PATTERNS OF SELF-ASSESSED HEALTH IN SCOTLAND: PEOPLE, PLACES AND INCOME INEQUALITY

Neil Craig
Public Health and Health Policy Section
Division of Community-Based Sciences
University of Glasgow
1 Lilybank Gardens
GLASGOW
G12 8RZ

0141 330 5017
0141 330 5018
Neil.Craig@udcf.gla.ac.uk

WORK IN PROGRESS: DO NOT QUOTE WITHOUT THE AUTHOR'S PERMISSION

ACKNOWLEDGEMENTS

The data used in this paper were made available through the UK Data Archive. The data were originally collected by NFO Social Research and MORI Scotland on behalf of the Scottish Executive. Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here. I am also grateful to Josephine Dean of the Scottish Household Survey for providing the electoral ward identifiers. I would like to thank Danny Mackay of the University of Glasgow for considerable assistance in equivalising the income data and Keith Murray of the University of Glasgow for help in preparing the electoral ward data. Thanks are also due to Matt Sutton of the University of Glasgow for useful discussions during the analysis on which this paper is based. Finally, I would like to thank Alastair Leyland of the MRC Social and Public Health Sciences Unit for advice on multilevel modelling.

ABSTRACT

Background
Associations between area characteristics (context) and health are significant but small relative to the associations between health and the characteristics of individuals living within areas. A contextual variable of particular interest is income inequality. A growing body of work suggests, firstly, that the association between income inequality and health found in many studies reflects individual or area level characteristics with which income inequality is associated, rather than an independent association with income inequality per se, and, secondly, that the nature and strength of the association is country-specific. An unresolved methodological issue is the geographical level at which to model contextual associations.

Methods
Multilevel logistic regression analysis was used to estimate the association between self-assessed general health (SAGH) and local authority (LA) income inequality, LA mean income and individual socio-economic characteristics. Random effects models in MLwiN estimated between-place variation in SAGH at local authority and electoral ward level before and after adjusting for individual characteristics. Data were from the 1999-2000 Scottish Household Survey.

Results
Socio-economic status and income inequality were positively associated with SAGH. More unequal areas were more affluent but mean LA income ceased to be positively associated with SAGH after adjustment for individual-level characteristics. Local authority and electoral ward-level variation in SAGH were small compared to the variation at the individual level but remained significant after adjustment for individual-level characteristics.

Implications
The income inequality:health association is country-specific suggesting different causal pathways in different countries. Further work is required to identify the area-level characteristics that explain the between-place variation in SAGH and to understand the policy implications.

Keywords
Income inequality, self-assessed health, multilevel, Scotland

Word count for main text: 6,365 words

INTRODUCTION

A substantial literature demonstrating a negative association between income inequality and mortality, both between and within countries, has focussed attention on the potentially damaging effect of income inequality on population health. The robustness of this association in cross-country studies has been challenged (Judge, 1995; Lynch, Davey Smith, Hillemeier, Shaw, Raghunathan, & Kaplan, 2001; Mellor & Milyo, 2001) and debate continues as to the causal mechanisms underpinning the relationship. Some authors suggest it reflects diminishing returns to absolute incomes at the individual level (Gravelle, 1998). Others implicate the physiological and psychological consequences of low social status in societies where incomes and other material resources are very unequally distributed (Wilkinson, 1996; Wilkinson, 1997). Others emphasise systematic variations between areas in the availability of private resources and health-related public infrastructure, variations that are related to income inequality because both share a common source in broader economic and social policies (Lynch, Davey Smith, Kaplan, & House, 2000; Coburn, 2004).

Explanations suggesting that the association between income inequality and health may not reflect the causal effect on health of income inequality *per se* are supported by a number of recent review articles (Wagstaff & Van Doorslaer, 2000; Mackenbach, 2002; Deaton, 2003; Lynch, Davey Smith, Harper, Hillemeier, Ross, Kaplan, et al, 2004) and empirical analyses (Deaton, 2001; Mellor & Milyo, 2002; Deaton & Lubotsky, 2003; Blakely, Atkinson, & O'Dea, 2003; Mellor & Milyo, 2003). The majority of studies in this field come from the US where the evidence on the income inequality hypothesis is mixed. A the growing literature from other affluent countries provides limited support for the hypothesis that income inequality exerts a generalisable, independent effect on health.

One line of argument that has developed in response (Subramanian, Blakely, & Kawachi, 2003a) to this challenge to the income inequality hypothesis is to highlight reasons why results in non-American studies might be expected to differ from those in the United States. These include lower *levels* of income inequality *within* areas, which would reduce the impact of income inequality if there is a threshold effect (Shibuya et al, 2002); smaller *differences* in income inequality *between* areas within countries, which would reduce the power of studies to detect the impact of income inequality (Gerdtham & Johannesson, 2001; Blakely et al, 2003); and the impact of more comprehensive programmes of welfare provision than in the US, which might offset the impact of income inequality (Osler et al, 2002).

The implication of this line of argument is that we need studies from a wider range of countries combining different levels of and variations in income inequality with different patterns of welfare provision. A recent study from Chile, for example, where income inequality is high, found that income inequality was significantly and negatively associated with individual health after adjusting for individual-level socio-economic and demographic status (Subramanian, Delgado, Jadue, Vega, & Kawachi, 2003b).

A second line of argument concerns the considerable methodological challenges faced in testing the impact on health of income inequality, in particular the appropriate modelling of 'multilevel' effects. The term 'multilevel' has been used in a general way to mean studies that use both individual and area-level data to model health effects of inequality. Generally such studies have found that associations between health and individual-level socio-economic status are stronger than those between health and income inequality (and other area-level variables) and that the inclusion of individual-level variables greatly attenuates the association between income inequality and health, often to the extent that it disappears altogether.

