rigby & Gatrell, 2000). Although some of the work on “place effects” on health has been undertaken by geographers (not all of whom have been convinced that there are any “place effects” (Dorling, 1997; Duncan, Jones, & Moon, 1995), much of it has been undertaken by researchers from outside geography and has not been informed by any specifically geographical perspective (Ecob & Macintyre, 2000; Haan, Kaplan, & Camacho, 1987; Hart, Ecob, & Davey Smith, 1997). Although there has been a series of debates over this issue (Hayes, 1999; Jones & Moon, 1993; Kearns, 1993), there is very little work that has forged an explicitly geographical voice to inform and conceptualise hypotheses concerning “place effects” on health (an exception is Curtis & Rees Jones, 1998).

Although there has been considerable research interest in social stratification and its impact on human health, this has tended to focus on individuals rather than on the environments (physical or social) to which individuals are exposed. Major debates within this tradition of research have focused on the relative importance of health selection, artefact, cultural or behavioural, or material explanations for inequalities in mortality or morbidity; the relative importance of psychosocial and material factors; and the stage of the life course at which inequalities are generated or maintained (Macintyre, 1997a). A more North American, and particularly USA, focus has been on racial inequalities in health (Krieger, 2000; Krieger & Fee, 1996).

There has been some resistance to this tendency to methodological and theoretical individualism. This resistance has been expressed in the development of what has been called “the new public health”, which attempts to redirect the attention of public health theorists and practitioners back towards structural and environmental influences on health and health beha-
viours (Martin & McQueen, 1989; Baum, 1998), and in calls to look upstream at the causes of poor health and inequalities in health, rather than downstream at their expression in individual behaviours or ill-health (Krieger, 1994; McKinlay, 1993, 1999); or to combine upstream and downstream explanations (Susser, 1994, 1998; Susser & Susser, 1996a, b). (It should be noted that this version of “the new public health” is actually a reorientation back to what some would regard as “the old public health”, that is, the nineteenth and early twentieth century attempts to clean up the dirty cities; and to move away from the late twentieth century concern with chronic disease and contributory individual lifestyles, which some would regard as “the new public health” (Porter, 1994; Rosen, 1993). Usage of the term “new public health” may differ in different countries.)

Places and people

Since the early 1990s there has been a resurgence of interest in the role of place in shaping people’s health experiences, and a new debate has developed about the relative importance of people or place characteristics. In 1993 two of the present authors collaborated in publishing a paper which pointed out that area level data were commonly used as surrogates for missing individual level data (for example, exposure to specific pathogens, or poverty). Much information supposedly about areas was actually based on the aggregated properties of individuals as measured in censuses or other surveys. (Useful distinctions have since been made between derived or aggregate area variables, which summarise the characteristics of individual residents, such as the percentage unemployed, and integral or global variables, which describe features not reducible to characteristics of the group, such as de-industrialisation (Diez-Roux, 1998; Mitchell, Gleave, Bartley, Wiggins, & Joshi, 2000)). We also noted that it was frequently assumed that any associations observed between places and health stemmed from the population characteristics of residents in the areas (for example, the age or social class composition). Our conclusion from this review was that there was a need to study directly those features of the local social and physical environments which might promote or inhibit health, and that improvements in public health might be achieved by a greater focus on places (Macintyre, Maciver, & Sooman, 1993).

A paper was published in the following year which reached the opposite conclusion, namely that it is important to focus more on people (Sloggett & Joshi, 1994). Sloggett and Joshi used the above-mentioned Longitudinal Study to follow up nearly 300,000 people, aged between 16 and 65 at the 1981 census, for all cause mortality over the following nine years. A deprivation measure for ward of residence was developed, based on the proportion of the labour force who were unemployed, the proportion of households with no access to a car, the proportion of households not owner occupied, and the proportion of employed men and women in social classes IV and V. In a model controlling for age, this deprivation score was linearly related to mortality among both men and women. However, when individual level versions of this deprivation score (that is, whether unemployed, housing tenure, car access, and social class) were introduced, the odds ratio per unit of area deprivation score became non-significant for men and was strongly attenuated for women. The authors concluded that the evidence does not confirm any social miasma whereby the shorter life expectancy of disadvantaged people is further reduced if they live in close proximity to other disadvantaged people... Deprivation appears to be adequately assessed by personal or household circumstances, which are themselves associated with income. Area based measures of deprivation are not efficient substitutes. For maximum effectiveness, health policy needs to target people as well as places (Sloggett & Joshi, 1994, pp. 1473, 1474).

Subsequently, a number of studies have attempted to investigate whether area differences in a variety of health outcomes were due to the composition of the resident population, or to features of place not captured by individual, compositional, properties. Initially, several studies concluded that there were no effects of area of residence over and above key individual predictors of health behaviours or health. For example it was argued that mortality differences between people in the West of Scotland and the civil service in London, England, were not because of differences between these areas, but because of differences between these two areas in the distribution of deprivation among individuals (Davey Smith et al., 1995).

