PERGAMON

Health & Place 8 (2002) 15–24

HEALTH & PLACE

www.elsevier.com/locate/healthplace

Population production and modelling mortality—an application of geographic information systems in health inequalities research

Richard Mitchell[a,*], Danny Dorling[b], Mary Shaw[c]

[a] Research Unit in Health, Behaviour and Change, University of Edinburgh Medical School, Teviot Place, Edinburgh EH8 9AG, UK
[b] School of Geography, University of Leeds, Leeds LS2 9JT, UK
[c] School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK

Abstract

Much research into spatial inequalities in health has sought the balance between compositional and contextual influences on observed patterns. Research published recently by the authors sought to determine whether composition of areas alone might account for the changing geography of mortality in Britain, between 1983 and 1993. The research required data describing Britain in terms of the numbers and location of people with every possible combination of age group, sex, social class and employment status. This paper describes the approach and the custom written geographic information system which estimated these data, and their subsequent application. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Mortality; Inequality; GIS; Data estimation

Background

Much research has demonstrated growing spatial inequalities in health in Britain, at many different spatial scales (Shaw et al., 1999). Britain's health 'gap' (between sick and well) is wider now than at any time since records began (Dorling, 1997; Shaw et al., 2000). Broadly speaking, theories as to the cause of these inequalities are divided into two contrasting themes (Macintyre et al., 1993, forthcoming; Shaw et al., 2001). The first explanation, commonly referred to as 'compositional', suggests that area level mortality or morbidity rates reflect the risks of ill health which the resident individuals carry with them. The relationships between individual level factors such as social class and employment status, and the risk of mortality or morbidity, are well documented, powerful, and very robust. The composition thesis thus argues that places with apparently high levels of sickness or death rates are those in which a higher proportion of the residents are at higher risk of sickness or death. The second explanation, commonly referred to as 'contextual', suggests that the nature of day-to-day life in an area can exert an influence on the mortality risk of the resident population, over and above their individual characteristics. The influences might, for example, stem from the social or physical environment. Somehow, life in an area raises or lowers the risk of ill health for the resident individuals so that they experience different risk of illness from that which they might experience living somewhere else.

In September 2000 the authors published a report "Inequalities In Life and Death—What If Britain Were More Equal" (Mitchell et al., 2000a) to widespread media, public and government interest (see for example The Times, 26.09.00 Poverty 'is killing 10,000 every year' and The Guardian, 26.09.00 Redistribution of wealth 'will save lives'). The report looked back in time to trace and explain the development of geographical

---

*Corresponding author. Tel.: +44-131-651-1283; fax: +44-131-650-6902.

E-mail address: richard.mitchell@ed.ac.uk (R. Mitchell).

PII: S1353-8292(01)00032-6

inequalities in health in Britain between the early 1980s and early 1990s. Using a model which related population characteristics to risk of mortality, the report demonstrated how an understanding of change in the nature of residential populations across Britain accounts for virtually all the changes in Britain's geography of mortality over this time period. It also looked forward to see the potential effect on patterns of mortality of possible future scenarios in which a government delivers the eradication of child poverty, full employment and a mild redistribution of income.

The report thus suggests that, at this geographical scale, compositional influences on the changing geography of mortality are certainly dominant and possibly exclusive for many areas. Indeed, its projections of future policy scenarios and their impacts relied on compositional dominance. However, like all research of this kind, the findings are highly dependant on the quality of the population data with which the models were fuelled. Behind the report lay a model of the British population created by a custom written data manipulation and geographic information system (GIS). The GIS combined existing aggregate and disaggregate data sources with a neighbourhood analysis function to produce four items of information (age, sex, social class (I–V) and employment status) for all residents of every parliamentary constituency in Britain. In effect, two models were produced—one for Britain as it was in 1983 and one for Britain as it was in 1993. Creating these models required (what we believe is) an unprecedented approach to population modelling.

In this paper, we explain how and why the population model was created and applied. We begin by establishing why such a comprehensive and complex data set was essential for the research question at hand, and then explain our approach.

The nature of the task

Most research into spatial health inequalities recognises that differences in area composition are responsible for the great majority of area variation in health, at least in statistical terms (Macintyre et al., forthcoming). If the (vast) majority of geographical differences in mortality rates at any one point in time are accounted for by differences in the individual characteristics of resident populations, the development of those patterns through time ought also to be accounted for by changes in (i) the nature of the residential population's characteristics and (ii) the relationship between individual's characteristics and their risk of mortality. Our research was designed to test this hypothesis for changes in Britain's geography of health during the period between the early 1980s and the early 1990s, with mortality as the health outcome of interest. We wished to try and account for every premature death, in every part of Britain. Where deaths rose or fell in an area we wanted to determine whether changes in population and/or population risk could account for it.

