The value of linked data for research into the social determinants of health: An international perspective

Michael Goldacre

Professor of Public Health, Unit of Health-Care Epidemiology, Department of Public Health, University of Oxford, Institute of Health Sciences, Old Road, Oxford, OX3 7LF, email michael.goldacre@dphpc.ox.ac.uk

In this paper I cover some examples of insights that the use of linked use of data have given into health, and particularly mortality, in relation to occupational social class, other social measures of advantage and disadvantage, unemployment, income inequality, ethnicity, and growing old.

Occupationally defined social class

The Black Report[1,2] on health inequalities in England defined social class as "segments of the population, sharing broadly similar types and levels of resources; with broadly similar styles of living; and some shared perception of their common condition". Table 1 illustrates the classification of English occupational social class.

I	Professional, e.g. doctor, lawyer, clergyman
II	Intermediate, e.g. manager, nurse, schoolteacher
III nm	Skilled non-manual, e.g. accounts clerk, typist, secretary
III m	Skilled manual, e.g. tailor, carpenter, toolmaker
IV	Partly skilled, e.g. painter, roofer, postman
V	Unskilled, e.g. cleaner, docker, roadsweeper

Table 1 Occupational social classes I to V

In England, as in many other developed countries, our information about occupation, social class and their associations with mortality generally come from analysis of death certificates (which provide the data for the numerators) around the time of the decennial census (which provides the data for the population denominators). These are cross-sectional data, around a single point in time which, although useful, have some limitations. One is the difficulty of addressing the potential biases of health-related changes of occupation and consequential social class 'drift'. This is important because people who become sick may change jobs and, as a consequence, move down the social class scale. As I shall describe, findings from cross-sectional data have been complemented in valuable ways by analyses from a national linked Longitudinal Study of census data and mortality data in England.[3,4]

The Longitudinal Study (LS) comprises a sample of all people born on one of four days in each year, as recorded in the census. The census data on the people in the sample are then linked prospectively to their subsequent records (if any) of cancers, emigration, the next census, and mortality. The linked LS has a number of advantages over the cross-sectional decennial analyses as a source of information about social factors and health. First, it can be used to analyse the time relationships between social circumstances, such as occupation and employment, and subsequent ill-health. Second, the census in England only takes place once every ten years; and, a few years on from the census, there is a natural tendency to question whether findings from its time still apply. Through continuous linkage, the LS can be used to provide regular updates in inter-censal years. Third, death certificate data, alone, on occupation, and hence on social class, can be poor for women and retired people. The census data tend to be more reliable for these groups and their inclusion by linkage to death data enriches the study of occupation and mortality. Finally, the census includes social data that are not recorded on death certificates, such as employment status and housing tenure. These can be analysed as factors associated with mortality in the LS.

Analysis of data about men by their social class, with social class for each individual recorded in the 1971 census, shows strong class gradients in mortality analysed for the years 1976–80, 1981–5 and 1986–92 (Table 2).[4]

Social class	1976-80 1981-85 1986-92
I, II	621 539 455
III, nm	860 658 484
III, m	802 691 624
IV, V	951 824 764
Ratio IV, V to I, II	1.53 1.53 1.68

Table 2 Age-standardised mortality rates per 100,000 men aged 35–64 years, by social class recorded in 1971, for deaths in 1976–1992

Death rates in all classes have fallen across the time periods; but rates in higher classes have fallen more than those in lower classes. The social class gradient widened over time. The data shown in Table 2 exclude deaths in the first five years after the census. If life-threatening illness caused much downward migration through the social classes, one would expect its effect on mortality rates to be most prominent in the early years after any change of job (and class); and to wear off with time, as those affected die. The data in Table 2 provide convincing evidence that the effect of social class on mortality is not just a short-term effect that might be explained by recent illness-associated downward migration. A similar, but slightly shallower, gradient is seen for women (Table 3).

