Has regional variation in mortality rates declined since 1931, and in all age groups, in Britain? A re-analysis using formal statistical modelling

Russell Ecob, Chris Robertson, Graham Watt

Abstract

Study objective—To examine changes in regional variance in all cause mortality rates in Great Britain from 1931-91 using formal statistical modelling procedures, and to follow up the suggestion by Illsley and Le Grand that there has been a reduction over time in the regional variance in younger but not older age groups.

Design, setting, and participants—Data were the age and sex specific death rates around each census from 1931-91 for the British regions, reconstructed to make them comparable with the 1981 regional definitions. Regional variance was modelled using bootstrap simulation tests and by age-period and age-cohort models.

Measurements and main results—While there was some evidence of a decline and levelling off of the regional variance over time in older age groups (over 35), the decline in younger age groups was more marked. This broadly confirms previous findings. Parametrising the period effect into linear and quadratic components, with allowance for an increase in regional variance in the war years, gave broadly comparable fit to the data as a model with period as a factor. Models for the changes in regional variance which were based on period effects seemed to provide a better description of the observed variances than those based on birth cohort effects. In the younger (but not older) groups there was evidence of a rise in the regional variance between 1981 and 1991.

Conclusions—The decline in regional variance is larger in younger than in older age groups when allowance is made for the increase in regional variance over the war years. Statistical modelling can provide insights into the data which are not always detected by descriptive analyses. Moreover, they provide a capacity for generalisation beyond the particular data; relationships found can form the basis for studies of replicability, for example, in other countries.

(*J Epidemiol Community Health* 1997;51:502-509)

The study of mortality differentials between regions over time has gained impetus in recent years partly as a result of a dissatisfaction with the use of social class as an invariant indicator over time of underlying socioeconomic factors. Direct methods of examining changes in the

relation of mortality to particular socioeconomic variables over time are bedevilled by two factors—the change in the measurement scale and the inconsistency of numerator and denominator populations. Though absolute levels of mortality have declined over time, there is substantial evidence that existing differentials have not decreased, and even that they have increased over time.¹⁻⁵ However, it has been argued that the changes in the occupational and social structure of Britain, involving, in particular, the reduction in the size and composition of the lowest social class (V), means that the measurement scale has been “stretched . . . artificially creating an increase in the mortality inequality”.^{6,7} Moreover, unless the denominators (population) and the numerators (deaths) making up the mortality rate are taken from the same source (as, for example, in the OPCS longitudinal study) it is also likely that biases will arise in the calculation of such differentials through differences between self report of occupation at census and the later, proxy report by next of kin at death.⁸

Though there has been much research into regional differences in mortality at particular points in time,⁹ little research has examined systematically the differences between regions on a national basis over time.^{10,11}

A substantial body of research now suggests that the differences in mortality between areas of high and low deprivation have increased over the recent period (1981-91).^{12,13} Though no evidence seems to have come so far from national UK data sources, such an increase has been found in Scotland,¹⁴ northern England,¹⁵ Yorkshire,¹⁶ London,¹⁷ and in Glasgow.¹⁸

Persistence of the regional differentials between nine regions in England and Wales and between these and Scotland was found from 1971-81, though the reduction in mortality was found to be larger in the non-manual than in the manual groups.¹⁹ Moreover, a comparison between 1961 and 1981 of Glasgow and Edinburgh, two cities with contrasting socioeconomic composition, showed differences in mortality which were increasing over time and predicted a further increase in 1991.²⁰

Illsley and Le Grand¹ were, to our knowledge, the first to examine this issue in Britain as a whole, differentiating by sex and age groups. They concluded that there was evidence of different patterns of regional inequality in mortality at different age groups, that this was due to a substantial convergence between regions in age specific death rates between younger but not in older age groups, and that the historic

MRC Medical
Sociology Unit.
Glasgow University, 6
Lilybank Gardens,
Glasgow G12 8RZ
R Ecob

Division of
Epidemiology and
Biostatistics,
European Institute of
Oncology, Via
Ripamonti, 435, 20141
Milan
C Robertson

Department of
General Practice,
Glasgow University,
Glasgow, G20 7LR
G Watt

Correspondence to:
Mr R Ecob

Accepted for publication
January 1997

north/south gradient had disappeared in younger age groups though it persisted in older age groups. These conclusions were based on a graphical analysis of the variances in log mortality rates over the 10 regions, separately for males and females.