This was not a task suited to samples or survey data. Since everyone in Britain is at risk of death, we required information on everyone. Previous work had shown that parliamentary constituency is an appropriate spatial scale for this type of research due to the size and stability of the areas. Previous research (Shaw et al., 1998) had also shown that the most efficient and effective predictor of an individual's risk of mortality is their combination of age (5 yr age group is sufficient), sex, social class and employment status. Government statistics provide quantification of the mortality rates associated with all possible combinations of age, sex, social class and employment status, and how these changed between the early 1980s and early 1990s. 'All' we needed was a data source to describe Britain in terms of the number and locations of people with every possible combination of these characteristics for both time points! No data source existed which could provide all the detail needed. Our technical challenge was thus to manipulate and combine those data which *were* available to estimate those which we required.

In more formal terms our aim was to generate a plausible data set which described the number of individuals with every possible combination of age group, sex, social class and employment status, in every parliamentary constituency in Britain in 1983 and then again in 1993[1]. Although the relationship between characteristics and mortality takes place at an individual level, we did not need to create a genuinely individual level data set (in which a record would be held for every person in Britain). Since their combination of age, sex, social class and employment status would be enough to ascribe a mortality rate to a person, we only needed to know the counts of people who possessed each possible combination of those characteristics, within each constituency area.

Methods

The task was split into three main components: (a) the creation of disaggregate age/sex/class records from aggregate source data, (b) the allocation of a social class (I–V) and employment status to all individuals for whom that information was missing following stage (a), and (c) the application of appropriate mortality rates to

---

[1] When we measure mortality rates, a 5 yr average is used to reduce the influence of epidemics such as flu or meningitis. Our death data covered the periods 1981–1985 and 1991–1995. Population data for the mid-points (1983 and 1993) were thus required.

individuals and subsequent geographical analysis of the results. Code was written to complete most of components (a) and (b). Where appropriate, further manipulation was carried out in Microsoft Excel with text-based files used to port data from one platform to another. Component (c) was carried out in ArcView, with additional analysis in Excel. This was very much an approach based on the principles of GIS rather than on an off the shelf software solution. The geographical aspects of the work were simple in nature. Complexity stemmed from the size of the data set.

Creating disaggregate data

Problem

The UK census provides population counts by age and sex at ward level for Britain at the two points in time. These counts are provided in the small area statistics (SAS). For stage (a) we generated much of our data at ward level using the SAS, and this enabled us to use features of the neighbourhood to enhance our estimation procedures. The ward level figures were subsequently aggregated to parliamentary constituency for analysis. The census also (separately) provides limited information on the number of individuals in each social class and on the number of individuals in each possible category of employment status. However, since these data are all aggregate, there is no direct means of determining which classes or employment groups the individuals of a given age and sex belong to.

Fig. 1 illustrates this problem. The data which need to be estimated are represented as ?'s.

It seemed prudent to determine what other work had been carried out in this field which might help us. The usual route to disaggregate data which describe individual characteristics or behaviour is microsimulation (Mitchell, 1997; Mitchell et al., 1998). Here, synthetic populations can be created and then followed as they live out their artificial lives. There are numerous examples of successful microsimulation models which run close to the scale (in terms of population size) of our task, for example the CORSIM model under development at Cornell University since 1986 (Caldwell and Keister, 1996; Caldwell et al., 1998). CORSIM has been used to model wealth distribution in the United States over the historical period 1960–1995 and to forecast wealth distribution in the future. A further example of microsimulation modelling is the work of the Spatial Modelling Centre[2] in Sweden (Holm et al., 1996; Vencatasawmy et al., 1999). The SMC built on previous microsimulation modelling efforts (Holm et al., 1996) and constructed total population simulation models (TOPSIM) and systems for visualising economic and regional influences governing the environment (SVERIGE), which simulate the entire population of Sweden. In addition, SVERIGE is aimed at studying the spatial consequences of various national, regional and local-level public policies. The database used for this model comprises longitudinal socio-economic information on every resident of Sweden for the years 1985–1995 and in that sense it is based on the kinds of data which were not available to us. Nonetheless, it provided some encouragement that a national scale population model might be feasible.