It has been suggested, however, that the methods used within the multilevel literature have not always been suited to multilevel analysis (Subramanian, Blakely, & Kawachi, 2003a). Where the study population falls naturally into a multilevel structure, for example, i individuals nested within j geographical areas, they are more likely to share certain characteristics than individuals in different areas. This imposes a correlation structure on the data that invalidates assumptions of independence common to many traditional regression techniques. Procedures are available that estimate standard errors adjusted for the clustering of data but these processes of adjustment merely treat intra-cluster correlation as a 'nuisance factor' to be controlled for rather than indicating a relationship or set of relationships of substantive interest in understanding the links between income inequality, health and other area and individual-level variables (Rice & Jones, 1997; Merlo, 2003).

Fixed effect models are often used in this context, controlling for area-level effects at one level that confound estimates of either area effects at other levels or the effects of individual-level variables with which higher level effects are correlated. However, fixed effect models have limitations as a way of exploring contextual effects in multilevel datasets. Firstly, they require a separate parameter to be estimated for (j – 1) dummy variables assigned to the areas. This can be problematic where there is a large number of areas. Secondly, in fixed effects models, area level effects of specific variables such as income inequality cannot be estimated at the level to which the dummy variables are assigned. Such effects are subsumed within the area level fixed effect as a whole, which means that the sources of area-level variation cannot be explored.

Thirdly, by definition, area-level variables assigned to individuals cannot vary between individuals in the same area. Therefore the number of observations for each area-level variable is the number of areas rather than the number of individuals within the sample. Fixed effect models that treat each individual as providing one independent observation of the area-level variable underestimate the standard errors on the estimated area-level parameters which has implications for statistical inference regarding parameter estimates.

An alternative approach is to use multilevel modeling techniques, sometimes called *inter alia* random effects models or variance components models. Such models correctly estimate standard errors allowing for within-area clustering and, at the area-level, the number of areas for which data are available. They are a parsimonious way of estimating area effects because they do not require a separate parameter to be estimated for each area (Subramanian et al, 2003). They allow the variance to be partitioned into the component attributable to differences between areas, before and after estimating the effects of the independent variables, which gives an indication of the importance and causes of area-level effects. The proportion of the total variance attributable to differences between areas gives an indication of the extent to which the health of people within the same area is similar compared to the health of those in other authorities (Merlo, 2003). This is referred to as the intracluster correlation.

These arguments for multilevel modeling do not mean that it is always the most appropriate modeling strategy to use. The choice of approach depends partly on the questions of interest. If the parameter estimates for particular areas or groups are of interest rather than the extent or cause of the area/group effects, fixed effects models are appropriate (Subramanian et al, 2003a). Fixed effect models also tend to lead to more consistent parameter estimates where individual-level variables and area effects are correlated (Rice & Jones, 1997).

In practice, the estimates derived from fixed and random effects models can differ substantively. For example, Subramanian et al (2003a) reworked the results of a three-level analysis of the effects of income inequality undertaken in the US (Mellor and Milyo, 2003). Individuals were nested in states, which were nested in nine 'census divisions'. The original study used a fixed effects model with dummy variables assigned to the divisions to adjust for divisional-level confounding of the state level relationship between income inequality and health. The original fixed effects specification failed to find a significant effect of state-level income inequality on health whereas Subramanian et al's multilevel analysis found "modest" but statistically significant effects on health of state level income inequality. (For discussions of the appropriate circumstances and the effects of using fixed and random effects models, see Rice & Jones, 1997, and Blundell & Windmeijer, 1997).

Even if multilevel modelling is considered to be the most appropriate strategy to use, key methodological issues remain unresolved, one of which is the choice of geographical level at which area effects are estimated. It has been suggested that differences between studies in the strength of the association between income inequality and health may reflect the different levels of geographical aggregation at which analyses have been carried out (Weich et al, 2002; Lynch et al, 2004). Ideally, the choice of levels should be determined by clear hypotheses about the mechanisms by which income inequality is thought to affect health and the levels at which these mechanisms might operate. However, this remains a contested area (Wagstaff & Van Doorslaer 2000; Wilkinson, 1997). In addition, the levels at which analyses can be carried out in practice are constrained by data availability.

This paper therefore has two main objectives. Firstly, it uses data from the Scottish Household Survey (SHS) to test the hypothesis that income inequality at the level of local authorities in Scotland is associated with individuals' health, after accounting for individual demographic and socio-economic status. Secondly it measures intra-area correlation in self-assessed health at different geographical levels to identify at which level it might be useful to explore contextual effects on health. A subsidiary aim is to test whether estimates of the extent of any association between income inequality and health are sensitive to the way in which the multilevel structure of the data is addressed.

METHODS

Data

Data come from the first two years of the SHS, 1999-2000. The SHS is a continuous cross-sectional survey based on a systematic random sample in areas of high population density and cluster random sampling in the remaining areas (Anderson, Hope & Martin, 2001). The SHS provides data on household income, health and other socio-economic and demographic variables at the individual level and by local authority. The SHS survey manager also supplied electoral ward identifiers. These enabled people from the same ward to be grouped but the ward could not be identified for confidentiality reasons.

The survey is in two parts. The highest income householder or their spouse or partner answers part 1 on behalf of all members of the household. 30,227 Part 1 interviews were completed. The overall response rate was 66%, varying from 79% in the Western Isles, a remote low-income rural area, to 59% in East Renfrewshire, an affluent suburb of Glasgow. Part 2 is answered by an adult aged 16 or over selected at random from the household members. 28,340 'random adult' interviews were carried out.