This paper aims to subject to further analysis the conclusions of Illsley and Le Grand¹ regarding the changes in the inter-regional variances within age groups over time. We therefore asked whether reductions in regional variance over time were found in younger but not in older cohorts. Illsley and Le Grand's analysis considered only period effects, but we considered whether the different patterns of variability in the regional log mortality rates in different age groups were more consistent with changes in variance associated with birth cohort.

Methods

Population and mortality data were available from 10 regions of the United Kingdom around the censuses of 1931-91, obtained from annual reports of the Registrar General and of the OPCS. These comprised a single year's deaths at each occasion. The regions were South West, South East, East Anglia, East Midlands, West Midlands, Yorkshire and Humberside, North, North West, Wales, and Scotland. Both deaths and population data were grouped for publication into age groups, from which the age groups used here (< 1, 1-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+) were derived.

The data are those used, and obtained, by Illsley and Le Grand,¹ and with their reconstruction of regions before 1981 to render them comparable with 1981 definitions, but with the following modifications. Firstly, we used the 1991 census figures and 1991 mortality figures, whereas Illsley and Le Grand used death rates based on an average of the years 1987-89 (the latest available at the time). No changes in regional definition occurred between 1981 and 1991.

Secondly, the two youngest age groups, those less than 1 and those 1 to 4 years old, were omitted to facilitate the formation of synthetic, 10 year birth cohorts. These were assumed to be centred on the following years, 1851, 1861, . . . 1981. The remaining data thus comprised seven time periods, eight age groups, and 14 birth cohorts.

The first step in analysis was to test if there was evidence that the regional variance in the log mortality rates had changed over the seven time periods. This was accomplished by the use of "bootstrap simulation" tests.²² The second step was to model the between region variances using the following generalised linear model.²³

$$s^2_{kij} = \mu + S_k + A_i + P_j, \quad 2.1$$

where s^2_{kij} was the between regional variance in log mortality rates in sex k age group i and time period j . The parameters, represented by S_k , A_i , and P_j , were the sex, age, and period effects, respectively.

KEY POINTS

- The decline in variance in mortality rates in UK regions between 1931 and 1981 is greater in younger (5-34 years) than older (35+ years) age groups.
- In younger but not older age groups the decline in regional variance reduces over time, the regional variance increasing between 1981 and 1991.
- For these data models with period as a factor produces a more parsimonious explanation of temporal changes in regional variance than models with cohort as a factor.
- Statistical modelling provides a concise and straightforward summary of the data and a basis for comparison of trends over time in other countries.
- The increase in regional variance in younger age groups between 1981 and 1991, reversing the trend up to that time, is a source of concern and should be monitored using data at the 2001 census.

One of the main hypotheses was that the changes in variance associated with young and "old" age groups over time period were not the same. A new variable was created where the first level represented the younger age groups 5-34 years, while the second level represented the older ages 35+ years. This factor was denoted "old" age and was wholly subsumed within the main "age factor". The particular age range of this variable was decided on by inspection of the parameter estimates forming the age-period interaction (model 3, table 3). Little difference was found in practice between this age division and one based on 45 years. This was chosen primarily to reduce the heterogeneity within the younger age group, this now comprised children and young adults. This hypothesis required the extension of equation 2.1 to consider interactions between the old factor (denoted O , below) and period. This was the baseline model for the analysis and was:

$$s^2_{kij} = \mu + S_k + A_i + P_j + O.P_j, \quad 2.2$$

The model was an additive model for the variance and could yield predictions of negative variances. A Box-Cox transformation²⁴ was used to test the assumption that the additive model was valid. Gamma errors were assumed for this model as the variances will have χ^2 distributions, assuming the log mortality rates in each age group and time period follow a normal distribution with constant variance.²³ All tests of specific hypotheses used likelihood ratio tests based on the differences of scaled deviances.

Age-cohort models were also fitted to the data. These were of the same form as equations 2.1, 2.2 but with period replaced by cohort.

Results

The log mortality rates for males and females for the United Kingdom as a whole are presented in figure 1. The figures are plotted to show individuals in the same synthetic cohort

Figure 1 Log mortality rates in relation to birth cohort for males and females.

in a vertical line, values for a given age (for different cohorts) being linked together. These serve to provide a background for the discussion. They show, overall, decreasing mortality with time within each age group, with the rates in younger people decreasing fastest. In comparison with the period before the war, however, there are increases in mortality rates

over time for males at ages 15-34 and a halt in the reduction in the mortality for females aged 5-34 and males aged 5-14 over the war period. The cohort representation illustrates the increase in mortality in the war period (looking vertically upwards in fig 1) compared with the subsequent age for cohorts of both sexes born around 1910 and 1920. The variances in the log rates are presented in table 1. On the basis of an inspection of these variances, though using data up to 1989 only, Illsley and Le Grand¹ came to their conclusion that there was evidence of a decreasing variance in younger age groups but that the variability in middle and older age groups was wide and increasing in later periods.