Duncan et al reported that there were few effects of aggregate levels of deprivation in the area of residence, once individual levels of deprivation are taken into account, on health-related behaviours (Duncan, Jones, & Moon, 1993) or psychiatric morbidity (Duncan et al., 1995). Sloggett and Joshi found no area deprivation effect on adverse fertility events (stillbirth, underweight birth, birth to teenage mother, and sole registered birth) (Sloggett & Joshi, 1998). However, even the studies by Sloggett and Joshi observed some area or regional variations not explained by individual deprivation: for example, a North/South difference within England even after controlling for population characteristics, and a significant, although small, remaining excess mortality
among women in poor neighbourhoods after controlling for individual deprivation (Sloggett & Joshi, 1994).

In the late 1990s, however, most studies which tried to partition area effects into compositional or contextual explanations tended to find that there was some residual effect of area having taken into account a number of compositional features. In Britain this was observed for mortality, long-standing illness, health-related behaviour, and cardiovascular risk factors (Macintyre, 1999). For example, a study in the West of Scotland found that area based and individual socioeconomic indicators both independently contributed to mortality risk (blood pressure (BP), cholesterol, height, body mass index, respiratory function, smoking, and coronary heart disease (CHD) symptoms and signs) (Davey Smith et al., 1998).

Similar findings have been reported from North America. An important early study was that of Haan et al., which found that residents in a poverty area experienced higher mortality over a follow-up period than residents in non-poverty areas. This increased risk of death persisted with multivariate adjustment for age, sex, baseline health status, race, income, employment status, education, access to medical care, health insurance coverage, and a whole range of behavioural factors. The authors concluded that these results support the hypothesis that properties of the socio-physical environment may be important contributors to the association between low socioeconomic status and excess mortality, and this contribution is independent of individual behaviours (Haan et al., 1987, p. 989).

Such findings have been confirmed in larger scale studies using more areas, and multilevel models. For example, Waitzman and Smith examined the effect of poverty area residence on mortality risks among adults from the National Health and Nutrition Examination Survey who were studied in 1971 through to 1974, and followed up for death until 1981. Controlling for household income, years of formal education completed, race, marital status, smoking behaviour, alcohol consumption, exercise, baseline health status, cholesterol level, hypertension, and body mass index, among those under 55 years, those living in poverty areas showed a significant excess of mortality (all cause, cardiovascular, cancer, external causes, and other causes) (Waitzman & Smith, 1998). Similarly, Diez-Roux and colleagues found that neighbourhood characteristics in four US communities predicted CHD prevalence and risk factors after controlling for individual socioeconomic characteristics (education, occupation, income, and house value) (Diez-Roux et al., 1997).

Thus, although cautioning that the finding of contextual effects may be due to unmeasured individual effects, most investigators have tended to conclude that where you live matters for health, although probably not as much as who you are (Pickett & Pearl, 2001).

An important observation in some studies of the late 1990s, particularly those using multilevel models, was of interactions between area and individual characteristics. For example, Waitzman and Smith found that although living in a poverty area predicted mortality among those under 55, it did not do so among older people, after controlling for individual characteristics (Waitzman & Smith, 1998). Diez-Roux observed interactions with gender, race and social class. In one community, women living in a relatively better off neighbourhoods had higher serum cholesterol with worsening neighbourhood indicators, while those living in relatively worse off neighbourhoods had lower serum cholesterol with worsening neighbourhood indicators. Among African American men in the same community, CHD decreased as neighbourhood characteristics worsened; and low social class was associated with high serum cholesterol in “richer” neighbourhoods and low serum cholesterol in “poorer” neighbourhoods (Diez-Roux et al., 1997).

UK studies have found variations in the associations between social class and health depending on the type of area, particularly steep health gradients, between more deprived and affluent individuals or small areas, being observed in more affluent districts or regions (Shouls, Congdon, & Curtis, 1996). Associations between area and health-related behaviour have been observed to vary according to social class (for example in a study in the West of Scotland, there was a relationship between “bad” diet and area deprivation but only among more affluent households (Ecob & Macintyre, 2000)). Gender and age differences have also been observed in patterning by area (Ellaway & Macintyre, 2001), as have differences in the likelihood of observing “area effects” according to the health measure of choice. For example, also in the West of Scotland, body mass index was linearly related to deprivation of neighbourhood after individual controls in women but not men, while respiratory function was linearly related to deprivation of area of residence in both men and women (Davey Smith et al., 1998).

Thus, a more differentiated picture has tended to emerge, in which rather than there being one single, universal “area effect on health” there appear to be some area effects on some health outcomes, in some population groups, and in some types of areas.

Composition and context: a useful distinction?