Measure of income and income inequality

The SHS provides data on household income from earnings, benefits and a variety of miscellaneous sources of the highest income householder and their spouse after tax and other deductions. Income data are available on 29,574 of the 30,227 households for which interviews were carried out. The data *exclude* the incomes of other members of the household.

The income data were equivalised using the McClements scale before housing costs and deflated to December 2000 prices using the retail price index including housing costs. Respondents were categorised into quintiles on the basis of the income of the household in which they lived. Income inequality by local authority was measured using the Gini coefficient. Area-level income was measured by mean rather than median income to avoid over-controlling for income inequality (Blakely & Kawachi, 2001). For both mean area income and income inequality, local authorities were divided into 4 categories: greater than or less than one standard deviation below the mean for Scotland as a whole and greater or less than one standard deviation more than the mean. To reduce the effect of outlying values, both mean local authority income and the Gini coefficients were calculated excluding the top and bottom 1% of the income distribution. This left 28,962 cases on which the mean area income and income inequality estimates were based, ranging from 464 in East Renfrewshire to 3001 in Glasgow.

Measure of health

The measure of health is self-assessed general health (SAGH). The SHS asks whether over the previous 12 months the respondent's health has been good, fairly good or not good. The measure has been shown to be a consistent independent predictor of mortality, even after controlling for other health status indicators and covariates of ill health (Idler & Benyammi, 1997). Responses were dichotomised into "good" SAGH and "less than good" SAGH. "Less than good" SAGH includes respondents rating their health as either "fairly good" or "not good". This approach has been used in many studies, although results regarding trends or differences in inequalities in health can be influenced by where the cut-off between "good" and "less than good" health is drawn (Van Doorslaer, Wagstaff, Bleichrodt, Calonge, Gerdtham, Gerfin, et al., 1997).

In the SHS, self-assessed health is only asked of the random adults in Part 2 of the survey. Therefore, for all variables except income, the analyses below use random adult data.

Non-income measures of socio-economic status and demographic variables

Age, sex economic status and highest educational qualification were included as independent variables. The analyses were restricted to the 16-64 age group. Economic status was defined as three levels: 'working', 'unemployed', and 'economically inactive'. The 'working' category includes those on government training schemes at work or college. The 'unemployed' category includes those actively seeking work. Highest educational qualification was categorised into four levels: degree or equivalent, higher or equivalent, O grade or equivalent, and none. The proportion of the sample with no qualifications and the proportion either unemployed or economically inactive were calculated for each local authority and electoral ward.

Inherent in cross-sectional analyses is the possibility of reverse causation. Low income may be a result rather than a determinant of poor health. To reduce the potential scale of this effect, respondents who attributed their economic inactivity to a limiting long-standing illness or disability, that is people whose poor health is likely to have preceded their low incomes, were excluded from the analysis, a method used in previous studies of the association between income and health (Aberg Yngwe, Diderichsen, Whitehead, Holland, & Burstrom, 2001; Benzeval, Judge & Shouls, 2001).

Statistical analysis

Multilevel logistic regression analysis in MLwiN was used to calculate the log odds of poor SAGH by income inequality category, by local authority mean income category and by levels of the individual-level socio-economic variables. The lowest inequality category, the highest local authority income category, the richest quintile, degree or equivalent, and economically active (in training or employment) formed the reference categories.

Area-level variation was estimated at local authority (level 3) and electoral ward (level 2) with individual respondents at level 1. Variance was partitioned into the components attributable to differences between local authorities and wards, from which the intracluster correlation coefficient was estimated by dividing the area-level variance at each level by the total variance based on the approximation that level 1 variance in a binomial model is $\phi^2/3$ (Snijders & Bosker, 1999; Goldstein, Browne & Rasbash, 2002).

All estimates were derived using the second-order penalised quasi-likelihood (PQL) options provided in MLwiN. First order marginal quasi-likelihood procedures tend to underestimate variance parameters where there are few level 1 units within higher-level units (Rasbash, Browne, Goldstein, Yang, Plewis, Healy et al, 2002). In such circumstances, it is also more appropriate to use restricted iterative generalised least squares (RIGLS) rather than iterative generalised least squares (Rice, 2001). The results below are based on RIGLS. Estimates of log-likelihoods for models using these methods on binary data are unreliable so inference is restricted to parameter estimates and associated variances (Rice, 2001).

RESULTS

Table 1 gives mean income, the prevalence of poor SAGH and Gini coefficient by local authority. Gini coefficients range from 0.261 in North Lanarkshire to 0.302 in Edinburgh with a Gini coefficient for Scotland as a whole of 0.287. There was a significant positive correlation between mean local authority income and the Gini coefficient (Pearson's r# #0.495, p=0.0040 and a significant negative association between the Gini coefficient by local authority and the rate of poor self-assessed health (Pearson's r# #40.629, P<0.001). The association between mean area income and the prevalence of poor self-assessed health was negative and significant (Pearson's r# #40.430, p=0.014)[1]. This association strengthened markedly if the Western Isles were excluded. The Western Isles are an outlier in that both the mean income and the prevalence of poor SAGH in are the lowest of all Scottish local authorities.

Exclusion of those who are economically inactive due to a limiting longstanding illness or disability left 20,304 16-64 year olds, 18,466 of whom had data on, income, self-assessed general health and the other socio-economic status and demographic variables included in these analyses. Sixty-three cases were missing data on electoral ward and a further 195 individuals from electoral wards that straddled local authority boundaries were excluded. For the multilevel analysis this left 18,208 individuals nested in 1206 electoral wards (range 1-49 people per ward) in 32 local authorities (range 17-79 wards per LA). Nearly 38% (37% if the data are weighted[2]) of these respondents have less than

[1] Correlation coefficients were calculated after weighting to take account of the relative sample size in each local authority. Unweighted correlations were slightly weaker but remained statistically significant, except the correlation between local authority mean income and the prevalence of poor self-assessed health when the Western Isles were included: r = -0.227, p=0.212. Excluding the Western Isles, this association was significant.