Assuming that log rates follow an approximately normal distribution, a likelihood ratio test can be constructed to test the equality of the variances over the seven periods.²⁵ The test statistics are presented in table 2. If the samples were large enough, a χ^2 distribution on 6 degrees of freedom (df) would be expected to follow. In this case the samples are not large enough as there are only 10 observations used in the calculation of each variance.

The p values presented in the table were obtained by simulation. These were based on a normal assumption for the log rates and a Poisson assumption for the numbers of deaths in each region. The log rates were simulated under the assumption of equality of variance over the seven time periods and the log likelihood ratio statistics were calculated 999 times to give the p values in table 2. These suggested that in the two youngest age groups for both sexes, the assumption of equality of variance over the periods was not valid, nor was it for the oldest group for females. This conclusion is affected to some extent by the war in 1941. The p values calculated by omitting 1941 are therefore also given in table 2. There is now no evidence of changes in the variance for the 15-24 age group in males.

The above simulations were based on a single age group within each sex and so did not utilise the information in adjacent age groups

Table 1 Regional population variances in log (mortality rates)

Age group (y)	Time period						
	1931	1941	1951	1961	1971	1981	1991
Males							
5-14	0.0471	0.0366	0.0075	0.0093	0.0034	0.0052	0.0093
15-24	0.0258	0.0759	0.0091	0.0140	0.0049	0.0067	0.0148
25-34	0.0272	0.0580	0.0275	0.0175	0.0122	0.0126	0.0089
35-44	0.0254	0.0360	0.0249	0.0273	0.0193	0.0252	0.0154
45-54	0.0224	0.0300	0.0212	0.0180	0.0194	0.0232	0.0262
55-64	0.0145	0.0218	0.0141	0.0141	0.0140	0.0164	0.0167
65-74	0.0180	0.0234	0.0079	0.0122	0.0078	0.0089	0.0122
75+	0.0094	0.0085	0.0043	0.0044	0.0020	0.0026	0.0021
Females							
5-14	0.0403	0.0480	0.0240	0.0146	0.0042	0.0038	0.0182
15-24	0.0384	0.0639	0.0431	0.0033	0.0154	0.0091	0.0137
25-34	0.0251	0.0406	0.0367	0.0169	0.0061	0.0128	0.0299
35-44	0.0227	0.0255	0.0173	0.0150	0.0127	0.0132	0.0105
45-54	0.0211	0.0186	0.0089	0.0090	0.0112	0.0177	0.0122
55-64	0.0207	0.0175	0.0125	0.0142	0.0115	0.0197	0.0190
65-74	0.0242	0.0271	0.0142	0.0140	0.0073	0.0108	0.0148
75+	0.0121	0.0080	0.0070	0.0076	0.0012	0.0011	0.0021

Table 2 Log likelihood test statistics and Monte Carlo p values

Gender	Age group (y)	All 7 periods		Omitting 1941	
		Log likelihood ratio statistic	Monte Carlo p value	Log likelihood ratio statistic	Monte Carlo p value
Males					
	5-14	9.99	0.001	24.76	0.007
	15-24	28.94	0.001	8.91	0.221
	25-34	13.11	0.065	5.22	0.489
	35-44	2.11	0.947	1.14	0.962
	45-54	0.93	0.995	0.44	0.996
	55-64	0.82	0.997	0.16	0.999
	65-74	5.63	0.565	2.82	0.775
	75+	12.07	0.113	9.42	0.152
Females					
	5-14	24.98	0.001	19.50	0.033
	15-24	26.71	0.001	19.59	0.011
	25-34	11.41	0.138	9.32	0.183
	35-44	3.16	0.840	1.87	0.909
	45-54	3.79	0.765	3.27	0.716
	55-64	1.61	0.964	1.58	0.932
	65-74	5.95	0.530	3.87	0.644
	75+	24.89	0.003	23.99	0.003

Table 3 Deviances and degrees of freedom (df) for the Gamma error generalised linear model for the variances