Most of the research described in the previous section aimed to establish the relative importance of composi-
tional and contextual explanations for the geographical patterning of health. Just as much research on socio-economic inequalities in health was based on debates between what were seen as mutually exclusive competing explanations (for example, selection vs. causation, or material vs. psychosocial (Macintyre, 1997a)), compositional and contextual explanations have tended to be seen as mutually exclusive, competing, and culturally and historically universal. That is, investigators have tended to seek to establish both whether there is any explanatory role for context after taking population composition into account, and how much of the observed geographical variation this context might explain.

We would like to highlight three problems with this common approach. Firstly, the distinction between composition and context may not be as conceptually clear or as useful as may appear at first glance. The properties of individuals or households which are used in many multilevel models are themselves shaped by the properties of the locality used in the same models. For example, social class as based on occupation is likely to be constrained and shaped by the local labour market economy, educational achievement may be based on local school standards, and housing tenure on the local housing market (Macintyre & Ellaway, in press).

Secondly and relatedly, the individual controls introduced into multivariate analysis may well be intervening variables on the pathways between place and health, not “confounders” as they are so often treated (Yen & Kaplan, 1999). This applies to household characteristics, health related behaviours, and to physical or mental functioning. Commonly used individual “confounding variables” include baseline health status (which might already have been influenced by features of the environment), smoking, drinking, or exercise (ditto), and body mass index, BP, height, and respiratory function (ditto) (Waitzman & Smith, 1998; Yen & Kaplan, 1999). These potential confounders are usually introduced without any explicit justification; and while in some studies health behaviours and physical or mental functioning are treated as outcome measures (Davey Smith, Hart, Watt, Hole, & Hawthorne, 1998; Diehr et al., 1993; Duncan, Jones, & Moon, 1996; Hart et al., 1997), in others they are treated as control variables, the only outcome used being death (Waitzman & Smith, 1998; Yen & Kaplan, 1999).

Thirdly, and again relatedly, a crucial problem is the lack of any clear theorising about the mechanisms which might link area of residence and health behaviours or health, and which might form the basis for the selection and interpretation of variables. “Composition” and “context” are frequently treated as unproblematic and obvious distinctions, and underlying causal models are often implicit. For example, in rejecting the role of contextual aspects of deprivation, Sloggett and Joshi said that they had refuted the idea of there being a “social miasma” by which living near to deprived people decreases one’s own health; but this notion of “social miasma” was not spelled out in their own paper nor in any of the papers they cite (Sloggett & Joshi, 1994). (However, more recently there has been some interest in the possibility of “pull up” or “pull down” effects on health of adjacency to different types of people or places (Boyle et al., 1999; Graham, MacLeod, Johnston, Diben, & Briscoe, 2000).)

Despite pleas from several quarters to do so (Curtis & Rees Jones, 1998; Davey Smith et al., 1998b; Kaplan, 1996; Macintyre et al., 1993; Diez-Roux, 1998), few investigators have attempted to hypothesise what features of the local social or physical environment might influence health, and then tested these hypotheses. As Mitchell et al. have noted

Lack of theory has often resulted in a choice of variables with which to characterise an area which is guided more by what is available ‘off the shelf’ than by careful theoretical considerations. It also sidelines the important question of what is the appropriate spatial scale for analysis....Without an explicit consideration of how the characteristics of an area are expected to influence health, it is no surprise that contrasting results have been obtained in the search for area effects (Mitchell et al., 2000, p. 68)

“Context” is thus often treated as a residual category, containing those factors influencing human health behaviours or health which remain once every imaginable individual characteristic is taken into account. It is indeed a black box, an unspecified “miasma” which somehow, but we do not know how, influences some aspects of health, health-related behaviour or health risks in some population groups.

This implied model of area effects is evident in attempts to establish whether there are any area differences in health, having controlled for many of the characteristics of individuals (or areas) which could plausibly be seen as lying on the causal pathway between area and health. For example, in a study of limiting long-term illness among men in England and Wales as reported in the 1991 census, Wiggins et al controlled for individual level characteristics (age, education, ethnicity, social mobility, social class of origin, and migration), and for an area classification of County districts (this includes types such as “coalfields”, “coast and countryside”, “manufacturing”, “mixed economies”, “most prosperous”, “ports and industry”, “resort and retirement” (Wallace et al., 1995)). After these controls, there was a marked reduction in amount of unexplained variation between County districts, a result described by the authors as showing that
overall, there now appears little evidence for a geographical area divide. Put another way, the divide exists only in the absence of information about the occupational and migration histories of the men living there and the type of geographical environment (Wiggins et al., 1998, p. 191).

Although the authors may have meant to imply that they had uncovered the reasons for the originally observed geographical variation, this statement could be taken to imply that there are no such geographical variations, because they have been “explained away” by using information about the areas.