[2] Weighting is necessary in some analyses based on SHS data to take account of design effects and non-

good SAGH, less than in Table 1 because of excluding the elderly and those who were economically inactive due to a long standing illness of disability.

Table 2 gives the results of three random effects models: Model 1, a 'null' model with no independent variables in the fixed part and variance components for electoral wards (level 2) and local authorities (level 3); Model 2 which adds individual and household-level predictors to Model 1; and Model 3 which adds local authority level predictors (mean income and income inequality) to Model 2. In columns 4 and 5 the results are expressed in terms of the log odds of less than good health relative to the reference category for each variable. The final column expresses the results from Model 3 in terms of the odds ratios for less than good self-assessed health relative to the reference category derived by taking the exponential of the estimates of the log odds in column 5.

The null model (Model 1) shows that local authority and electoral ward-level variation (the 'variance components') in SAGH were significant but small compared to the variation at the individual level. Area-level variances of 0.016 and 0.085 imply that only 0.5% and 2.5% of the total variance in SAGH is accounted for by variance at the local authority and ward levels respectively.

Model 2 shows that the measures of individual-level socio-economic status were positively associated with health, as would be expected. The log odds of less than good SAGH health increased with lower socio-economic status. The inclusion of individual-level covariates explained just over 21% of the variance at the ward level but the area-level variance remained significant. Local authority level variance actually increased slightly, again remaining small but significant.

In Model 3, the inclusion of mean local authority income inequality and mean income had little effect on the variance components, suggesting they explain little of the residual heterogeneity at the electoral ward or local authority level after inclusion of individual-level socio-economic status variables. Greater income inequality and higher mean area income were associated with a *lower* likelihood of poor health although none of the fixed parameters on the income inequality or mean local authority income were significant. Area-level variances of 0.02 and 0.067 imply that only 0.6% and 2.0% of the total variance in SAGH is accounted for by variance at the local authority and ward levels respectively, after adjusting for individual level and area level variables.

A single-level specification for Model 3 was also estimated. The estimated coefficients changed little but the standard errors were reduced such that the positive association between income inequality and health became significant and the odds ratio in the lowest mean local authority income category became significantly different from 1.

Models 2 and 3 suggest that mean LA income and LA income inequality do not account for the small but significant LA-level variation. Other LA-level variables, some derived from the SHS survey data (proportion of respondents with no qualifications, proportion

---

response. The impact of weighting on both descriptive statistics and the estimated models (data not shown) was small.

of respondents not economically active) and one from other data sources (the Scottish Index of Material Deprivation), were also introduced as independent variables but the estimated coefficients on these variables were not significant and they had negligible impact on either the variation at the LA or ward level, or on the coefficients on the income inequality categories.

Exploring electoral ward-level variation was restricted by the design of the survey. The small number of respondents in many of the wards limited the calculation of ward-level variables from the SHS data. For example, 13 of the 1206 wards had 1 respondent, 14% had 10 respondents or less. The median number of respondents per ward was 18. In addition, ward identifiers were 'anonymised' so ward-level variables from sources other than the SHS could not be attached to the individuals within the SHS.

To explore ward level variation further in light of these constraints, respondents in wards with 18 respondents or more (n=9,686, 53% of the sample) were selected and ward-level variables (mean ward income, proportion of respondents with no qualifications, proportion of respondents not economically active) were calculated. Models 1 to 3 were re-estimated on this reduced dataset and an additional model including these three ward-level variables was also estimated.

The results of Models 1 to 3 on the reduced dataset were very similar to those on the full dataset. For example, in the null model, Model 1, the estimated coefficients (standard errors) for the constant term, the local authority variation and the ward-level variation were –0.524 (0.036), 0.014 (0.010) and 0.087 (0.020) respectively. The corresponding figures in Table 2 based on the full dataset were –0.506 (0.030), 0.016 (0.007) and 0.085 (0.015). In Model 3, the estimated coefficient on the highest income inequality category was -0.179 (0.152), giving an odds ratio of 0.836 (95% CI 0.633 – 1.070) relative to the lowest inequality category. This compares with –0.195 (0.134), odds ratio 0.823 (95% CI 0.621 – 1.126) based on the full dataset.

Inclusion of the ward-level variables in Model 3 had little impact on either the estimated ward-level variation or the estimated association between income inequality and SAGH. Having fallen from 0.087 (0.020) in Model 1 to 0.064 (0.018) in Model 3, the ward level variation only fell to 0.061 (0.018) when the ward-level variable were added. The estimated coefficient on the highest inequality category was –0.171 (0.156).

DISCUSSION

The study had three aims, firstly, to test the hypothesis that income inequality at the level of local authorities in Scotland is associated with individuals' health; secondly, to explore intra-area correlation in self-assessed health at different geographical levels to try to understand at what level, if any, contextual effects may be significant; and thirdly, to test whether estimates of the extent of any association between income inequality and health are sensitive to the way in which the multilevel structure of the data is addressed. Each is discussed in turn.

Income inequality hypothesis
This study finds little support for the hypothesis that areas with higher income inequality have worse health, after adjusting for individual-level associations between socio-economic status and health. If anything, the results suggest the reverse. In Scotland, local authorities with higher income inequality have higher incomes and lower rates of poor self-assessed health, even after adjusting for individual socio-economic status, although this relationship is weak. An unexpected but weak *positive* association between mean area income and the likelihood of poor health was also found. Individual-level associations are strong and have the expected sign.