Social class	1976-80 1981-85 1986-92
I, II	
	338 344 270
III, nm	
	371 387 305
III, m	
	467 396 356
IV, V	
	508 445 418
Ratio IV, V to I, II	
	1.50 1.29 1.55

Table 3 Age-standardised mortality rates per 100,000 women aged 25–64 years, by social class recorded in 1971, for deaths in 1976–1992

Considering common individual causes of death, the class gradient is strong, for example, for ischaemic heart disease and lung cancer; but it is also found for most other major conditions. In summary, the LS shows that mortality rates have fallen in all social classes; that those in higher social classes have fallen the most; and that the social class gradient has widened. Selection of ill people into lower social classes is unlikely to have caused much of the gradient.

Specific occupational exposure and general social advantage/ disadvantage

Is much of the social class effect a consequence of specific occupationally-related factors? The answer is generally no; but I will comment briefly on a few interesting examples of exceptions. Doctors now have very low standardised mortality ratios (SMRs) for some conditions,[5] like lung cancer and coronary heart disease, because of a past occupational tendency to act on their own advice and stop smoking. SMRs for lung cancer in doctors in England are now about one-quarter of the level in the general population. SMRs for lung cancer in publicans are about twice, and SMRs for cirrhosis of the liver are about ten times, the level in the general population.[5] If we had no idea what caused liver cirrhosis, we could guess that it might be something in the pub!

A colleague in my research unit, Stephen Roberts, has recently studied mortality from accidents at work by occupation.[6] He has a particular interest in seafarers. People who are resident in Great Britain, but who die at sea, are not registered as deaths through the same registration processes as deaths on land. They are registered through a maritime register and do not get into the conventional occupational mortality statistics. Stephen Roberts did his study by linking occupational records with maritime registration records. He showed that by far the most dangerous occupations in Great Britain are fishing (50 times higher mortality rates for work-related deaths than the general working population) and other seafaring (Table 4),[6] but that is not evident from routine occupational mortality statistics. So, there are some specific occupational issues.

Occupation	Relative Risk 95% CI
Trawler fisherman	
	52.6 43.5 – 63.8
Merchant seafarers	
	29.1 25.1 – 33.7
Scaffolders	
	12.9 9.8 – 17.0
Coal Miners	
	4.8 3.5 – 6.5

Table 4 Mortality rates from fatal accidents at work, expressed as relative risks (RR) compared with all workers (RR=1) in Great Britain

For the most part, however, occupational social class gradients reflect more general social advantage and disadvantage. Findings from the LS make this clear beyond reasonable doubt. Table 5 shows some of the evidence. The census includes data on factors like car ownership and housing tenure. As linkage to mortality data shows, these are surprisingly good predictors of mortality gradients.[4] These proxies for social advantage and disadvantage corroborate the evidence from the occupational social class data that, in respect of premature mortality, we are far from being a society of equals.

	Men Women
Car	
Owner occupier	72 70
Private tenant	83 82
Public council tenant	96 93
No Car	
Owner occupier	99 91
Private tenant	129 105
Public council tenant	120 125

Table 5 Standardised mortality ratios for men and women aged 45–64 years (general population, each sex = 100), classified by car ownership and housing tenure in 1971 (for deaths that occurred in 1971-92)

We have been undertaking work in the South East England Public Health Observatory using data about so-called Super Profile Lifestyle Groups.[7] These are classifications of postcoded addresses, households, and neighbourhoods. Data about small-area neighbourhoods, used to classify them according to 'lifestyle', are compiled from the census as a starting point, are supplemented by data from the electoral roll and housing data, and, in particular, they incorporate market research and sales data. The neighbourhood profiles have been developed and compiled commercially to aid the targeting of sales marketing strategies. We have used these characterisations of neighbourhoods to link postcoded mortality and other health-related data to them. Though the Super Profile Lifestyle Groups were never intended to discriminate households by health status, they certainly do. The South East of England is wealthy and healthy, compared with the rest of England but, even within it, there are three-fold differences in coronary heart disease mortality between affluent and poor neighbourhoods.[7] As further examples, there are also strong gradients across neighbourhoods for lung cancer incidence (studied by linkage with cancer registry data) and tuberculosis (studied by linkage with infectious disease notifications). The data for road traffic accidents, studied by linkage with hospital data, show an association with rural living (one of the Lifestyle Profile subgroups) with relatively high admission rates for road-related accidents in country areas. The Lifestyle Profiles even pick up the reverse gradient of breast cancer: they show breast cancer to be more common in affluent than poor neighbourhoods. Thus, the Super Profile Lifestyle data add further evidence to the conclusion that social advantage/disadvantage are important determinants of health status, at least in England.