In this regard, the status of “place” is similar to that of a “class” in some epidemiological research. Rather than seeing “social class” as a summary term encompassing a number of specific social, psychological, and material exposures which might influence health, the concept of class is sometimes treated as though it was some mystical single entity which directly influences health, or “gets under the skin” (Anderson & Armstead, 1995, p. 214).

The collective dimension

In a previous paper, we suggested three types of explanation for geographical variations in health: compositional, contextual, and collective (Macintyre, 1997b). Compositional explanations draw our attention to the characteristics of individuals concentrated in particular places; contextual explanations draw our attention to opportunity structures in the local physical and social environment; collective explanations draw our attention to socio-cultural and historical features of communities. This last type of explanation emphasises the importance of shared norms, traditions, values, and interests, and thus adds an anthropological perspective to the socioeconomic, psychological, and epidemiological perspectives often used to examine area effects on health.

For example, children in deprived areas may not play in the open air because their families do not have gardens or the resources to take them to play parks (a compositional resource based explanation); because too few public play parks are provided, and there are no good public transport links to those that do exist (a contextual resource based explanation); or because within the prevailing local culture play is not seen as something which is important to children, or it is not considered desirable or safe for children to play with strangers in public places (a collective explanation). Thus even if families were to be provided with their own gardens or with easily accessible public play parks, children might still not play in the fresh air (we are indebted to Dr. John Davies of the University of Edinburgh for this observation). Another example would be that areas may exhibit high rates of poor health because they contain lots of individuals whose personal characteristics pre-dispose them to smoke (Law & Morris, 1998); because they have lots of cigarette retail outlets, advertisements for cigarettes, and low-price cigarettes (Hackbarth, Silvestri, & Cosper, 1995; Pucci, Joseph, & Siegel, 1998); or because local norms and traditions are relatively pro smoking (Duncan et al., 1996).

Since the collective properties of local residents (such as being fairly pro smoking) are part of the context facing any individual living in that place, we no longer think it sensible to view collective explanations as being separate from contextual ones. However, we think drawing attention to features of collective social functioning and practices is timely in the current state of research on health and place, given the emergence in the 1990s of a lively literature on the role of income inequalities, social capital, and social cohesion on health (Lynch, Davey Smith, Kaplan, & House, 2000a; Lynch, Due, Muntaner, & Davey Smith, 2000b; Lynch, Kaplan, Cohen, Tuomilehto, & Salonen, 1996; Lynch et al., 1998; Wilkinson, 1997a, b; Wilkinson, 1996, 2000; Coburn, 2000). Within this research field, contextual material and institutional resources and opportunity structures have been less the focus of research than collective psychosocial characteristics. Indeed yet another new polarity has developed in research on place and health; namely between the social cohesion/social capital theorists (Wilkinson, 1996) and the neo-material neo-Marxist theorists (Lynch et al., 2000b).

We are struck by the extent to which collective explanations for area differences in health have recently tended to be confined to a very narrow range comprising psychosocial constructs such as social cohesion, social capital, and perceived position in social or economic hierarchies. This ignores many other aspects of collective, shared, social functioning such as ethnic, regional or national identity, religious affiliation, political ideologies and practices, legal and fiscal systems, shared histories, kinship systems, domestic division of labour, gender, age and caste appropriate roles, and so on (Dressler, Balieiro, & Dos Santos, 1999; Kawachi, Kennedy, Gupta, & Prothrow-Stith, 1999; Neelaman, Halpern, Leon, & Lewis, 1997). It also tends to ignore a century or so of research on community and neighbourhood ties and social relations within urban sociology (Donnelly & Majka, 1996; Toennies, 1957). Rather than thinking simply of a distinction between, for example, material infrastructure and investment in areas on the one hand, and social capital among the residents on the other, we think it would be more useful to think about the distinction between material infrastructure and collective social functioning and practices.
We do not wish to introduce yet another (false) polarity. We see no a priori reason why poor health in certain places cannot be the result both of levels of collective social functioning and of material infrastructure (Frohlich, 2000; Frohlich, Potvin, Chabot, & Corin, submitted-a; Frohlich, Potvin, Gauvin, & Chabot, submitted-b). Further, under different historical or socioeconomic circumstances, for different health outcomes, and for different population groups, collective social functioning and material infrastructure explanations may differ in importance, or operate additively or interactively.

A key issue in urban regeneration is whether to focus on material features of the place (for example, building new houses or sports facilities) or on psychosocial features of the residents (for example, enhancing empowerment and collective efficacy). However, most of the evidence suggests that an exclusive focus on either is counter-productive; there is no point investing in new sports facilities if the local population cannot afford to use, or to travel to, them or sports participation is not a valued activity, but there is equally no point investing in encouraging collective action among the local population if no tangible advantages (such as jobs or facilities) are being offered (Forrest & Kearns, 1999).

