

What socioeconomic factors shall I measure and include in my analyses?

Tony Blakely MBChB, MOH, PhD, FNZCPHM

Department of Public Health, University of Otago, Wellington, New Zealand

Occupational or social class, income, and education are classically the three main measures of socioeconomic position in quantitative sociology. But that has changed with the advent of measures of deprivation. The first and most prominent set of deprivation measures are those that use indices (e.g. from principal components analysis) of 'average' socioeconomic position for people living in an administrative area used in censuses. Examples are the Townsend and Carstairs indices in the UK.^{1,2} In the 1990s Salmond, Crampton, and Sutton developed such an area-based index of deprivation in New Zealand³—the so-called NZDep index that is now ubiquitous in New Zealand research, funding formulas, and advocacy. The second set of deprivation measures are actually calculated at the individual level, and can be likened to measures of (usually) material hardship. They are obviously strongly associated with income or poverty, but they are not necessarily the same; some people in poverty 'get by' without necessarily being deprived or in hardship and, conversely, some people with reasonable incomes struggle.

Salmond and Crampton present a paper in this edition of the Journal⁴ comparing and contrasting their new individual-level measure of deprivation (NZiDep)⁵ and the Ministry of Social Development's ELSI measure of living standards,⁶ the existing and entrenched small-area measure of deprivation (NZDep), and two of the three 'classic' measures of socioeconomic position (education and income). Their aim is to provide guidance to researchers—especially primary health care researchers—as to what is the optimal selection of socioeconomic indicators to include in research. Their answer? NZiDep trumps ELSI (mainly in terms of simplicity to measure), NZDep explains a bit more again than NZiDep alone, and actually you might get away with ditching measures of

income and education and only need to measure NZDep and NZiDep.

At this point you might well say, 'They would say that as they created both measures'. Declaring my own conflict of interests, I direct the Health Inequalities Research Programme (www.uow.otago.ac.nz/HIRP-info.html) that hosts the Socio-economic Deprivation Indices project. I use both NZiDep and NZDep measures (as well as income and education) in my research (but what researcher in New Zealand does not use NZDep?). However, I have not been directly involved in the development of either NZiDep or NZDep. I broadly agree with Salmond and Crampton's conclusions. First, the NZDep index is so easy to assign to most research files, so long as address is collected—which it almost always is. Furthermore, if you can code your addresses to mesh-block level (about 100 to 150 people in a 'small' neighbourhood), then conceptually the NZDep index is getting pretty close to an individual-level measure. Here I deviate somewhat from Salmond and Crampton in interpretation.

Consider when we ask someone's income on the census form. They are expected to accurately recall their previous year's income, add on benefits and family tax credits, and then tick the right box. And by 'right box', we mean that applying to true gross income in the last 12 months—not lifetime or other periods of income. Therefore, the income measure is prone to substantial measurement error. You could argue that your personal socioeconomic position is actually better predicted by the deprivation of the (small) neighbourhood you live in, which presumably reflects the culmination of many socioeconomic processes such as lifetime income, asset wealth, education (a key determinant of cultural habitus in Bourdieu jargon—or 'who you want to live

J PRIM HEALTH CARE
2012;4(4):268–270.

CORRESPONDENCE TO:

Tony Blakely
Professor, Department of Public Health, University of Otago, Wellington, New Zealand
tony.blakely@otago.ac.nz

next to'), and so on. So—as any research student walking through my door in the last 10 years can attest to—one can make a strong case for NZDep being as good a proxy for individual-level socioeconomic position as any of the classical measures of class, income, and education in the New Zealand setting. And, if I could only choose one measure of socioeconomic position in my primary health care research, I too would choose NZDep. And I would interpret any association of NZDep with a health outcome as largely reflecting individual-level socioeconomic position, with a bit of 'contextual overlay' (e.g. just as Salmond and Crampton speculate that smoking has a likely association with neighbourhood-level deprivation, so would I).

But what if I can include two or more measures of socioeconomic position in my research? Now you have a tricky optimisation problem, and you

commentary); and to explicitly determine the (sometimes causal) association of a given socioeconomic factor with a given health outcome. Regarding the former, the task at hand is to adjust as fully as possible for the complex construct of socioeconomic position. Having, say, adjusted for NZDep, the next socioeconomic factor you ideally want to adjust for is that measure of socioeconomic position that captures as much as possible of the remaining (residual) confounding by socioeconomic position. Strictly speaking, this requires watching what happens to the association of your exposure with your outcome as you add (theoretically justified) potential confounders. If we use Salmond and Crampton's paper as guidance here, we are assuming that the 'additional R^2 gain' from adding sequential measures of socioeconomic position will be a good correlate of what really matters—how much further residual confounding of your exposure–outcome associa-

Broadly speaking, there are two types of reasons for measuring and adjusting for socioeconomic position in your analyses: to adjust for socioeconomic position as a confounder (and sometimes a mediator—but I put that aside in this commentary); and to explicitly determine the (sometimes causal) association of a given socioeconomic factor with a given health outcome.

need to really nut through what it is you are trying to do. And here I am circumspect about how much weight we can place on the paper by Salmond and Crampton. They make comparisons of the R^2 —essentially a measure of how much variation in the outcome is explained by your model and its included independent variables. That is all well and good, informative, and statistically interesting. But what clues and guidance does this type of R^2 analysis provide applied researchers?