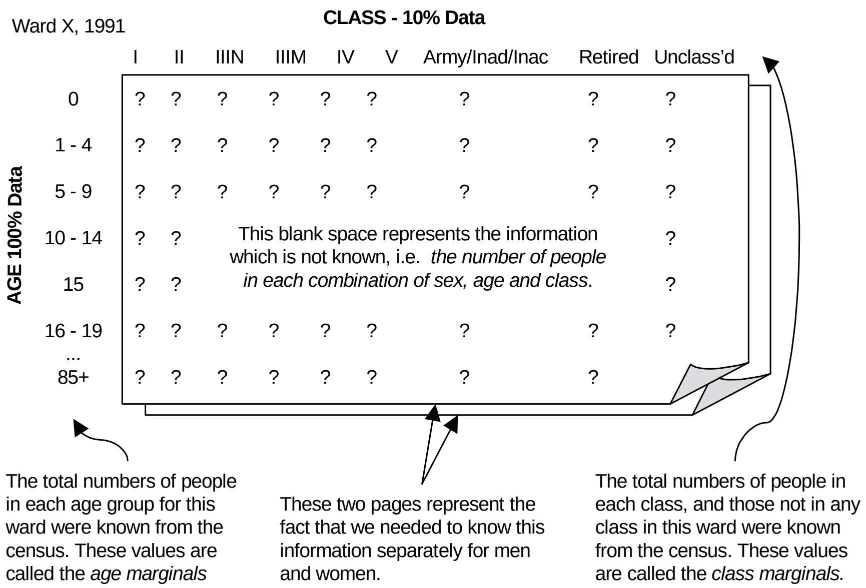
There is a lot of microsimulation work at the UK national level. For example, Falkingham and Lessof (1992) present LIFEMOD which is an example of dynamic cohort microsimulation, simulating the life histories of a cohort of 2000 males and females. Each individual 'experiences' major life events such as schooling, marriage, childbirth, children leaving home, employment etc. Another example of a UK national dynamic microsimulation model is PENSIM (Hancock et al., 1992) which models the influences of policy change on pensioner income distribution up to the year 2030. Recent work on the eradication of child poverty in Britain has been using a simulation called POLIMOD to determine the effects of New Labour Government benefit policies on the number of children lifted out of poverty (Piachaud and Sutherland, 2000; Sutherland, 2000). Ballas and Clarke (2001) and Clarke (1996) provide useful looks at this subject in more depth and with greater knowledge.

Although it was encouraging, none of the work reviewed provided a precise blueprint to follow because our research needed very much more simple outputs. The work we identified as being closest to our aims (yet still very much more complex) was that of Birkin & Clarke (1988) and Williamson et al. (1998). Birkin and Clark's work concerned the production of a microdata sample for the population of Leeds Metropolitan District. The synthetic population was to consist of individuals placed together in households and conforming as closely as possible to known aggregate parameters. The data to be simulated were effectively an individual, disaggregate set. Birkin and Clarke proceeded by generating households using sample distributions and table marginals from the SAS, combined through iterative proportional fitting (Johnston and Pattie, 1993).

The addition of other attributes to the household heads was carried out stage by stage. In each case, data were required that would yield a frequency distribution of the attribute to be added, based upon some or all of those attributes already assigned to the household. Following the production of a probability distribution for the feature in question, Monte Carlo methods were employed to assign the developing households the attribute. Iterative proportional fitting (IPF) constrained the generated 'synthetic' population to fit the known parameters of the real population.

---

[2] http://www.smc.kiruna.se



The total numbers of people in each age group for this ward were known from the census. These values are called the *age marginals*

These two pages represent the fact that we needed to know this information separately for men and women.

The total numbers of people in each class, and those not in any class in this ward were known from the census. These values are called the *class marginals.*

Fig. 1. Diagrammatic representation of the problem posed by only having access to aggregate table margins.

It was clear that IPF applied in a similar manner to Birkin and Clarke, would provide the best means of converting our aggregate SAS data to a disaggregate model in which combinations of individual level characteristics were known. Our version would be very much more simple since we were not attempting to physically create synthetic individuals and our generation of counts would have no probabilistic component. In fact, it is debatable whether the approach we adopted can be called a microsimulation at all. We will expand further on this notion following discussion of our methods.

Solution

Since this is a paper about the geographic aspects of our techniques, we will not dwell too much on how IPF works in practice but will draw out how geography and GIS were involved in getting the most plausible results. A complete step-by-step explanation of how the IPF technique was implemented is available (Mitchell et al., 2000b).

The principle of IPF is to find a set of table cells which, when aggregated, match a known set of table margins. In this case we knew the numbers of men and women of each age (forming the row marginals in Fig. 1), and we knew the numbers of men and women in each social class (forming the column marginals in Fig. 1). To get plausible results from IPF, seed values were needed. Seed values describe typical proportional distributions for the table. In this case, we needed seed values which described how the people in each age group are usually distributed amongst the classes. Seed values are necessary since it would be wrong to assume that there is no relationship between age group and social class.

The first geographical augmentation to the standard IPF procedure came in making sure we had appropriate seed values. The IPF was seeded using an age-sex-class distribution drawn from the disaggregate ONS longitudinal study (LS) (Hattersley and Creeser, 1995). The LS is a sample of individuals drawn from the census, for whom the combination of age, sex and class are actually known. However, we could not simply apply the age/sex/class sampling distribution from the LS to every ward in Britain since we also knew that the relationships between these variables differ in different parts of the country. It is widely suggested that Britain's regions contain quite distinct types of population, some richer, some poorer, some older, some younger, some more working class, some more professional. To get the best possible seed values, a simple geographical lookup function within our custom written code matched each ward in Britain to its (larger) Registrar General's Standard Region. The LS provided a separate age/sex/class sampling distribution for each Standard Region (with the English North being used as a surrogate source for Scotland which is not covered by the LS). The seeds for each ward were thus congruent with the wider social and economic circumstances of the region.