We should also point out that collective social functioning and practices need not be geographically restricted to residential neighbourhoods. Whereas, as Forrest & Kearns note, “original notions of community were based around the shared geographical spaces of neighbourhood” (Forrest & Kearns, 1999), collective identity and opportunity may be as much related to interest groups and other sorts of spaces such as the family and the workplace, which cut across “traditional” geographical space and are only anchored in “place” by common meeting points (work, the home) (Giddens, 1984). Some communities of interest may indeed be extremely widely dispersed geographically. For example, the norms, values and shared economic interests of the descendants of immigrants to the United Kingdom from areas of the Indian sub-continent may be more similar to those of current residents in the area of origin than to those of their neighbours in the current area of residence; and lesbian and gay people from all over the world may feel at home at the Gay Mardi Gras celebrations in Sydney, Australia.

The black box of places

In the remainder of this paper we try to shed light on the interior of the “black box of places” and to suggest a conceptual framework and set of measures which might be used to test specific hypotheses about the role of place, taking into account the ideas described above about specific contextual aspects of the local physical and social environments which might influence health or health-related behaviours in particular population groups.

What constitutes a healthy neighbourhood?

In our own work in the West of Scotland, we have been trying to look directly at features of local areas that might be health promoting or health damaging. We have been using as an organising framework the following five types of features of local areas which might influence health:

1. **Physical features of the environment shared by all residents in a locality.** These include the quality of air and water, latitude, climate, etc. and are likely to be shared by neighbourhoods across a wide area. In Glasgow, for example, all the drinking water for a city of nearly a million comes from the same loch, so the two and a half fold differences in death rates between neighbourhoods cannot be explained by variations in drinking water.

2. **Availability of healthy environments at home, work and play.** Areas vary in their provision of decent housing, secure and non-hazardous employment, safe play areas for children, etc. These environments may not affect everyone living in an area in the same way that air and water quality do; they may affect the employed more than the unemployed, families with children more than elderly people, and so on.

3. **Services provided, publicly or privately to support people in their daily lives.** These include education, transport, street cleaning and lighting, policing, health and welfare services. Again, how these affect people may depend on personal circumstances. Public transport may matter more if you do not have a car.

4. **Socio-cultural features of a neighbourhood.** These include the political, economic, ethnic and religious history of a community: norms and values, the degree of community integration, levels of crime, incivilities and other threats to personal safety, and networks of community support.

5. **The reputation of an area.** How areas are perceived, by their residents, by service or amenity planners and providers, by banks and investors, may influence the infrastructure of the area, the self-esteem and morale of the residents, and who moves in and out of the area (Macintyre et al., 1993 pp. 220–221).
The first three of these categories can best be seen as material or infrastructural resources. We have conceptualised features such as these as “opportunity structures”, that is, socially constructed and socially patterned features of the physical and social environment which may promote or damage health either directly, or indirectly through the possibilities they provide for people to live healthy lives. An example of a direct effect would be if polluted air or water compromises the health of residents; an example of an indirect effect would be the local availability of affordable and nutritious food (since not all individuals are totally dependent on local food supplies, some being able to range further afield and thereby manage to achieve a healthy diet). In our work to date we have tried to measure directly a number of these features: for example we have examined the number of dwellings whose water supplies exceeded the EC guidelines for lead content (Macintyre et al., 1993), and we have also studied the distribution of food retail outlets and pricing of different types of foodstuffs in different areas in Glasgow (Cummins & Macintyre, 1999; Macintyre et al., 1993; Sooman, Macintyre, & Anderson, 1993).

The last two categories in this typology relate to collective social functioning and practices, and we have also studied features of these (for example, people’s use of their local area, perceived social cohesion, aggregate and integral measures of social participation such as the number of local voluntary groups, and symbolic and actual representations of the areas via newspaper and other accounts of the area, including photographs (Macintyre, 1997b; Macintyre & Ellaway, 1998; Macintyre & Ellaway, 1999; Macintyre & Ellaway, 2000b)).

Although this framework has proved useful, it is limited in that it does not specify exactly what we would need to study within each of these broad domains in order to take further our understanding of the importance of specific aspects of the social and physical environment. In developing our work we now feel it is important to step back and learn from the literature on urban planning.

Much urban planning this century has focused as much on social aspects of the neighbourhood environment as on physical ones; meeting places, as well as drains. Forest and Kearns have noted that:

In its early twentieth century conception, the neighbourhood was cellular, bounded, inwardly focused and relatively self-contained. It emphasised environment as a major determinant of residential quality of life. The concept of neighbourhood, which was adopted in much American urban planning and in new town planning in Britain and elsewhere, derives from the fact that there were concerns in the first decades of the century about the effects of urbanisation, and the neighbourhood concept was an attempt to counter these by promoting neighbourliness, identity with primary groups in places, order and dignity. The neighbourhood was intended to create the physical place that was coterminous with a sense of community through providing a setting in which there were opportunities for leisure, recreation and social interaction, and an environment that was safe, secure and protected (Forrest & Kearns, 1999, p. 24).