Unemployment and health

Cross-sectional studies show relatively high mortality rates associated with unemployment. But how much of this is selection bias – a consequence of sick people becoming unemployed? Linked longitudinal data help clarify some of the relationships between unemployment and mortality. In the 1971 and 1981 censuses in England people were asked to categorise themselves as employed, not employed but seeking work, temporarily sick, permanently sick, retired, student, or looking after home. Studies undertaken at the time showed that people who were sick and unemployed tended to regard themselves, and record themselves, as sick rather than unemployed. Linked data from the LS shows that mortality rates were higher in those seeking work than in the employed (Table 6); and, as would be expected, still higher rates in those who classed themselves as sick.[4] These data add confirmation that unemployment may have an adverse effect on mortality that is independent of ill-health at the time of initial unemployment.

Employment status	Men
Employed (n=5551)	84
Seeking work (n=812)	132
Temporarily sick (n=247)	269
Permanently sick/disabled (n=908)	338
Total	100

Table 6 Standardised mortality ratios for men of working age, by employment status at the 1981 census (for deaths that occurred in 1981–92)

What about social class and unemployment? The LS can be used to link both social class and unemployment to mortality and to discriminate between them. The findings show a clear social class gradient within employment categories – the employed and the unemployed (Table 7); and a difference between the employed and unemployed within social classes.[4]

Social class	Employed Unemployed
1	56 92
2	74 102
3nm	82 118
3m	86 118
4	97 139
5	112 176

SMR for total male population=100

Table 7 Mortality of men by social class and economic activity (as they were at the 1981 census), expressed as SMRs for deaths in 1981–1992

What about unemployment at two successive censuses? The LS provides data on that too. Mortality rates were higher in men unemployed at the time of both censuses than in men unemployed at either one only.[4] In all, the data seem unequivocal in showing that involuntary unemployment can be bad for health.

Ethnicity and mortality

Linkage of birth certificates to death certificates has been done routinely in England for a number of years now. It can be used to study, among other things, infant and child mortality by mothers' place of birth. As an example, there are considerably higher infant mortality rates in infants born of Pakistani than UK-born mothers, which is not explained by social class (Table 8).[4]

Mother's country of birth Social class:	UK Pakistan
1	4.2 7.2
2	4.3 12.0
3nm	5.1 11.3
3m	5.7 11.9
4	6.2 11.5
5	7.5 16.0

Table 8 Infant mortality in England by mother's country of birth and social class: death rate per 1000 live births

Income distribution

A particularly thought-provoking topic of study has been that of whether communities with substantial inequalities in income and wealth tend to have higher mortality rates, other things being equal, than communities with more equitable distributions of income and wealth. These studies are particularly associated with the work of Richard Wilkinson from Sussex University.[8] There is empirical evidence, particularly from studies of communities in the United States, that areas with very unequal distributions of wealth tend to have generally high mortality rates. In other words: are unequal societies unhealthy societies? But is this a real finding or artefact through confounding? And if it is at least partly real, is it a universal phenomenon, or does it depend or how individual societies are organised? In a study published recently in the British Medical Journal, investigators studied birth cohorts, born a long way back in time and so now middle-aged, in different parishes across Denmark.[9,10] The investigators linked data from Danish registers on housing, income, occupation and education. They confirmed the well-established finding, within populations, that there is a relationship between income and mortality at the level of individuals – generally,

higher income is associated with lower mortality. At the level of geographical areas, they reported no association between the extent of inequality of income in the area and the level of mortality in the area. They speculated that the 'Danish welfare system might even out the effects of the pathways which might otherwise have connected inequalities and health'. I raise the question, which I will return to at the end, of whether these kinds of associations, or lack of them, may be country-specific and society-specific.