Custom written code ran the IPF procedure on one ward at a time, identifying the most appropriate sampling distribution from the LS, calculating seed values for the ward and iterating cell values so that when summed they were constrained to match each known table marginal to within 0.5%. This produced plausible tables describing the numbers of men and women, of

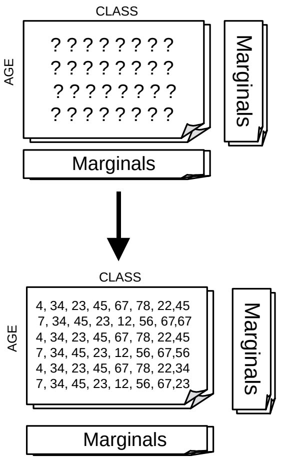


Fig. 2. Diagrammatic representation of the situation bridged by the first round of iterative proportional fitting.

each age group, in each social class, for every ward in Britain, in 1981 and 1991 (manipulating the data so that they refer to 1983 and 1993 came later in the process).

In Fig. 2 we recap the two situations which this process has bridged. At the top of the figure we see a representation of tables in which the marginals were known but the cell contents were not, and at the bottom we see a representation of tables for which we had estimated those cell values. By being aware of the likely differences in the age/sex/class relationship in different parts of the country and using geographical relationships to reflect those differences, we created the most plausible ward level information possible from the available data.

Allocating everyone to a social class I–V and an employment status

Problem

Not everyone is assigned to a social class in the census. Amongst the several reasons for this are the designation of class via occupation and distinguishing those in the army, those in retirement and those with inadequate descriptions of their occupation, from those easily allocated to a class I–V. However, while the circumstances of some people might disqualify them from certain categories on the census form, we know that no one lives in a vacuum outside social class structure. The influence of social position on health is felt by all. In addition, the mortality rates we wished to apply in the analysis phase of the work distinguished people by social class I–V and were based on *everyone* having a class within that range. It was essential that we allocated these 'classless' people to a class I–V in an appropriate way if we were to adequately model the effects of social class on Britain's geography of health. If we had been able to access individual level records for the whole population of Britain we might have allocated these classless individuals to a class based upon their spouses, or other household level relationships. However, without access to such information we had only one possible source of data.

Solution

We used the idea that people who live in the same neighbourhood tend to have similar social and economic characteristics for assigning people to an appropriate social class. Although this is a classic application of GIS our version included some original enhancements, most notably the application of a temporal axis to the estimation procedure.

In principle, those who did not have a class I–V from the census were assigned one based upon other people of the same age and sex who live in the same neighbourhood (ward). For example; if according to the census 25% of men aged 20–24 were in class IIIM in a ward, 25% of men in the same ward, also aged 20–24 but who were not given a class by the census, were placed in class IIIM. Note though that this is *not* the same as allocating a social class based on residence in an area. This was an allocation of class based on the characteristics of other similar individuals with whom neighbourhood residence was shared.

Fig. 3 illustrates this approach. The people holding little certificates are those who had a class I–V by virtue of their relationship with the labour market. Those without the certificates had no class I–V and were allocated one on an age/sex proportional basis. The GIS code identified both a spatial and a social relationship between the 'donor' and 'recipient' groups to maximise their similarity. The characteristics of neighbours and peers who had a class were used to designate the class for others.

Problems with this approach

This version of the approach proved very problematic. We were able to test our version of the age/sex/class (I–V) structure for England and Wales through comparing aggregated results with work by Hattersley (1999) from the LS. Using additional information about household relationships to which we did not have access, Hattersley had created her own population in which almost everyone has a class I–V. When we compared our version of the national age/sex/class (I–V) distribution to hers we found systematic differences between the two versions, pointing towards groups for whom our initial solution was inappropriate.

We found that our system failed when applied to areas in which there were lots of retired people, and more generally for middle and older aged women. The reasons were, with hindsight, quite obvious. We were trying to assign a class I–V to people without one based on those

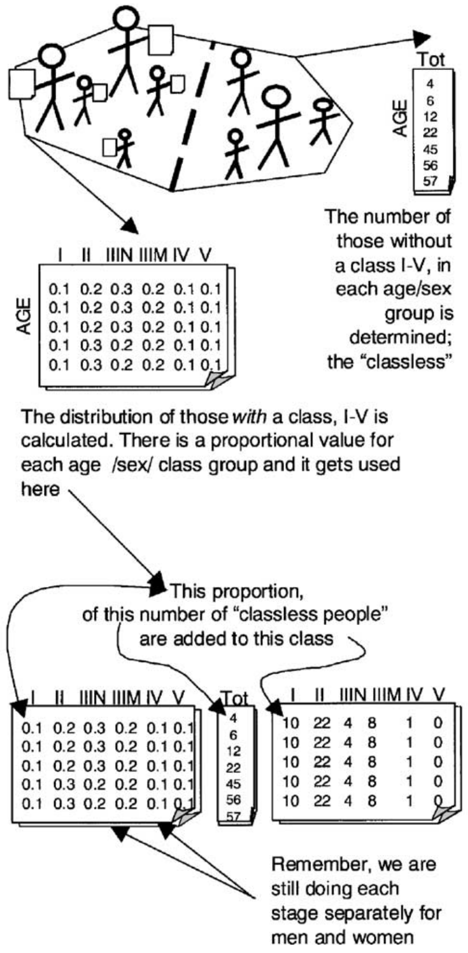


Fig. 3. An illustration of the principle by which those without a social class I–V were allocated one.