Model	Model terms	Scaled deviance	df
1	Sex + age	316.93	103
2	Sex + age + period	159.49	97
3	Sex + age*period	60.50	55
4	Sex + age + period + old.period	91.00	91
5	Sex + age + period + old.period + sex.period	84.40	85
6	Sex + age + period + old.period + sex.period + sex.old	79.21	84
7	Sex + age + period + sex*old*period	69.39	78
8	Sex + age + period + old.period + sex.old	86.65	90
9	Sex + age + P_L	234.45	102
10	Sex + age + $P_L + P_Q$	182.87	101
11	Sex + age + ($P_L + P_Q$)*old	117.97	99
12	Sex + age + ($P_L + P_Q$ + war)*old	107.64	97
13	age + (Sex + $P_L + P_Q$ + war)*old	101.99	96
14	Sex + age + cohort	148.10	90
15	Sex + age + cohort + old.cohort	119.78	85
16	Sex + age + C_L	234.45	102
17	Sex + age + $C_L + C_Q$	224.24	101
18	Sex + age + ($C_L + C_Q$)*old	156.34	99
19	Sex + age + ($C_L + C_Q$ + war)*old	131.87	97
20	age + (sex + $C_L + C_Q$ + war)*old	121.99	96

Table 4 Parameter estimates of the models for the variances

	Full model (model 8)		Linear quadratic model (model 13)	
	Estimate $\times 10^{-2}$	Standard error $\times 10^{-2}$	Estimate $\times 10^{-2}$	Standard error $\times 10^{-2}$
Intercept	3.46	0.58	0.79	0.12
Sex (females)	0.25	0.14	0.34	0.15
15-24	0.35	0.15	0.35	0.16
25-34	0.72	0.20	0.78	0.21
35-44	-0.63	0.65	1.31	0.25
45-54	-0.77	0.65	1.15	0.23
55-64	-0.97	0.64	0.96	0.22
65-74	-1.34	0.63	0.60	0.19
75+	-2.30	0.62	-0.37	0.14
1941	2.05	1.05		
1951	-1.42	0.68		
1961	-2.48	0.61		
1971	-3.10	0.58		
1981	-3.00	0.59		
1991	-2.22	0.62		
Sex by old age	-0.33	0.15	-0.39	0.16
Old age by 1941	-2.01	1.09		
Old age by 1951	0.85	0.72		
Old age by 1961	1.95	0.66		
Old age by 1971	2.14	0.62		
Old age by 1981	2.08	0.63		
Old age by 1991	1.36	0.67		
Period linear			-0.58	0.09
Period quadratic			0.22	0.04
Old age by period linear			0.41	0.10
Old age by period quadratic			-0.18	0.04
War			2.65	0.94
Old age by war			-2.38	0.97

which might have been expected to behave with a similar pattern. A generalised linear model was used to model the variances in table 1. Although the inverse link was the natural link for the gamma distribution,²⁶ it was considered desirable to use the identity link in this analysis as the trends in the variances were investigated on the identity scale. Furthermore, investigating the appropriate link function through a Box-Cox transformation²⁴ yielded the power transformation exponent as 1.4 with an approximate 95% confidence interval of 0.9, 2.1. Thus, the identity link and not the inverse link seemed to be the appropriate one, the latter exponent (-1) being outside the above confidence interval.

The baseline model required to describe these data is equation 2.2 and has a deviance of 11.824 on 91 df, leading to an estimate of the scale parameter of 0.1299. The scaled deviance

in table 3 (model 4, equation 2.2) were obtained using this estimate.²⁶ This is the reason for the equality between the scaled deviance and df for this model. To check on the consistency of the findings irrespective of the choice of the baseline model for the estimation of the scale parameter, the models were re-run using, as the baseline model, the previous model 3. The scale parameter (0.14) for this model was almost identical. No change in the model selection process resulted and neither were there any changes in the parameter estimates or standard errors in the final model chosen.

Investigation of the differences in the scaled deviances, which had an asymptotic χ^2 distribution assuming the baseline model (table 3, model 4) was a reasonable fit to the data, suggested that there was evidence to support the conclusion that there were different variances in the different time periods ($\chi^2 = 157.44$, 6 df: models 1 and 2, equation 2.1, of table 3). Also the differences in variance over time period were not independent of the age groups ($\chi^2 = 98.99$, 42 df: models 2 and 3 of table 3). The "old" by "period" interaction on only 6 df accounted for most of the differential effect as the residual χ^2 was 30.50 on 36 df (models 3 and 4). Thus, it was reasonable to conclude that the sex + age + period + old.period model was the most appropriate for these data (note that this is the same model as is presented in equation 2.2).