In an excellent text on typologies of (mainly urban) residential environments, Brower proposes that good housing should satisfy the following functions:

Shelter. The housing should be available, affordable, structurally sound, and compatible with physical health and mental well-being. It must provide effective protection against inclement weather and unwanted intrusions.

Housekeeping. The cost and effort involved in housekeeping tasks must be reasonable and within the limits of the householder’s resources. These tasks include preparing and storing food, providing regular care and maintenance, replacing worn or damaged parts, and disposing of waste.

Accommodation. The various spaces and facilities—their size, division, arrangement, and equipment—must be suitable for full range of domestic activities.

Connection. There must be appropriate connections to other households and places of employment and service. Site and occupancy conditions should create opportunities for co-operation and not promote friction among neighbours.

Meaning. The character, appearance, and conditions of use of the housing must reflect the values of the householder and foster a sense of belonging, attachment, and advocacy.

Recreation. Housing must provide householders with opportunity to relax, rest, take time out, recreate themselves. (Brower, 1996, p. 18).

The requirement to satisfy those functions could be extended beyond housing to residential neighbourhoods (Greenberg, 1999). However, when thinking about human health, we think it is important not to forget about aspects of the basic physical design and layout of residential neighbourhoods, such as clean air, water, street cleaning and lighting, and sewerage systems, which so exercised the 19th-century sanitary reformers (Porter, 1994; Rosen, 1993).

We suggest that a starting place for conceptualising and measuring area influences on health is to consider what humans need in order to live a healthy life, given
the particular socio-economic and socio-cultural context. The following list starts roughly in the order of the hierarchy of human needs (Maslow, 1968).

- **Air**
  - unpolluted air

- **Water**
  - clean water for drinking and cooking

- **Food**
  - adequate supplies of nutritious and non-poisonous food

- **Shelter**
  - protection from wind, cold, rain

- **Security**
  - protection from threats to the person or property

- **Hygiene**
  - protection from infectious or contagious disease and from toxins and pollutants

- **Education**
  - socialisation in the skills and information needed in a given society

- **Healing**
  - care and treatment for the sick and infirm

- **Housekeeping**
  - resources for food storage and preparation, cleaning (of people, clothes and homes), waste disposal

- **Work**
  - gainful labour

- **Means of exchange**
  - money, credit, or other forms of trading power

- **Information**
  - access to prevailing media of information and communication (books, newspapers, postal and telecommunications services, etc.)

- **Transport**
  - private and public transport, roads, railways etc

- **Personal relationships**
  - family life, intimate relationships, acquaintance and friendship networks

- **Religious**
  - spiritual or ritual practices

- **Involvement in group activities**
  - participation in political, social, or economic, activities

- **Play**
  - social, cultural, and physical recreation

How these needs are met depends on the state of technological development and wealth, and how these are distributed, in any given society. For example, books may now be less important than television or the World Wide Web as sources of information; conversely, bicycles may be an extremely important means of transport in China and parts of Africa, although they are mainly recreational playthings in parts of North America and Europe. It seems to us legitimate to ask, within a given society, how the resources available to meet these needs are distributed geographically, and whether this distribution is in turn related to the distribution of health.

This means operationalising measures, appropriate for the particular society and historical period, of the ways in which these human needs are met in particular places. For example, what is the quality of the air and drinking water and what regulations are in force to control air or water quality? How is food produced, regulated, priced and distributed? What is the standard of housing, local housing regulations, the proportion of sub-standard or damp houses? What is the location and quality of schools? What is the location, staffing, call out times, and policies of the police and fire services? What is the distribution, frequency, and quality of bus, tram, and railway connections? What is the location and opening hours of launderettes, dry cleaners, clothes and shoe stores, ATMs, fast food outlets, theatres, cinemas, pharmacies, etc.?

We are not suggesting that every study of area and health try to cover every single measure of the physical and social context which might influence human health. Rather, we wish to suggest that investigators should attempt to hypothesise and test specific pathways by which area might influence health. If the concern is with cardiovascular health, one might wish to identify major risk factors for cardiovascular disease such as obesity, diet, lack of exercise, smoking, and stress, and link these features of the material and social environments which promote such risk factors (for example, the price and availability of more and less healthy types of food, opportunities for exercise, tobacco advertising and cigarette outlets, local norms and values about certain types of behaviours, and environmental stressors such as crime, violence, noise, overcrowding, etc.). Some of this context may be international (eg the activities of multinational tobacco companies), while others may be national (eg the tax on tobacco products), or local (eg the location of tobacco retail outlets and the control of ostensibly non-smoking spaces (Frohlich, 2000; Frohlich et al., submitted-a,b)).