people who did have a class I–V, but in some areas the vast majority of people were not allocated a class by the census (retirement areas for example). Unlike the 1981 census, the 1991 census failed to report the previous occupations or classes of the retired. We, therefore, had no direct information through which to place the classless retired people into classes I–V in 1991. We also failed to recognise or cope with the radical change in the labour structure of women within the last few decades. Today many more women work in the formal labour market than at home as unpaid ‘housewives’, but since class in the census is based on formal occupation it fails to deal with women’s class position properly in those areas where many middle and older aged women still do not formally work. We were allocating these women a class based on their more unusual neighbours who *do* work formally. The net result of these two problems was a serious over allocation of middle and older aged people to lower social classes. Since this is a group of people at quite high risk of mortality we needed to get their class allocation as correct as possible! A more sophisticated version of the technique was required to deal with these complexities.

Geography was again the key to dealing with the problem. We needed to enhance the ‘donor’ information about class distribution from which the older and female classless population would be allocated to a class I–V. There were three possible sources of information available. We knew the I–V class distribution amongst slightly younger age groups within the ward. We also knew the age/class distribution for men. This could be applied to women at ages for which men’s class might be more representative of their own social position (because more of them were housewives). We also had access to the census data for the same ward from 1981 and 1971 and could therefore ‘look back in time’ to younger groups and / or look back in time to see what the neighbourhood class profile for older men and women was like in the 1970s and 1980s. These different ‘donor’ sources of information were blended together by adjusting their relative contributions according to the ‘recipient’ group in question.

For example, we combined one third of the class distribution of retired men for the ward in 1981 and two thirds of that for retired men in 1971 to give an appropriate distribution for men, aged 80–84 in 1991. Consider the group in question; nearly all of them will have been very close to retirement in 1971 and (we assumed) will then have exhibited a similar class profile to their slightly older, retired peers. However, since we only had information on the retired male population as a whole for 1971, and this included very much older men with perhaps a slightly different class structure again, we temper the influence of the 1971 retired with a contribution from the 1981 retired.

In some cases we used the class distribution of a younger age group to give class to an older group. This was particularly useful for the late 50s-early 60s age groups where there is a class bias amongst those who have taken early retirement. If we had classified those who were retired at 55 (and were thus classless) using those who were still working, we would have created too high a count of working class people. A higher proportion of those who retire early are drawn from the higher classes. The choice of weighting factors was initially based on intuition rather than any established criteria but it was tested and fine tuned in relation to Hattersley’s work.

In the end we used historical class structures and those of younger groups to differing extents for men and women of different ages, creating a complex set of rules for the code to follow in class allocation. Class allocation always used ‘donor’ information from the same spatial unit but blended that from different points in time and from different demographic groups. The full set of rules are specified in the technical report (Mitchell et al., 2000b).

Changing time—making data from 1981 and 1991 relevant for 1983 and 1993

At this point in the process we projected the age/sex/class data forward in time by 2 years to 1993 and 1983, respectively. The reader will recall that we needed data for these years because they fall in the middle of the two 5 year periods for which we had death records. The adjustment procedure was slightly different for the two time periods. To adjust the 1981 data to 1983 we combined the estimates from 1981 and 1991 using a simple formula. The number of people in a particular age/sex/class group in 1983 was estimated as being four times that in 1981, plus that in 1991, divided by 5. The formula had the effect of assuming that the changes which took place between 1981 and 1991, happened at the same rate each year. We literally "froze" that process of change in 1983. To adjust the 1991 data to 1993 was slightly more difficult since we did not have the luxury of a data set with identical structure for 2001. Instead, we were able to utilise ONS estimates of the numbers of people in each age/sex grouping in each ward for 1996, adjusted on a pro-rata basis to refer to 1993.

Employment status

With an appropriate social class (I–V) now allocated to everyone in the population model the last step was to further disaggregate the population aged 16–64 by employment status and distinguish between the employed, unemployed and economically inactive. For this stage, the data were aggregated to parliamentary constituency level. As before, there was a straightforward IPF element to this part of the process, enhanced with some geography. We will focus on the geographical element.

Disaggregation by employment status used the same type of IPF procedure as before but marginal values were derived from 1991 and 1981 census counts which describe numbers in each employment status category by age and sex, and by class and sex. These were converted to proportions and applied to aggregated counts by age and sex, and class and sex. The seed for IPF was provided by making an assumption that the rate of unemployment was uniform across age groups within the same social class.