Plots of the Pearson residuals against age group, time period, and the fitted values are presented in figure 2, along with the normal order plot of the residuals. None of these plots suggested that the model was invalid. There was a slightly wider spread of residuals for the 15-24 age group and some suggestion of outliers in time periods 1951-71, suggesting that the dip in 1961 in this age group in comparison with 1951 and 1971 (see Illsley and Le Grand¹, fig 2) was not taken into account, and a slight non-linearity in the normal plot of the residuals, but nothing to suggest any serious model discrepancies.

There was no evidence of any difference in the effect of period between the two sexes, model 4 and 5 ($\chi^2 = 6.60$, 6 df). This test was of interest because of the possible effect of the war on the regional variance in mortality rates. The high mortality rates in 1941 mainly affected the younger age groups and so we probed the effects of the war further by looking at the interaction between sex and "old" age and the three way interaction between sex, period, and "old" age. A slight improvement to this model could be made by considering an interaction between sex and the "old" age factor ($\chi^2 = 5.19$, 1 df model 5 and 6), but there was no evidence of any three way interaction (models 6 and 7).

The parameter estimates for model 8, the baseline model plus the sex by 'old' age interaction (as a comparison with model 6 confirmed no sex by period interaction), are presented in table 4. These are compared with a reference group comprising males, aged 5-14, in 1931. Fitted values for males are shown in

Figure 2 Plots of Pearson residual for model 4 in relation to time period, age group, fitted value, and normal order plot for residuals.

figure 3. Among the younger ages there was more regional variance in females than males. Among the youngest three age groups there was greatest regional variance in 1941, but the general trend thereafter was to decrease to a trough in 1971, followed by a slight rise. Among the older age groups there was less regional variation in log mortality rates in 1941, with the interaction effect cancelling out the high value among younger age groups. Taking the interaction estimates into account there was evidence of only a slight decline in regional variance among the older age groups (this ranges from a maximum of 0.0004 in 1941 to a minimum of -0.0098 in 1981 relative to 1931, compared with a range of 0.0205 in 1941 to -0.0310 in 1971 among the younger age groups). The fitted values for females were of the same form as for males but with the values at ages up to 35 slightly larger.

Of particular interest was the estimation of the change in regional variance from 1981-91, as a number of authors have suggested that increases have occurred in small areas within regions, differentiated according to deprivation. In the younger age groups, the variance in 1991 was greater than the variance in 1981 by

0.0079 (SE 0.0027), while among the older ages the difference was 6.39×10^{-4} (SE 7.82×10^{-4}). These estimates suggested that there had been an increase in the regional variance among the three youngest age groups between the last two censuses, though none had occurred in the older age groups.

Period can alternatively be considered in its linear and quadratic components, centred on 1961 only. Looking at the scaled deviances corresponding to these components it can be seen that they accounted for the greater part of the time period effect (models 1, 9, 10, table 3). The deviance associated with the linear effect was 84.48 on 1 df, that associated with the quadratic effect was a further 51.58 on 1 df while the interaction of these two terms with the older age groups accounted for 64.90 on 2 df (models 10 and 11).

There was still a significant excess compared with model 4 of 26.97 on 8 df. Inspection of the residuals (not presented) showed that the lack of fit was due to an effect of the war in 1941, and the quadratic time model did not take into account the excess mortality among the younger age groups compared with 1931 especially. Adding in a dummy variable to

Figure 3 Predicted value plots for the time period factor model (model 8); males only.

account for the excess mortality among the younger age groups in 1941 left a residual scaled deviance of 107.64 on 97 df (model 12). There was no significant improvement in fit in allowing the excess mortality in younger age groups in 1941 to vary according to sex. There

Male data quadratic time model

Figure 4 Predicted values for the linear and quadratic time model (model 13); males only.

was still some lack of fit compared with models 4 and 8 and this was due mainly to poor prediction of the variations in the rates for females aged 15-24 in 1951, 1961, and 1971. This was the sex by "old" age interaction previously noted (model 13)

The parameter estimates for model 13 are presented in table 3 and the fitted values for males are shown in figure 4. Fitted values for females showed the same general pattern. These confirmed the previous evidence for an increase in regional variability over the last time period in the youngest three groups. Moreover, within the older age groups the linear trend with period was significant (-0.0017 , SE, 0.00033), though at a lower value than the linear trend in the younger groups (-0.0058), as was the quadratic term (0.00038 ; SE, 0.00012).

