Counterfactual challenges to social epidemiology

Tony Blakely
Alistair Woodward
Department of Public Health, Wellington School of Medicine, University of Otago, Wellington, NZ
tblakely@wnmeds.ac.nz

The study of the socio-economic determinants of health is receiving much interest world-wide. The publication of the Black Report[1] in 1980, a UK government sponsored review of the socio-economic determinants of health, was a turning point. Since then, researchers and policy makers alike have been drawn to (or forced to) attempt to understand the relationship of socio-economic status with health, and ponder the policy implications. Nineteenth century epidemiology had a strong focus on the distribution of disease by social forces in society, but post World War II 'modern' epidemiology has focused more on individual-level associations and associated methodologies.[2-5] Recently, a growing cadre of epidemiologists are returning to the study of how health is patterned and caused by socio-economic forces in society. Why? Numerous reasons, including: [3,6-8]

- A growing awareness of the limited value of studying only individual-level risk factors (e.g. income, diet, cell-phone use), when the these risk factors are strongly 'patterned' by socio-economic status and other distal forces.[9]
- Following that landmark work of Rose[10] and infectious disease epidemiologists (e.g. Koopman and Longini[11,12]), researchers and policy makers have been provoked to think less in terms of individual-level associations, and more in terms of population-level dynamics.

Berkman and Kawachi (2000) define social epidemiology as "the branch of epidemiology that studies the social distribution and social determinants of states of health" (page 6).[13] As such, the orientation is similar to other branches of epidemiology that focus on exposures (e.g. environmental or nutritional epidemiology). Why launch this sub-discipline 'social epidemiology'? The most important justification is the need to identify the casual components of associations. Much of what passes for social epidemiology at the moment is descriptive; this is important work, if only for monitoring and raising awareness of the underlying determinants of health. But if 'social epidemiology' is to have a notable impact, then it must not just describe associations but move on to isolate causal mechanisms. For example, we should not be happy with just noting the strong gradient of health by income, but rather ask (then attempt to answer) what is the causal contribution of income on health as opposed to income being a proxy/correlate for numerous other variables.

This potential analytical contribution of social epidemiology, and the challenges to be overcome, is the subject of this review. We wish to briefly review a recently emerging body of literature on the application of a counterfactual model to causal inference in social epidemiology.[14-19] Our main purpose is to alert readers to what we think are interesting and challenging issues in epidemiology generally, but particularly social epidemiology. We also believe that these emerging issues will gain more widespread attention by epidemiologists in the next decade. Our own reflections on these are somewhat preliminary, so we would encourage dialogue with other interested epidemiologists in Australasia.

Causal inference in social epidemiology

Say a cohort study found that after controlling for the 'standard' confounders of age, sex, and ethnicity, low-income people have twice the mortality of high-income people in the first five years of follow-up. Imagine we could go back in time to the start of the study, and top-up all the low-income peoples' income to a high income – but change nothing else. In this 'counterfactual' thought experiment, would we expect the two-fold mortality risk between the two originally defined income groups to be reduced to nothing? Certainly not – for two reasons.

First, as socio-economic factors are distal factors in any causal process that eventually impacts upon health, it takes time for income to get 'under the skin' and affect health. For example, a higher income may afford you a better diet, which changes your lipid profile, which – over time – may affect your risk of coronary heart disease. Beyond that, there is a further delay from the pathophysiological beginnings of disease and clinical presentation or death. More generally, for any socio-economic factor (e.g. education, occupational class), it will probably take years, decades, or even generations for its impact on chronic disease mortality to be affected through the most likely pathways: psychosocial processes, behaviour, and other lifestyle factors.[20]

Second, high-income and low-income people are not 'exchangeable' in a counterfactual sense.[14] The counterfactual model[21,22] is perhaps the most useful way of thinking about confounding and attribution, and has been the foundation of much recent methodological work in epidemiology.[23-25] Simply, a counterfactual model in observational epidemiology states that the average causal effect of an exposure could be determined if we could go back in time, change individuals' exposure status (and nothing else), and determine the effect of this change on disease occurrence. Obviously, this is impossible. However, a large randomised trial approximates the counterfactual model; as exposure is randomised we can assume that the individuals in each treatment arm are very similar, and hence could be considered as 'exchangeable' between treatment arms. In observational epidemiology, the exchangeability assumption requires a bigger jump of faith – we assume that within strata of all the measured confounders, individuals are exchangeable, and hence we can determine causal effects for a given exposure. When individuals are not similar for unmeasured confounders within strata of the measured confounders, then the exchangeability assumption is violated and the observed exposure-disease association suffers from residual confounding.