Whilst the failure to find a positive relationship between income inequality and the risk of poor self-assessed health is not perhaps surprising, given the growing US and international literature with similar findings (Deaton, 2003; Lynch et al, 2004), the negative relationship was unexpected. Only two other studies to date have found better health in areas of higher inequality after adjusting for other individual and area characteristics (Wen, Browning & Cagney, 2003; McLeod et al, 2003). Both suggest causal explanations working through the impact of the tax contributions and other resources of those at the top end of the income distribution supporting the social infrastructure within neighbourhoods that enhances health for all. These explanations are consistent with the positive association between income inequality and mean local authority income noted earlier, an association also found at the regional level in the UK in two recent studies by Weich et al (2001, 2002). If these explanations are correct, the negative association between income inequality and health in Scotland may reflect unmeasured area-level income-related health-enhancing characteristics of the more unequal local authorities in Scotland, which tend to be the more affluent.

However, these explanations are not consistent with the result that, after adjusting for the other covariates, the significant negative association between area-level income and the likelihood of poor self-assessed health becomes a positive albeit insignificant association. This result is, however, consistent with the relative income hypothesis in that holding individual income constant, increasing local authority incomes corresponds to a fall in the individual's relative income (Wagstaff & Van Doorslaer, 2000). It is also possible that other area-level characteristics are offsetting the impact of low socio-economic status at the area level, such that having accounted for low status at the individual level, health is not as bad in poorer areas and is worse in some of the more affluent areas than might otherwise be expected.

Contextual effects
Area level effects were significant in this study but very small. The intracluster correlation suggests that less than 1% of the variation in poor self-assessed health arises from differences between local authorities and around 2-2.5% from differences between electoral wards. This is similar to the findings reported at the state-level in the United States (Subramanian, Kawachi & Kennedy, 2001; Subramanian et al, 2003a) and the neighbourhood level in Chile (Subramanian et al, 2003b) and is consistent with a review of multilevel studies by Pickett and Pearl (2001).

Two caveats need to be considered in interpreting these findings. The first is methodological. The estimates of the extent of clustering at the LA and ward levels uses an approach based on an approximation of the level 1 (i.e. individual level) variation ($\phi^2/3$). In logistic regression, the individual level variation is a function of the fixed part of the model, that is, the estimates for the intercept and the parameters on the independent variables, so the accuracy of this approximation is conditional upon the fixed part. Alternative methods exist for estimating the extent of clustering, which are discussed elsewhere (Goldstein et al, 2002). The sensitivity of the estimates of the intracluster correlation to alternative methods has not been tested. The second caveat is epidemiological. Small effects to which a lot of people are exposed may still imply a significant public health problem. The relatively small area level effects identified in this study should not be taken to mean that area-level effects are unimportant.

Modelling multilevel effects

The estimated parameters differed little in the single-level and multilevel random effects models but the statistical significance of the estimates changed. This may suggest that the dangers arising from not using random effects models in analyses of the effects of income inequality in Scotland may lie in the possibility of misinference about the statistical significance of contextual effects rather than parameter misestimation. Further analyses are required to assess the generalisability of this conclusion.

One factor upon which the generalisability might depend is the appropriateness of the geographical level at which income data have been aggregated. Ideally, of course, the choice of level should correspond to the level at which income inequality is thought to affect health. However, this remains a highly contested area (see, for example, Wilkinson, 1997; Wagstaff and Van Doorslaer, 2000). The choice of level in this study was driven mainly by the structure of the SHS, which only provides data at the level of local authorities. The ‘anonymised’ electoral ward identifiers prevented ward level variables from other sources being attached to the respondents in the SHS and it prevented adjacent wards being grouped into clusters representing larger geographical areas. This would have enabled different levels of geographical aggregation to be explored, and provided larger sub-samples in each area from which to calculate ecological variables based on the survey data, such as income inequality.

Another issue raised by the choice of area is whether they exhibit high enough levels of inequality and wide enough variation in levels of inequality to identify effects of inequality. What ‘enough’ means in this context is an empirical question, which can only be answered with more studies of this kind. The range of Gini coefficients (0.261 to 0.302) is comparable to the range estimated by Weich et al (2002) for the UK as a whole (0.258 to 0.307) with the exception of Inner London which had substantially higher income inequality (Gini coefficient 0.324).

Part of the rationale for using local authorities as the level of geographical aggregation in was that authorities have responsibility for a range of services thought to influence health. Social services, education, public sector housing and a range of environmental services are organised and financed at local authority level. However, this does not preclude the possibility that the availability and quality of services might vary between inhabitants *within* authorities, in particular in some of the mixed urban and rural authorities, hence the rationale for using a lower level of aggregation such as electoral wards. Other multilevel studies have found stronger intracluster correlation within smaller areas.

Limitations of the study

Firstly, within-country studies of this kind cannot rule out the possibility of effects on population health of income inequality at a national level (Blakely et al, 2003). Scotland's health may, for example, be compromised by income inequality across Scotland as a whole, or by income inequality across the UK both of which have increased over the last 25 years (Department of Social Security, 2000; Goodall, Fleming & Murray, 2002).

A second limitation is that the SHS was not designed to collect reliable economic statistics and the survey managers warn that economic data should only be used for selecting groups, such as the low paid, for further analysis or to provide background on individuals when analysing other topics. Specific concerns are a high proportion of imputed income data and the potential impact of the exclusion of incomes of adults other than the highest income householder and their spouse or partner.