Within this simplified period effect model there was no evidence of any differential regional variability between the sexes associated with the war ($\chi^2 = 0.47$ on 1 df) nor any differential regional variability between the sexes in young ages only associated with the war ($\chi^2 = 0.12$ on 1 df). There was also no evidence that the relation to period was not the same for males and females ($\chi^2 = 3.13$ on 2 df). Birth cohort is an alternative representation of temporal changes and the remaining six models (numbers 14 to 20) in table 3 used cohort in place of period. Note that the model (number 16) with age and cohort as a linear term is logically equivalent to that (number 9) with age and period as a linear term due to the linear dependence between these factors.²⁷ However, none of the other models with cohort are equivalent to or fit as well as the corresponding ones with period. Thus, our preference is for period as the factor to use in a parsimonious explanation of the temporal changes in the regional variances.

Comparison of the variances in table 1 with those in table 1 of Illsley and Le Grand¹ reveals some differences in the last period. In our case the regional variances tended to be larger in those age groups under 45. This can be explained by Illsley and Le Grand's use of the average of three years' data in the last period. They did so because there were few deaths in some of the regions in some of the age groups, but the effect of using the average was to reduce the expected value of the regional variance by comparison with all the other periods in which only 1 year's deaths were used. We used the single year (1991) to maintain consistency with the previous years.

Discussion

This paper has sought to model and describe changes in the regional variance in mortality rates within the United Kingdom from 1931-91. Of particular interest are changes associated with period. However, successive cohorts have certainly enjoyed increased expected lifespan and it is of interest also to see if regional variation has changed with cohort.

There is evidence to suggest that regional variance was not constant over the period of 60 years investigated here. Initially, it was greatest

among the younger age groups and it declined substantially over time among these age groups. Among older age groups the regional variance seemed to be lower at earlier time points (1950 and earlier), though not at later time points, than in the younger groups. However, there has still been a significant reduction in the older age groups over time up to 1971, though there have been no further reduction since then.

In the above respects the statistical modelling broadly confirmed the descriptive results of Illsley and Le Grand, though we found no evidence of an increase in variance between regions in older ages at later periods.¹ However, in addition, in younger age groups there was evidence of a recent increase in the last period (now 1981-91) examined in the variance in the regional log mortality rates. This continued a trend, operating in all years apart from the war ones, towards a reduction over time in the rate at which the between regional variance decreased. This was not apparent in the Illsley and Le Grand analysis, partly because of their unusual practice of using an average of three years' deaths in the last period in contrast to the one year in all previous periods. This practice was justified by them (p 440) as follows "... because by then the number of deaths in a single year had become small in the youngest age groups". This had the effect of reducing the observed variance between regions by substantially reducing the error variance at this level.

Birth cohort does provide an alternative explanation of patterns in regional variance but on its own does not explain as much of the residual variation in the models of section 3 as time period. This is not a valid reason for preferring age-period to age-cohort representation and concluding that the changes in the variance are a result of period effects. However, the analysis indicated that the variance had reduced in all of the younger age groups at the same time (ie 1941-71) and that the changes in the older age groups over the same period were not as great. These different patterns in two distinct age groups are not in accord with birth cohort effects.

The variance of the log rates was larger in the younger age groups which had lower mortality rates, and smaller in older age groups.²⁸ Thus, part of the estimated relationship between the variance of the log rates and age and period may be due to the different variances in the log rates which have not been taken into account in the models. If this were the only reason for the changing variance, a monotonic relationship would be expected and not the quadratic relationship found. Also, the variance would be higher in the younger ages over all the periods as the rates themselves are more variable when there are fewer deaths. At the time when there was greatest regional variation among the younger age groups (1931-51) the mortality rates were higher (fig 1). Taken together these points mean that under specification of the model is not an explanation for the recent increase in the regional variance in mortality rates among younger age groups.