Growing old

My final topic is that of ageing. From young adulthood, mortality rates rise exponentially with age. Life expectancy is increasing over time in industrialised countries outside the old Soviet Union (where life expectancy has decreased in recent years). Many industrialised countries are concerned about the rising numbers of elderly in their populations, and its implications. As life expectancy increases, does morbidity get compressed so that people live in reasonable health until shortly before death? Or is increasing longevity merely an extension of ill health? It is clear that the elderly are heavy users of health care. It is now also clear that, for many people, most of their 'lifetime' use of health care occurs in the last few months of life. It follows that increased use of health care by the elderly may result from their being typically closer to death than young people. A number of research groups have been interested to know: which is the more powerful determinant of use of hospital care – ageing or proximity to death?[11-13]

We and others have used linked data, linking deaths to antecedent hospital care, to try to address this. In our studies, we identified data on all people who had died aged 45 years or over.[12] We accumulated the number of days spent in hospital by them. At every age at death, there were heavy users and light users. Regardless of age, for the majority of people most of their use of hospital care was made in the last year or two of life. Use increased with advancing age at death; but not very dramatically. Summarising the evidence, typically, dying is 'high cost' in terms of medical care.[13] Proximity to death is a stronger determinant of care, and cost, than age. Elderly people are heavy users of hospital care, but this is because typically they are closer to death than younger people. The data suggest that, at least in respect of life-threatening disease requiring hospitalisation, morbidity is indeed compressed, at least to some extent, with increasing longevity.

Conclusions

Concluding, I hope that I have shown that linked data can give useful insights into determinants of health and of health care. With some of these determinants, I doubt that there are universal truths that transcend all societies. For example, the impact on health of occupation, social class, unemployment, ethnic origin, income inequality, and perhaps even growing old, may not necessarily be the same in Australia as it is in England or in the Scandinavian countries. It follows that, if one is interested in these issues, these are relationships that may need to be tested afresh in different places and at different times. Linkage of routinely available data can be a useful and economical way of gaining at least some initial insights into them.

References

1. Black D (Chairman). Inequalities in Health. Report of a Research Working Group. London: DHSS 1980.

2. Townsend P, Whitehead M, Davidson N (eds). Inequalities in Health: the Black Report and the health divide. London: Penguin, 1992.

3. Hattersley L, Creeser R. Longitudinal Study 1971-1991. Office of Population Censuses and Surveys series LS No. 7. London: HMSO, 1995.

4. Drever F, Whitehead M. Health Inequalities. Decennial Supplement. Office of National Statistics Series DS No. 15. London: The Stationery Office, 1997.

5. Drever F. Occupational Health. Decennial Supplement. Office of Population Censuses and Surveys. Series DS no.10. London: HMSO, 1995.

6. Roberts SE. Hazardous occupations in Great Britain. Lancet 2002; 360: 543 544.

7. http://www.sepho.org.uk

8. Wilkinson RG. Income distribution and life expectancy. British Medical Journal 1992; 304: 165-168.

9. Osler M, Prescott E, Gronbaek M et al. Income inequality, individual income, and mortality in Danish adults: analysis of pooled data from two cohort studies. British Medical Journal 2002; 324: 13-16.

10. Mackenbach JP. Income inequality and population health. British Medical Journal 2002; 324: 1-2.

11. Henderson J, Goldacre MJ, Griffith M. Hospital care for the elderly in the last year of life. British Medical Journal 1990; 301: 17-19.

12. Himsworth RL, Goldacre MJ. Does time spent in hospital in the last year of life, increase with age at death? A population based study. British Medical Journal 1999; 319: 1338-1339.

13. Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: a red herring? Health Economics 1999; 8: 485-496.