Despite these concerns, the Gini coefficient in this study of 0.287 for Scotland as a whole is similar to figures of 0.281 in Gravelle and Sutton (2003) for 1995 based on GHS data and 0.285 in Weich et al (2002) for 1991 based on Wave 1 of the British Household Panel Survey. Both studies excluded the first and 99th centiles of the income distribution to reduce the effect of outliers. In addition, prior to estimating the multilevel random effects specification of Model 3, a single level specification was estimated using only the cases for which no income data were imputed. The impact on the results was negligible (data not shown).

Another limitation is the rather ad hoc method for diminishing the potential impact of reverse causation (i.e. from health to incomes) that might explain the income:health associations observed in cross-sectional data. However, a number of studies suggest that this explains a small proportion of the observed association between incomes and health (Benzeval & Judge, 2001; Davey Smith, Neaton, Wentworth, Stamler, & Stamler, 1996; Ettner, 1996; Shaw, Dorling, Gordon, & Davey Smith, 1999; Whitehead, Burstrom & Diderichsen, 2000; Wilkinson 1989). Its potential impact in this study was reduced by the exclusion of those who were economically inactive due to a longstanding illness or disability.

This is a stringent test of the income:health association. Just as it cannot be assumed that the direction of causation runs from low income to poor self-assessed health, nor can we assume the reverse. Long-standing illness may be one way in which the impact of low income is felt and even amongst those whose income *is* compromised by a long-standing illness, there may be subsequent effects on health of the low income arising from this illness. If so, analyses excluding those with a long-standing illness will underestimate the income:health association attributable to a causal link running from income to health. Its potential impact on the income inequality:health association is less clear, especially given that the excluded proportion of the sample varied markedly between local authorities. It was highest in three of the lower income, worse health authorities (Glasgow, Inverclyde and North Lanarkshire), two of which were also amongst the least unequal authorities (Inverclyde and North Lanarkshire).

Finally, the models adjust for a limited range of socio-demographic and socio-economic variables. The most obvious omission is social class. In the SHS, the occupational information required to derive social class is only gathered if the person is in employment or has been in the past five years. There are only around 15,500 cases with social class data in the 16-64 age group after excluding those who are economically inactive due to a limiting long-standing illness, and the missing data are not representative of the sample as a whole in terms of either health or income. Therefore, social class was excluded.

CONCLUSIONS

This study provides further support for the argument that the income inequality:health association is place-specific. There are many reasons why the results of studies of the income inequality:health association may differ between countries, but perhaps the most fundamental is that the characteristics of the more unequal areas differ in different countries. In Scotland, for example, in contrast to states in the US, higher income inequality is more a characteristic of affluent areas than poorer areas.

The small area effects in this study are consistent with the wider literature. Nevertheless, the area-level variation, even after adjustment for individual and area-level predictors, was significant at both ward and local authority levels, and the direction of the association between self-assessed health and the area-level variables was unexpected, although the associations were not significant.

Finally, this study also suggests that methodological choices regarding the ways of estimating contextual level effects on health of income inequality and mean area income may not make a substantive difference to the results when contextual effects are small in relation to individual level effects.

This study is very much work in progress. Further research is required to explore possible reasons for these findings and to establish the generalisability of the conclusions regarding the strength and nature of contextual effects on SAGH and on health generally in Scotland and the UK. Questions of particular interest include:

1. the interpretation, explanation and implications of the results of this study;
2. alternative specifications of the local authority and ward-level effects;
3. possible directions for future research, in terms of both analytical techniques and data sources.

REFERENCES

1. Aberg Yngwe, M., Diderichsen, F., Whitehead, M., Holland, P., & Burstrom, B. (2001). The role of income differences in explaining social inequalities in self rated health in Sweden and Britain. *Journal of Epidemiology and Community Health 55*, 556-561.
2. Anderson, S. Hope, S., & Martin, C. (2001). *Scotland's People: Results from the 1999/2000 Scottish Household Survey (Volume 4: Technical Report)*. Edinburgh: Scottish Executive. http://www.scotland.gov.uk/stats/bulletins/00116-00.asp
3. Benzeval, M., Judge, K., & Shouls, S. (2001). Understanding the relationship between income and health: how much can be gleaned from cross-sectional data? *Social Policy and Administration 35* (4), 376-396.
4. Benzeval, M., & Judge, K. (2001). Income and health: the time dimension. *Social Science & Medicine 52* , 1371-1390.
5. Blakely, T. A., Atkinson, J., & O'Dea, D. (2003). No association of income inequality with adult mortality within New Zealand: a multilevel study of 1.4 million 25-64 year olds. *Journal of Epidemiology and Community Health 57*, 279-284.
6. Blakely, T. A., & Kawachi, I. (2001). What is the difference between controlling for mean versus median income in analyses of income inequality? *Journal of Epidemiology and Community Health 55*, 352-353.
7. Blundell, R., & Windmeijer, F. (1997) Cluster effects and simultaneity in multilevel models. *Health Economics 6*, 439-443.
8. Coburn, D. (2004). Beyond the income inequality hypothesis: class, neo-liberalism, and health inequalities. *Social Science & Medicine 58*, 41-56.
9. Davey Smith, G., Neaton, J. D., Wentworth, D., Stamler, R., & Stamler, J. (1996). Socioeconomic differentials in mortality risk among men screened for the Multiple Risk Factor Intervention Trial: I. White men. *American Journal of Public Health 86* (4), 486-496.
10. Deaton, A. (2001). *Relative Deprivation, Inequality and Mortality*. Cambridge, Massachusetts: National Bureau of Economic Research, Working Paper 8099.
11. Deaton, A. (2003). Health, inequality and economic development. *Journal of Economic Literature*, *41* (March), 113-158.
12. Deaton, A., & Lubotsky, D. (2003). Mortality, inequality and race in American cities and states. *Social Science & Medicine 56*, 1139-1153.
13. Department of Social Security. (2000). *Households Below Average Income 1994/5 to 1998/9.* London: Department of Social Security.
14. Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of Health Economics 15*, 67-85.
15. Gerdtham, U.-G., & Johannesson, M. (2001). *Absolute income, relative income, income inequality and mortality*. Stockholm: Stockholm School of Economics.
16. Goldstein, H., Browne, W. J and Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics 1*: 223-232.
17. Goodall, J., Fleming, A., & Murray, R. (2002). *Scottish Economic Statistics 2002 A4 Income Inequality in Scotland.* Edinburgh: Scottish Executive.
18. Gravelle, H. (1998). How much of the relation between population mortality and