This analysis has modelled the variance in the log mortality rates over regions as a function of age group, time period, and sex. We need to address the question of how much of this variance was due to random sampling variation and how much to systematic differences among the regions. It is possible to show that the expected values of the regional variance in the log mortality rates is equal to the sum of a component due to the systematic differences between the regions and a component which is due to the random variation in the log mortality rates. As the variance of the log mortality rate is estimated as the inverse of the number of deaths,²⁸ this second component is estimated as the average of the inverses of the number of deaths over the regions. Consequently, it will be larger in the age groups and time periods where there are fewer deaths, namely the younger age groups. There was no relationship between the regional variance in the log mortality rates and the average of the inverses of the number of deaths over the regions, the correlation being -0.12. Moreover, there was no pattern to a plot of residuals from the final model (model 8 in table 3) against the average of the inverses of the number of deaths. Furthermore, the inclusion of the average of the inverses of the number of deaths as an explanatory variable in model 8, table 2 did not result in a significant reduction in deviance, even when stratified by age group. Consequently, we believe that the effect of increased random variation due to increasingly small numbers of deaths in the younger age groups does not fully explain the recent increase in regional variance.

The causes of mortality differed between the age groups, with the youngest age group (5-14) who were still subject to childhood causes showing marked differences to the other age groups. Also, the oldest age group had an open ended interval and we assumed that it could be associated with a 10 year birth cohort. Similar results were obtained in the models when the youngest and the oldest age groups were omitted from the analysis. Thus, the conclusions from the quadratic model for the variances did not depend upon the inclusion of the two extreme age groups, for either sex.

The log mortality rates decreased with time and the decrease was greatest in regions which had high rates among younger age groups in the initial period of study. The convergence in the log mortality rates over the regions occurred because the rates in regions where they were highest decreased at a faster rates than those in regions where they were lower.

The explanation for the apparent increase in regional variance recently among younger people requires elaboration. It is not a consequence of a very steep decline from 1941-51 and a much shallower decline from 1951-61, which leads to an upturn in the future. Firstly, the quadratic model on data from 1950 onwards still shows the recent upturn in the predicted regional variance among younger ages in the last two periods. Secondly, this upturn is present in the fitted values in figure 3 which are based on a model which makes no

assumption about a quadratic model. Thus this effect is not a consequence of very fast decreases in regional variance in the early part of the series.

The increase in the between regional variance among the younger age groups between 1981 and 1991 may be related to the finding of an increase over a similar period in the relation between mortality and the deprivation of areas, shown for example in northern England,¹⁵ Scotland,¹⁴ and in Glasgow.²⁹

The model predictions for the variances (in fig 4) are based on a quadratic model. This is only a description of the observed trends and should not be taken as an indication of future trends as they would predict a continued increase in the variance in the rates, without any limit. The quadratic components are in the model to reflect different changes in the variances of the rates with time among different age groups—ie a faster decline and recent increase among younger age groups and a slower decline among older age groups.

We should note, however, that a quadratic relationship between regional variance and period could arise as a result of the reduction in the systematic component of the between regional variance to very low levels in recent periods with the increasing random variation due to the smaller numbers of deaths leading to an upturn. The models used in this analysis were not able to detect this, and the residual analysis did not suggest that this was the case but we have addressed the issue (Robertson, Ecob and Watt, personal communication).

This analysis has used the variance in the log mortality rates as a summary statistic at each time point for each age group and sex. An alternative analysis is to model the log mortality rates and the variance in these rates simultaneously through a multilevel model (Robertson, Ecob and Watt, personal communication).

Conclusions

While there was some suggestion of a decline and levelling off of the regional variance over time in the older age groups (35+), there was evidence that the decline in younger age groups (5-34) was more marked. This provides broad confirmation of Illsley and Le Grand's findings.¹ Parametrising the period effect into linear and quadratic components, with allowance for an increase in regional variance in the war years, gave broadly comparable fit to the data as a model with period as a factor. This model is a relatively concise description of the data and can form the basis for comparative research in other countries. The evidence, found here, of a rise in regional variance between 1981 and 1991 in the younger (but not the older) groups; a continuation of a trend operating over a longer period, was not detected by Illsley and Le Grand.¹ Though it is possible that this is due to the earlier final time point (1989) of their data it is more likely that this is because of their practice of combining three years of mortality data at this time point while only using data for one year at each of the

previous census points. This has the effect of reducing the observed variance between regions by eliminating a greater portion of the (random) variation between regions over adjacent years at the last time point. It is necessary in studies such as these to have comparable data at each time point.

These data (except the year 1991) have kindly been provided by Raymond Illsley whom we would like to thank for his cooperation in the early stages of this project. Valuable comments on earlier drafts were provided by Steven Cummins, George Davey Smith, Geoff Der, Anne Ellaway, and Sally Macintyre and by three anonymous referees.

Funding: Russell Ecob is employed by the Medical Research Council.
Conflicts of interest: none.