Our earlier conception of opportunity structures is still important in considering this list of basic human needs and how they may be met. For example, public transport facilities such as buses and train routes may be available to those living in a particular residential area, but may not be used for a variety of reasons: because it is more convenient to use private transport; because public transport may be dirty, dangerous, and stigmatising; or because it is not part of the sub-cultural repertoire (Ellaway, Hiscock, Macintyre, & Kearns, submitted; Hiscock, Macintyre, Ellaway, & Kearns, in press).

There may be many places of worship available locally, but this does not necessarily imply that local residents are religious; the distribution of places of
worship may reflect historical trends, and they may now hardly be used. A better measure of religiosity in an area might be the frequency of, and attendance at, religious services; funds donated to religious institutions; and/or participation in activities related to religion. Thus we can conceive of the number of places of worship locally as being an environmental input, a resource that people may choose whether or not to use, and measures of religious participation as reflecting an output of the activities of local residents.

Similarly, the local availability of police stations and police staff, and current policing policies, are measures of policing input, while crime, delinquency and vandalism rates are measures of residential output; the provision of fire stations and the response times of fire vehicles to emergency call outs is an input measure, the number of fires is an output measure; public and private investment in the school system is an input measure, educational achievement rates locally an output measure; and the location and opening hours of public libraries are an input measure, local book borrowing rates are an output measure.

This distinction between “inputs” and “outputs” is not absolute, and of course may be more complex than is implied above, but we think it is important not to confuse measures of material infrastructure and investment in residential areas with measures of the collective social functioning of local residents. We also think it important for investigators to be explicit about whether they are thinking of various area-based measures as being input or output measures.

It is also important to recognise the dynamic and interactive nature of relationships between features of place and features of the local population, and to add a time dimension to our theorising. A high ratio of places of worship in relation to number of local residents may reflect levels of religiosity, and of wealth sufficient for people to donate funds to build churches, in the late nineteenth century; thus the supply of churches may be seen as a response to a demand at that time for churches. Local retailers are sometimes criticised in the UK for not providing high-quality and low-cost fruit and vegetables in poor areas, as if they are wilfully culpable of failing to provide poor people with the resources to consume an adequate diet. However, retailers to whom we have spoken in socio-residentially deprived areas point out that they would like to be able to sell fruit and vegetables, but there is no local demand for them, and since these goods are perishable, shopkeepers cannot afford to stock them (Horwitz, 1992).

The ways in which access to money is made available locally may vary according to the financial situation and credit-worthiness of the local residents. For example, building societies, bank branches, and ATMs may be present in areas containing people with monthly salaries, bank accounts, and credit or debit cards, while cheque cashing agencies, money lenders, credit unions and pawnbrokers may be present in areas in which people are paid weekly in cash, do not have bank accounts, and are unable to obtain credit cards. Thus “access to the means of exchange” may be supplied differently, and has different social meanings, in different types of areas.

Differing modes of meeting human needs may have varying relationships with health or mortality because they reflect a number of characteristics of the area and local population. Thus it is important to recognise the interplay of supply and demand for services to meet a range of human needs. It is also important to recognise that “more of everything” is not necessarily better for health. Yen and Kaplan, for example, found that there was a positive relationship between mortality and the number of commercial stores (pharmacies, beauty salons, laundries, supermarkets) in the locality: that is the probability of dying was higher in areas in which many such resources were located (Yen & Kaplan, 1999). (The presence of pharmacies or laundrettes may reflect the fact that there is poor local provision of primary medical care, or that the residents are too poor to have washing machines in their homes.)

Methodological issues

In trying to collect data on local material infrastructure and the social context in areas, three major methodological issues are what spatial scales are appropriate for meeting different needs, at what spatial scale or level of aggregation information is actually available, and what might be the appropriate time interval between environmental exposures and any effects on health. Although much epidemiology and medical geography uses administrative or postal delivery areas such as census tracts, electoral wards/districts, counties, or postcode sectors/zipcodes, it is often acknowledged that these may not be appropriate scales to use for different types of human activities. Just because something is an appropriate size of territory for one census enumerator or mail delivery person does not necessarily mean it is the appropriate scale for analysing the provision of grocery stores or theatres. It could be argued that every household should have access within walking distance to a grocery store or a telephone, but few people would argue that every household should be within walking distance to an opera house, or to a fire, ambulance or police station. While every community might need a primary (elementary) school or access to primary health care, we would not expect every community to have a University or specialist cancer centre. Thus, we need to think carefully about the appropriate spatial scale and range of resources to meet different human needs.
Even if we were to specify, *a priori*, the appropriate spatial range for different activities (food buying, schools, access to money), we might not find that the data are available at that spatial resolution. Some data, for example those available in the Yellow Pages, might be postcoded (zip coded) and so be available at a small area level and capable of being aggregated to many spatial scales; but data on investment in secondary education, or in municipal functions such as street lighting, street cleaning, garbage disposal, and water or sewage treatment might only be available at a much larger scale.