Research has shown that the detrimental effect of unemployment on an individual's health does not occur after one day or even one month of unemployment, but after longer spells out of work (Bartley, 1994; Klein-Hesselink and Spruit, 1992). The influence of unemployment on death rates is substantial and it was important that we modelled the numbers of people who might have experienced *longer term* unemployment, and the change in those numbers over time, as accurately as possible. The great disadvantage of using the census to estimate these numbers is that it is only a 'snapshot' in time. For example, people who said that they were employed on census night in 1981 might have been made redundant the day after the census. Censuses also occur at particular points in the country's economic cycle. The 1991 census coincided with a particularly deep recession in the south. Had we used it as an indicator of how many people had been unemployed since the 1981 census our estimates would have been far too high. Geography influences the utility of the snapshot since the chances of long-term unemployment vary so markedly across the country. Making a simple blanket adjustment to every parliamentary constituency's unemployment rate to compensate for the national economic cycle would have led to greater inaccuracy.

Estimates of counts by age/sex/class/employment status which resulted from the IPF procedure were thus enhanced by an adjustment based on their geographical location. This allowed the estimates to better reflect the chances of unemployment for people living in each constituency over the previous 12 yr. The 12 yr figure was chosen because research on the LS, from which we drew our death rates, used a 12 yr time lag between recorded unemployment and mortality. To make the adjustment we compared the total number of people within each parliamentary constituency estimated to have been unemployed with the average number for the preceding 12 years, drawn from figures published by NOMIS. The proportional difference between the estimated counts for the single year and the 12 yr average from NOMIS was calculated. Each age/sex/class grouping of unemployed people was adjusted by adding or removing people relative to the proportional difference value. The same number of people were removed or added, respectively, from the same age/sex/class group who were employed. In this way, the number of people in the constituency stayed the same—people were just moved between the unemployed groups and the employed groups to better reflect the long-term labour market.

This process was automated within the custom written code. The common geographical basis between our figures and those from NOMIS allowed us to add temporal context to our geographic information.

At last, the end

So, at the end of this lengthy and complex process we had created as accurate a set of age/sex/class/employment status counts as possible. The entire process for 1983 and 1993 took approximately 0.3 billion calculations.

What is this process called?

Although we spent much time reviewing other approaches to population modelling and have drawn from several of them, it is quite a hard task to 'label' our approach. What we did probably does not qualify as *microsimulation* for two reasons. First, we did not actually generate or 'grow' synthetic individuals in the manner commonly associated with microsimulation. Second, there was no probabilistic or random component in the generation of the population counts. If we ran our population generation process again, we would get the same results. This is in contrast to microsimulation which usually includes some kind of random or probabilistic components so that no two populations generated by the same processing sequence will be precisely the same. A better term for our approach might be *estimation* because we have, in effect, estimated counts for groups of people with certain combinations of characteristics using known counts and relationships as a basis.

Are the results accurate?

It is important to consider the accuracy of the estimated counts. Of course, the fact that we needed to estimate population counts in this way because they are not available from any other source means that we have no directly comparable data with which to assess our model—we have estimated the unknown. In these situations a reasonable approach is to consider the possible sources of error. Perhaps the most significant source of error stemmed from our allocation of a social class to those people without one, based on the class distribution amongst peers within the neighbourhood and from earlier time periods. Since the influence of social class on mortality rates is so great it was very important for the wider research questions to get the right numbers of people in the right classes and in the right locations. We were able to ensure our class structures for Britain as a whole matched those produced by Hattersley (1999), and thus those on which the mortality rates we applied to the population counts were based. The potential for error was thus noted and minimised.

A further source of error might have stemmed from the accuracy of the initial census counts themselves. It is widely accepted that the 1991 UK census missed over 1 million people (Simpson and Dorling, 1994) and more people will have been missed in some areas and from particular social and demographic groups than others (Martin et al., forthcoming). Where possible, corrected counts for each age/sex group were applied. Again, this was designed to minimise the potential for error.

The IPF process itself also had the potential to be a source of error. Areas with quite extreme socio-demographic structures will not have been treated well by a system attempting to modify typical age/sex/class distributions to match local situations. When working at ward level, something akin to a large army barracks was enough to prevent the IPF converging on a solution. Error seeking routines in the computer code alerted us to wards in which the IPF approach was producing highly unlikely results or no results at all because of a failure to converge on a solution. Only 11 (of over 10,000) wards proved problematic in this way and we were forced to adopt the simple solution of imposing typical age/sex/class structures on those.

Whilst every effort was made to minimise potential sources of error stemming from the data and techniques employed, we felt it inappropriate to try and calculate a single statistic to represent our confidence in the resulting population model. We could only identify sources of error and had no meaningful way to quantify their impact. The best means of judging the quality of the estimates was to see how they performed. Following a description of how we used the population estimates, we will explain how this process strengthened our belief in their accuracy.

The application of appropriate mortality rates and subsequent geographical analysis of the results

Having explained the genesis of the data set, it would seem appropriate to provide a brief explanation of how it was employed. We derived mortality rates for the early 1980s and early 1990s for each age/sex/class/employment status group and by applying these rates to the counts of people with each combination of characteristics, in each parliamentary constituency, we were able to compute the number of deaths that constituency was expected to yield. By comparing the expected number in 1983 with 1993, we computed an expected change in mortality rate for that constituency. We were then able to compare this change value with that which actually took place in the numbers of deaths recorded in the constituency at each point in time. Our results showed that in 95% of constituencies we could account for the change in numbers of deaths to within 5% (Mitchell et al., 2000a).