- 1 Illsley R and Le Grand J. Regional inequalities in mortality. *J Epidemiol Community Health* 1993; 47: 444-49.
- 2 Townsend P and Davidson N. *Inequalities in health (The Black report)*. Harmondsworth, Middlesex: Penguin, 1982.
- 3 Wilkinson RG. Socio-economic differentials in mortality: interpreting the data on their size and trends. In: RG Wilkinson ed. *Class and health; research and longitudinal data*. London: Tavistock, 1986.
- 4 Wilkinson RG. Class mortality differentials, income distribution, and trends in poverty 1921-1981. *Journal of Social Policy* 1989; 18: 307-35.
- 5 Harding S. Social class differences in mortality of men: recent evidence from the OPCS longitudinal study. *Popul Trends* 1995; 80: 31-37.
- 6 Illsley R. Occupational class, Selection and the production of inequalities. *Quarterly Journal of Social Affairs* 1986; 2(2): 151-65.
- 7 Anson J. Regional mortality differences in Britain, 1931-87; a two dimensional analysis. *J Biosocial Science* 1993; 25: 383-95.
- 8 Fox AJ, Goldblatt PO. *Longitudinal study; socio-demographic mortality differentials*. Series LS No 1. London: HMSO, 1982.
- 9 Gardner M, Donnan S. Life expectancy: variations among regional health authorities. *Popul Trends* 1977; 10: 10-12.
- 10 Hollingsworth JR. Inequality in levels of health in England and Wales, 1891-1971. *J Health Soc Behaviour* 1981; 22: 268-83.
- 11 Drever F, Whitehead M. Mortality in regions and local authority districts in the 1990s: exploring the relationship with deprivation. *Popul Trends* 1995; 82: 19-26.
- 12 Davey Smith G, Morris J. Increasing inequalities in the health of the nation; government action at last? *BMJ* 1994; 309: 1453-54.
- 13 Davey Smith G, Eggar M. Socio-economic differentials in wealth and health; widening inequalities in health - the legacy of the Thatcher years. *BMJ* 1993; 307: 1453-54.
- 14 McLoone P, Boddy FA. Deprivation and mortality in Scotland, 1981 and 1991. *BMJ* 1994; 309: 1465-69.
- 15 Phillimore P, Beattie A, Townsend P. Widening inequality of health in northern England, 1981-91. *BMJ* 1994; 308: 1125-28.
- 16 Staines A, Cartwright R. Deprivation and mortality in Yorkshire: changes from 1981-91. *J Epidemiol Community Health* 1995; 49 (Suppl) S79.
- 17 Congdon P. Deprivation in London wards: mortality and unemployment trends in the 1980's. *The Statistician* 1988; 37: 451-72.
- 18 McCarron PG, Davey Smith G, Womersley JJ. Deprivation and mortality in Glasgow: changes from 1980 to 1992. *BMJ* 1994; 309: 1481-82.
- 19 Marmot MG, McDowall ME. Mortality decline and widening social inequalities. *Lancet* 1986; 2: 274-76.
- 20 Watt GCM, Ecob R. Mortality in Glasgow and Edinburgh: a paradigm of inequality in health. *J Epidemiol Community Health* 1982; 46: 498-506.
- 21 Illsley R, Le Grand J, Mullins C. *Regional inequalities in health. Welfare state*. Programme discussion paper no 57. London: London School of Economics: STICERD, 1990.
- 22 Efron B, Tibshirani RA. *Bootstrap methods*. London: Chapman and Hall 1994.
- 23 McCullagh P, Nelder J. *Generalised linear models*. 2nd ed. London: Chapman and Hall, 1989.
- 24 Box GEP, Cox DR. An analysis of transformations (with discussion). *Journal of the Royal Statistical Society (B)* 1964; 26: 211-52.
- 25 Mood AM, Graybill FA, Boes D. *Introduction to the theory of statistics*. 3rd ed. Tokyo: McGraw Hill, 1974; 439.
- 26 Francis BF, Green M, Payne RD, ed. *GLIM 4 The statistical system for generalized linear interactive modelling*. Oxford: Clarendon Press, 1994.
- 27 Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics* 1983; 39: 311-24.
- 28 Esteve J, Benhamou E, Raymond L. *Statistical Methods in Cancer Research Volume IV: Descriptive Epidemiology*. Lyon: IARC Publications, 1994; 53.
- 29 McCarron PG, Davey Smith G, Womersley JJ. Deprivation and mortality in Glasgow: changes from 1980 to 1992. *BMJ* 1994; 309: 1481-82.