unequal distribution of income is a statistical artefact? *BMJ 316*, 382-385.
19. Gravelle, H., & Sutton, M. (2003). Income-related inequalities in self-assessed health in Britain: 1979-1995. *Journal of Epidemiology and Community Health 57*, 125-129.
20. Idler, E. L., & Benyammi, Y. (1997). Self-rated health and mortality: a review of 27 community studies. *Journal of Health and Social Behaviour 38*, 21-37.
21. Judge, K. (1995). Income distribution and life expectancy: a critical appraisal. *BMJ 311*, 1282-1285.
22. Lynch, J. W., Davey Smith, G., Hillemeier, M., Shaw, M., Raghunathan, T., & Kaplan, G. A. (2001). Income inequality, the psychosocial environment and health: comparisons of wealthy nations. *The Lancet 358*, 194-200.
23. Lynch, J. W., Davey Smith, G., Kaplan, G. A., & House, J. S. (2000). Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions. *BMJ 320*, 1200-1204.
24. Lynch, J.W., Davey Smith, G., Harper, S., Hillemeier, M., Ross, N., Kaplan, G.A., Wolfson, M. (2004) Is income inequality a determinant of health? Part 1. A systematic review. *Milbank Quarterly 82*, (1), 5-99.
25. Mackenbach, J. P. (2002). Income inequality and population health. *BMJ 324*, 1-2.
26. McLeod, C. B., Lavis, J. N., Mustard, C. M., & Stoddart, G. (2003). Income inequality, household income, and health status in Canada: a prospective cohort study. *American Journal of Public Health 93* (8), 1287-1293.
27. Mellor, J. M., & Milyo, J. (2001). Re-examining the evidence of an ecological association between income inequality and health. *Journal of Health Politics, Policy and Law 26*, 487-522.
28. Mellor, J. M., & Milyo, J. (2002). Income inequality and individual health: evidence from the Current Population Survey. *Journal of Human Resources 37*, 510-539.
29. Mellor, J. M., & Milyo, J. (2003). Is exposure to income inequality a public health concern? Lagged effects of income inequality on individual and population health. *Health Services Research 38* (1), 137-151.
30. Merlo, J. (2003). Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *Journal of Epidemiology and Community Health 57*, 550-552.
31. Osler, M., Prescott, E., Gronbak, M., Christensen, U., Due, P., & Engholm, G. (2002). Income inequality, individual income, and mortality in Danish adults: analysis of pooled data from two cohort studies. *BMJ 324*, 13-16.
32. Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health 55*, 111-122.
33. Rasbash, J., Browne, W., Goldstein, H., Yang, M. Plewis, I. Healy, M., Woodhouse, G., Draper, D., Langford, I. & Lewis, T. (2002) *A users guide to MLwiN: Version 2.1d for use with MLwiN 1.10*. London: Centre for Multilevel Modelling, Institute for Education, University of London.
34. Rice, N. (2001) Binomial regression. Chapter 3 in Leyland, A.H. & Goldstein, H. *Multilevel Modelling of Health Statistics*. Chichester: John Wiley.

35. Rice, N. & Jones, A. (1997) Multilevel models and health economics. *Health Economics 6*, 561-75.
36. Shaw, M., Dorling, D., Gordon, D., & Davey Smith, G. (1999). *The widening gap. Health inequalities and policy in Britain.* Bristol: The Policy Press.
37. Shibuya, K., Hashimoto, H., & Yano, E. (2002). Individual income, income distribution, and self rated health in Japan: cross sectional analysis of nationally representative sample. *BMJ 324*, 16-19.
38. Snijders, T., & Bosker, R. (1999) An introduction to basic and advanced multilevel modelling. London: Sage Publications.
39. Subramanian, S.V., Kawachi, I., & Kennedy, B.P. (2001). Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Social Science & Medicine 53,* 9-19.
40. Subramanian, S. V., Blakely, T. A., & Kawachi, I. (2003a). Income inequality as a public health concern: where do we stand? Commentary on "Is exposure to income inequality a public health concern?" *Health Services Research 38* (1), 153-167.
41. Subramanian, S. V., Delgado, I., Jadue,L., Vega,J., & Kawachi, I. (2003b). Income inequality and health: multilevel analysis of Chilean communities. *Journal of Epidemiology and Community Health 57*, 844-848.
42. Van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U-G., Gerfin, M., Geurts, J., Gross, L., Hakkinen, U., Leu, R.E., O'Donnell, O., Propper, C., Puffer, F., Rodriguez, M., Sundberg, G., & Winkelhake, O. (1997). Income-related inequalities in health: some international comparisons. *Journal of Health Economics 16*, 93-112.
43. Wagstaff, A., & Van Doorslaer, E. (2000). Income inequality and health: what does the literature tell us. *Annual Review of Public Health 21*, 543-567.
44. Weich, S., Lewis, G., & Jenkins, S. P. 2001. Income inequality and the prevalence of common mental disorders in Britain. *British Journal of Psychiatry 178*, 222-227.
45. Weich, S., Lewis, G., & Jenkins, S. P. 2002, Income inequality and self-rated health in Britain. *Journal of Epidemiology and Community Health 56*, 436-441.
46. Wen, M., Browning, C.R. & Cagney, K.A. (2003). Poverty, affluence and income inequality: neighbourhood economic structure and its implications for health. *Social Science & Medicine 57*, 843-860.
47. Whitehead, M., Burstrom, B., & Diderichsen, F. (2000). Social policies and the pathways to inequalities in health: a comparative analysis of lone mothers in Britain and Sweden. *Social Science & Medicine 50*, 255-270.
48. Wilkinson, R. G. (1989). Class mortality differentials, income distribution and trends in poverty 1921-1981. *Journal of Social Policy 18* (3), 307-335.
49. Wilkinson, R. G. (1996). *Unhealthy Societies: The Afflictions of Inequality.* London: Routledge.
50. Wilkinson, R. G. (1997). Commentary: Income inequality summarises the health burden of individual relative deprivation. *BMJ 314*, 1727.