Fig. 4 maps this information as a cartogram. Each parliamentary constituency is represented as a circle, shaded according to the degree of error between our model of change and that which actually occurred. Black shows those constituencies in which we underestimated the persistence of high, or rise in, mortality rates and the lightest grey show those constituencies in which we underestimated the persistence of low, or fall in mortality rates.

Finally, the application of the population data provided an opportunity to assess their accuracy. If the model was accurate, the national total number of expected deaths ought to have matched the actual number of deaths recorded. Any sizeable total difference or systematic difference (for particular regions or

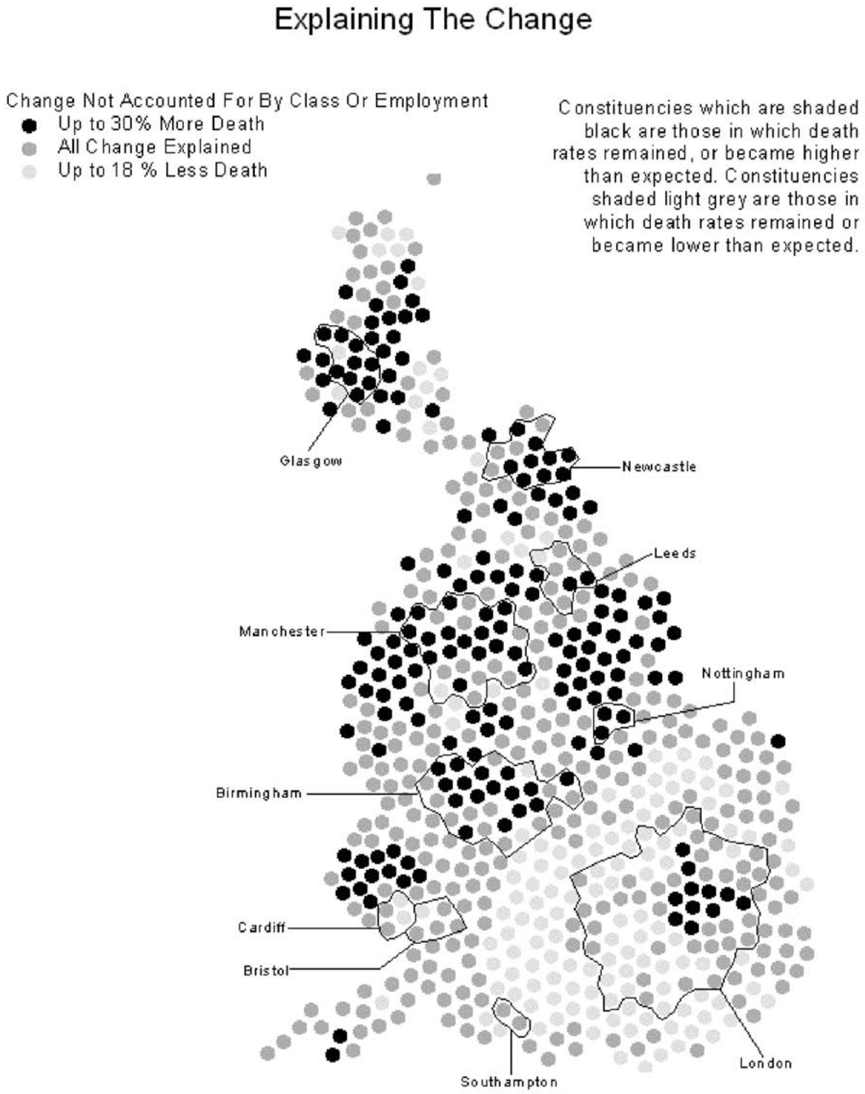


Fig. 4. Map showing results from the application of the population model to our research question.

socio-economic groups) between these two values would have indicated an inaccurate population model or inaccurate mortality rates. The differences found were small and random.

Summary and conclusions

Two individual level models of the whole British population were estimated from aggregate and disaggregate data sources using a custom written GIS. The primary role of geography and GIS here was to enhance the application of IPF and its results to estimate best the characteristics of resident populations. Both temporal and social aspects of neighbourhood and constituency character were exploited to create the best possible model. None of the techniques were particularly complex. Any complexity stemmed from the size of the task and the datasets involved. GIS continues to be developed as both a tool and a concept for use in health inequalities research.

Acknowledgements

We are grateful to Dimitris Ballas for assistance with the microsimulation literature. We are also grateful to the Joseph Rowntree Foundation for funding and support during this project. RM is employed by the Research Unit in Health, Behaviour and Change (RHUBC). RHUBC is funded by the Chief Scientist Office of The Scottish Executive Health Department

(SEHD) and the Health Education Board for Scotland (HEBS). The opinions expressed in this paper are those of the author(s) not of SEHD or HEBS.

References

Ballas, D., Clarke, G.P., 2001. Modelling the local impacts of national social policies: a spatial microsimulation approach. Environment and Planning C: Government and Policy 19 (4), 587–606.