Broadly speaking, there are two types of reasons for measuring and adjusting for socioeconomic position in your analyses: to adjust for socioeconomic position as a confounder (and sometimes a mediator—but I put that aside in this

tion you are capturing. This assumption might be reasonable, but not always so. Assuming it is for now, Salmond and Crampton suggest that using NZiDep (the individual-level measure of deprivation) might be the next best thing to adjust for after NZDep, in that adjusting for other socioeconomic factors over and above both NZDep and NZiDep explains little further variation in the data. But this is not quite what most researchers are interested in. For example, a researcher might be interested in the association of eating fruit and vegetables (exposure) with the onset of ischaemic heart disease (IHD; outcome). In this case, there are many potential confounders of this exposure–outcome association (e.g. physical activity), and many of these potential confounders are

patterned in turn by socioeconomic position. So adjusting for (numerous) socioeconomic factors can achieve near-full adjustment for confounding. In this case, the theoretically pure advice to researchers is 'adjust for as many measures of confounding and socioeconomic position as you can'. More practically, Salmond and Crampton suggest that once you have adjusted for NZDep, measure enough covariates to adjust for NZiDep. Such advice is reasonable. But, regardless, the plague of measurement error and residual confounding means that you will probably never quite fully capture these confounding constructs (collider and other introduced biases⁷⁻¹⁰ aside for now). So, if after adjusting for NZDep your odds ratio association of not eating fruit and vegetables with IHD reduces from 1.4 to 1.2 (putting confidence intervals aside), you would be most unwise to conclude that there is a true 20% excess risk of IHD from not eating fruit and vegetables with IHD; the remaining 20% could well be due to residual confounding by other socioeconomic factors (and/or other more proximal confounders). Adjusting for further measures of socioeconomic position will assist you gain confidence in the causal nature of any remaining association, assuming you have sufficient power, no major problems of collinearity and you are not making matters worse by adjusting for variables that actually introduce more bias than they reduce. These assumptions stated, I would be happy with a research team's decision to measure and adjust next for NZiDep, then income—in accord with Salmond and Crampton.

Regarding the latter task of measuring the intrinsic association of socioeconomic position with a health outcome, one should be less driven by 'what socioeconomic measurement has the strongest association with my health outcome' or 'what variable explains the most variation in my data'. Rather, one should be driven by theory. For example, if you are positing that income level is associated with your ability to afford a healthy diet, then you should be measuring income. Or if you were looking at the role of knowledge and education in determining smoking among Tongans, then measure education.

Summarising, Salmond and Crampton clearly show that one relatively easy-to-collect measure

of individual-level deprivation—NZiDep—clearly has a strong association with a range of health outcomes, and often independently of other socioeconomic factors. (Indeed, we too are finding strong associations of NZiDep with psychological distress in forthcoming work, and—closer to causality—changes in NZiDep over time tend to have stronger associations with changes in health than, say, changes in income.) Thus, more research on individual-level deprivation, probably using NZiDep, is clearly indicated and likely to be policy-relevant. And if your objective is to fully adjust for socioeconomic position as a confounder in your research, then definitely adjust for NZDep and also adjust for as many other socioeconomic factors as you can.

References

1. Carstairs V. Deprivation indices: their interpretation and use in relation to health. *J Epidemiol Community Health*. 1995;49(Suppl 2):S3–S8.
2. Carstairs V, Morris R. Deprivation and mortality: an alternative to social class? *Community Med* 1989;11(3):210–19.
3. Crampton P, Salmond C, Sutton F. NZDep91: a new index of deprivation. *Soc Policy J NZ* 1997;9:186–93.
4. Salmond C, Crampton P. *J Prim Health Care*. 2012;4(4):271–280.
5. Salmond C, Crampton P, King P, et al. NZiDep: a New Zealand index of socioeconomic deprivation for individuals. *Soc Sci Med*. 2006;62(6):1474–85.
6. Jensen J, Spittal M, Crichton S, Sathiyandra S, Krishnan V. Direct measurement of living standards: the New Zealand ELSI scale. *Nga Whakaaturanga Ahuatanga Noho*. Wellington: Ministry of Social Development; 2002.
7. Hernan M, Hernandez-Diaz S, Robins J. A structural approach to selection bias. *Epidemiology*. 2004;15(5):215–55.
8. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60(7):578–86.
9. Glymour M, Greenland S. Causal diagrams. In: Rothman K, Greenland S, Lash T, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p.183–212.
10. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14(3):300–06.