Much of the literature on neighbourhoods and health relates to urban or city neighbourhoods. We also need to think of ways of modifying our measures and their spatial scale to take account of rural or sparsely populated areas. Then the issue might be, not whether there is a grocery store or primary school within walking distance of every household, but whether there is adequate provision of public transport giving access within a reasonable time to a grocery store or primary school, or whether there are peripatetic services visiting communities to supply food, post, or library or dentistry services.

Most studies of area effects on health are cross sectional; that is, measures of the place and of the residents’ health are collected at roughly the same time. When one starts to think about socially and biologically plausible causal pathways by which place might influence health these cross-sectional designs often appear inappropriate. While someone’s diet or smoking behaviour at any given time might be influenced by features of the local neighbourhood at the same time, their weight and respiratory function are likely to be the result of cumulative exposures to food, tobacco products and air pollution over many years or decades. The effects of the environment on adult mortality is likely to have a particularly long time lag, so if we seek area-based explanations for mortality rates in the 1990s we should perhaps be looking at characteristics of the place of residence in the 1940s, or even 1920s. However, with one or two notable exceptions (Barker & Osmond, 1987; Barker, 1991), this is rarely done. Instead, a number of different health outcomes are often analysed in relation to contemporary measures of place, as if there is likely to be an instantaneous impact of levels of social trust, income inequalities or local material deprivation on aspects of human health as disparate as death, height, blood pressure or smoking (Kawachi, Kennedy, Lochner, & Prothrow-Stith, 1997; Kennedy, Kawachi, & Prothrow-Stith, 1996; Davey Smith et al., 1998a).

As Blakely and Woodward have pointed out, a zero time lag between exposure and outcome is usually implausible in social epidemiology, and one therefore needs to think about the stability over time of the ecological exposure (Blakely & Woodward, 2000). The challenge is to include in testable hypotheses explicit assumptions both about plausible time intervals during which area influences on health are likely to manifest themselves in health outcomes, and about the consistency over time of area characteristics.

**Conclusion**

The recently renewed interest in place or community effects on health has been largely data driven and has rarely been underpinned by explicit, testable, hypotheses about the mechanisms by which place might influence specific aspects of health in specific population groups, and the appropriate time scale over which these operate. Conflicting evidence about the extent and magnitude of area effects on health may be due to differing conceptualisations and operationalisations of area effects, and in particular to differences in whether certain features of individuals or local areas (such as baseline health status, smoking, or cardiovascular risk factors) are seen as “confounding” or “intervening” variables. Some of the conflicts in the evidence may be more apparent than real, in the sense that there is no *a priori* reason to assume that similar area effects will be observable for all spatial scales, population subgroups, health outcomes, risk factors, time scales or socio-economic and cultural contexts.

We have suggested that it might be helpful, firstly, to distinguish between compositional and contextual explanations for spatial variations in health; secondly, to include collective social functioning and social practices as candidate contextual mechanisms; thirdly to expand our conceptualisation of collective social functioning beyond the confines of social capital/social cohesion, to include other features of non-material culture; fourthly, to derive measures of context from an analysis of basic human needs; and finally, to develop robust, testable, hypotheses about the potential impact of features of the local social and physical environment on human health, and to test these empirically.

In a research project recently funded by the UK Medical Research Council we will be collaborating with colleagues from University College London in attempting to measure contextual features (in particular, material infrastructural inputs from public and private services) and features of collective social functioning (including not only social capital and social cohesion but also other measures of social practices) in a large number of neighbourhoods in England and Scotland. We will then examine whether there are any associations between these data on the social and physical environments and data on the health and health-related behaviour of residents in the same areas. This is a major extension to our previous work on urban neighbourhoods in the West of Scotland, and we hope to
incorporate into this study many of the ideas sketched out above.

Acknowledgements

All authors are employed by the UK Medical Research Council. The new study to which we will be applying the framework developed in the latter half of this paper is funded by the MRC (grant number G9900620). We are grateful to Fiona McDonald for help with the references, and to Anne Kavanagh and an anonymous referee for helpful comments on earlier drafts.

References


Frohlich, K., Potvin, L., Gauvin, L., & Chabot, P. Disentangling contextual from compositional effects? Insights into
the I>We problem. *Journal of Epidemiology and Community Health* (submitted-b).


Macintyre, S. (1997b). What are spatial effects and how can we measure them? In A. Dale (Ed.), *Exploiting national survey data: the role of locality and spatial effects* (pp. 1–17). Manchester: Faculty of Economic and Social Studies, University of Manchester.


Macintyre, S., & Ellaway, A. Neighbourhoods and Health: Overview. In: I. Kawachi, & L. Berkman (Eds.), *Neighbourhoods and Health: Oxford University Press, in press.


Pinfold, V. (2000). Building up safe havens...all around the world; users’ experience of living in the community with mental health problems. *Health and Place, 6*, 201–212.