Table 1: Self-assessed health, mean income and income inequality[a] by Scottish local authority, 1999-2000

	N[b]	Less than good SAGH[c] (%)	Mean income	Gini
Aberdeen City	1156	46.2	16,450	0.291
Aberdeenshire	1115	44.2	16,109	0.294
Angus	604	44.7	15,141	0.281
Argyll & Bute	552	44.7	14,880	0.286
Borders	590	45.9	14,631	0.279
Clackmannanshire	567	47.9	15,079	0.286
Dumfries & Galloway	865	52.4	13,870	0.262
Dundee City	747	48.9	13,759	0.268
East Ayrshire	658	49.4	13,924	0.272
East Dunbartonshire	564	44.0	16,888	0.274
East Lothian	566	44.7	16,483	0.300
East Renfrewshire	464	44.2	16,556	0.274
Edinburgh City	2327	39.7	16,708	0.302
Falkirk	798	48.9	15,008	0.279
Fife	1824	49.1	14,775	0.284
Glasgow	3001	48.7	13,764	0.289
Highland	1080	44.0	14,675	0.281
Inverclyde	518	48.4	14,048	0.268
Midlothian	590	49.0	15,421	0.272
Moray	604	47.4	13,980	0.300
North Ayrshire	796	49.9	13,807	0.273
North Lanarkshire	1577	51.6	14,048	0.261
Orkney	514	42.1	13,761	0.284
Perth and Kinross	741	42.8	16,555	0.298
Renfrewshire	917	43.6	13,941	0.278
Shetland	532	44.1	15,718	0.283
South Ayrshire	653	43.5	15,164	0.291
South Lanarkshire	1541	46.0	15,180	0.279
Stirling	579	45.7	16,052	0.300
West Dunbartonshire	517	48.0	13,654	0.275
West Lothian	789	49.6	16,202	0.274
Western Isles	636	34.2	13,231	0.288
Scotland	28,982	46.5	15,572[d]	0.287

Notes

[a] Income based on equivalised net household income, before housing costs, inflation adjusted (Dec 2000 prices) excluding the top and bottom 1% of the income distribution

[b] N refers to the numbers on whom income data were available. The prevalence of poor health is based on all those for whom data on self assessed health were available (n=28,379).

[c] Weighted to take account of household size and response rates

[d] Weighted to take account of under- and oversampling across local authorities

Table 2: Log odds (standard errors) and odds ratios (95% CI) of less than good self-assessed general health by socio-economic status and local authority mean income and income inequality

Variable	*Level*	*Model 1* *Log odds (SE)*	*Model 2* [a] *Log odds (SE)*	*Model 3* [a] *Log odds (SE)*	*Model 3* [a] *OR (95% CI)*
	Constant	-0.506 (0.030)	-1.692 (0.078)	-1.498 (0.145)	
Fixed Parameters					
Income	1 Poorest		0.215 (0.058)	0.219 (0.058)	1.25 (1.11-1.40)
	2		0.329 (0.055)	0.331 (0.055)	1.39 (1.25-1.55)
	3		0.200 (0.052)	0.201 (0.052)	1.22 (1.10-1.35)
	4		0.106 (0.051)	0.106 (0.051)	1.11 (1.01-1.23)
	5 Richest		*	*	*
Economic Status	Inactive		0.533 (0.041)	0.534 (0.041)	1.71 (1.57-1.85)
	Unemployed		0.533 (0.071)	0.535 (0.071)	1.71 (1.49-1.96)
	Econ Active		*	*	*
Education Level	None		0.429 (0.046)	0.430 (0.046)	1.54 (1.41-1.68)
	O grade or equiv		0.269 (0.050)	0.267 (0.050)	1.31 (1.18-1.44)
	Higher or equiv		0.219 (0.047)	0.219 (0.047)	1.25 (1.14-1.37)
	Degree		*	*	*
Income inequality	Highest inequality			-0.195 (0.134)	0.82 (0.63-1.07)
	High inequality			-0.199 (0.110)	0.82 (0.66-1.02)
	Low inequality			-0.116 (0.109)	0.89 (0.72-1.10)
#	Lowest inequality			*	*
Mean LA income	Lowest income			-0.203 (0.134)	0.82 (0.66-1.01)
	Low income			-0.057 (0.110)	0.95 (0.78-1.14)
	High income			0.014 (0.097)	1.01 (0.84-1.23)
#	Highest income			*	*
Random parameters					
Variance	Local authorities	0.016 (0.007)	0.021 (0.008)	0.020 (0.008)	
Components	Electoral wards	0.085 (0.015)	0.067 (0.014)	0.067 (0.015)	

a Models 2 & 3 adjust for age and sex
* Reference category