Bartley, M., 1994. Unemployment and ill health—understanding the relationship. Journal of Epidemiology and Community Health 48, 333–337.

Birkin, M., Clark, M., 1988. SYNTHESIS—a synthetic spatial information system for urban and regional analysis: methods and examples. Environment and Planning A 20, 1645–1671.

Caldwell, S., Keister, L., 1996. Wealth in America: family stock ownership and accumulation, 1960–1995. In: Clarke, G.P. (Ed.), Microsimulation for Urban and Regional Analysis. Pion, London, pp. 88–116.

Caldwell, S.B., Clarke, G.P., Keister, L.A., 1998. Modelling regional changes in US household income and wealth: a research agenda. Environment and Planning C: Government and Policy 16, 707–722.

Clarke, G.P., 1996. Microsimulation: an introduction. In: Clarke, G.P. (Ed.), Microsimulation for Urban and Regional Policy Analysis. Pion, London.

Dorling, D., 1997. Death in BritainHow Local Mortality Rates have Changed: 1950s–1990s. Joseph Rowntree Foundation, New York.

Falkingham, J., Lessof, C., 1992. Playing God or LIFEMOD—the construction of a dynamic microsimulation model. In: Hancock, R., Sutherland, H. (Eds.), Microsimulation Models for Public Policy Analysis: New Frontiers. Suntory-Toyota International Centre for Economics and Related Disciplines—LSE, London, pp. 5–32.

Hancock, R., Mallender, J., Pudney, S., 1992. Constructing a computer model for simulating the future distribution of pensioners' incomes for Great Britain. In: Hancock, R., Sutherland, H. (Eds.), Microsimulation Models for Public Policy Analysis: New Frontiers. Suntory-Toyota International Centre for Economics and Related Disciplines—LSE, London, pp. 33–66.

Hattersley, L., 1999. Trends in life expectancy by social class—an update. Health Statistics Quarterly 2, 16–24.

Hattersley, L., Creeser, R., 1995. Longitudinal Study 1971–1991: History, Organisation and Quality of Data. HMSO, London.

Holm, E., Lindgren, U., Makila, K., Malmberg, G., 1996. Simulating an entire nation. In: Clarke, G.P. (Ed.), Microsimulation for Urban and Regional Analysis. Pion, London, pp. 164–186.

Johnston, R., Pattie, C., 1993. Entropy maximising and iterative proportional fitting. Professional Geographer 45, 317–322.

Klein-Hesselink, D., Spruit, I., 1992. The contribution of unemployment to socio-economic health differences. International Journal of Epidemiology 21, 329–337.

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.

Macintyre, S., Ellaway, A., Cummins, S., forthcoming. Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science and Medicine.*

Martin, D., Dorling, D., Mitchell, R., forthcoming. Linking censuses through time: problems and solutions. *Area.*

Mitchell, R., 1997. An Integration of aggregate and disaggregate data. Ph.D. Thesis, Southampton University.

Mitchell, R., Martin, D., Foody, G.M., 1998. Unmixing aggregate data: estimating the social composition of enumeration districts. Environment and Planning A 30, 1929–1941.

Mitchell, R., Dorling, D., Shaw, M., 2000a. Inequalities in Life and Death: What if Britain were More Equal? Joseph Rowntree Foundation, York.

Mitchell, R., Dorling, D., Shaw, M., 2000b. Inequalities In Life and Death—What if Britain were More Equal? The Technical Report http://www.social-medicine.com/ip-fexpl2/contents.html.

Piachaud, D., Sutherland, H., 2000. How effective is the British Government's attempt to reduce child poverty? CASE Paper 38, LSE/STICERD.

Simpson, S., Dorling, D., 1994. Those missing millions: implications for social statistics of non-response to the 1991 census. Journal of Social Policy 23, 543–567.

Shaw, M., Dorling, D., Davey Smith, G., 1998. Life Expectancy: A Research Report for Granada/Channel 4's 'The Drop Dead Show'. University of Bristol, Bristol.

Shaw, M., Gordon, D., Dorling, D., Davey Smith, G., 1999. The Widening Gap: Health Inequalities and Policy in Britain. The Policy Press, Bristol.

Shaw, M., Gordon, D., Dorling, D., Mitchell, R., Davey Smith, G., 2000. Increasing mortality differentials by residential area level of poverty: Britain 1981–1997. Social Science and Medicine 51, 151–153.

Shaw, M., Dorling, D., Mitchell, R., 2001. Health, Place and Society. Pearson Press, Harlow.

Sutherland, H., 2000. The British Government's attempt to reduce child poverty: a budget 2000 postscript. Microsimulation Unit Research Note No.36, Cambridge.

Vencatasawmy, C.P., Holm, E., Rephann, T., 1999. Building a spatial microsimulation model. Paper presented at the 11th Theoretical and Quantitative Geography European colloquium, Durham Castle, Durham, 3–7 September.

Williamson, P., Birkin, M., Rees, P., 1998. The estimation of population microdata by using data from small area statistics and samples of anonymised records. Environment and Planning A 30, 785–816.