Kaufman and Cooper (1999) argue that violation of the exchangeability assumption is particularly problematic in social epidemiology.[14] First, exposures and confounders in social epidemiology are often highly correlated. For example, even after controlling for the confounders of education and social class, income is likely to be correlated with factors such as asset ownership, and other less readily measurable variables that are linked in a general way with social position, and may act as confounders in a study of disease etiology. That is, even with many confounders measured, it is still likely that individuals are not exchangeable between strata of the exposure of interest - income, in this case. From a life-course perspective, it is unlikely that enough life-course variables (e.g. parental income, early childhood environment) could ever be measured to ensure that individuals were exchangeable within strata of, say, adult income. (We would add to this lack of ability to measure all potential confounders the problem of measurement error - measurement error of confounders will limit the ability to control for confounding by the same.[26-28]) Second, Kaufman and Cooper argue that this lack of exchangeability becomes even more problematic for variables that are deeply embedded attributes of individuals, such as ethnicity. Whilst it is plausible to imagine the same person with a different income, they argue it is less plausible to imagine a white American person as an Afro-American. If the exchangeability assumption cannot be sustained in social epidemiology, we must accept that there will remain residual confounding of measured exposure-disease associations. Kaufman and Cooper propose three alternative strategies for progress in social epidemiology:

- Seek causal explanations only for definable interventions (e.g. income, but not ethnicity)
- Borrow methods from infectious disease epidemiology to allow for lack of independence between individuals in outcomes (e.g. social interactions might be viewed as analogous to a contagion effect (see Blakely and Woodward[29]) that is essential to understanding the causes of disease (and policy responses), but violates normal analytical assumptions of statistical independence)
- Use a systems approach,[12,30] whereby we attempt to model the actual associations that exist, and test predictions against future observations. (In the absence of complex systems models, then at least consider examining the associations of distal and proximal variables to one another, and mapping outcomes by joint distributions of covariates (see Kaufman et al[31]).

Not all social epidemiologists agree with the diagnosis of Kaufman and Cooper - see the comments of Muntaner[15] and Krieger and Davey Smith[17], and the authors' replies.[16] Two concerns with the position of Kaufman and Cooper stand out to us. First, analytical models are never perfect, yet often provide useful inferential information. Standard epidemiological analyses (e.g. standardisation, regression modelling) may not fully satisfy the exchangeability assumption of a counterfactual model, but nevertheless provide useful information about the structuring of health in our societies. Kaufman and Cooper perhaps overstate the level of methodological rigour required to inform public health actions. For example, it is useful to know that the association of income with health is, say, halved after controlling for the 'confounder' education; but that does not mean we need causally infer that this residual association is equivalent to the change that would be affected by changing peoples' incomes. Second, even for variables that may be argued as attributes of individuals (e.g. ethnicity) rather than variables subject to change (e.g. income), it is not inappropriate to assess effect sizes. As argued by Krieger and Davey Smith (2000) and Muntaner (1999), ethnicity is not a fixed genetic trait of individuals, but a socially constructed fact.[15,17] Biological features may be fixed, but the social response is not. For example, it is not impossible to imagine a counterfactual world where people were not discriminated against on the basis of ethnicity. Whilst we must be cautious interpreting effect estimates for attributes like ethnicity, gender and age (and in particular, should resist the temptation to closely equate effect measures with attributable and avoidable fractions), they provide further clues to the patterning of socio-economic inequalities of health. Indeed, Susser argues that it is necessary to think of personal attributes as one of the several Chinese boxes in an eco-epidemiological approach.[8]

Putting observational social epidemiology aside for a moment, consider the possibility of a randomised trial of income supplementation on health. Evidence of health effects from such a study would be very welcomed; as a randomised trial it would generate a 'higher quality' of evidence, and more closely approximate the counterfactual model. Conducting such a trial may seem impractical given the cost and possible political implications. However, several income supplementation trials were conducted during the 1960s and 1970s in North America, but no reliable analyses of health outcomes were conducted. What is worse, the remaining raw data are not amenable to reliable analyses - a lost opportunity to assess health outcomes.[32] Returning to our present conjecture, if we could conduct a randomised trial of income supplementation, what would be the challenges and limitations?

First, and as suggested above, the effect from a change in income is likely to be much less than that measured in observational studies due to residual confounding. For example, the association of personal income with health in observational studies may be largely due to a correlation with (unmeasured) parental income and, more generally, parental socio-economic status. Second, as in any epidemiological study, the change in exposure has to be substantive enough in size, for a long enough duration, and applied to enough people to cause a measurable change in outcome. Thus, convincing a government agency to increase welfare benefits by $20 per week for one year among 1000 unemployed people is unlikely to effect any measurable change in health outcomes. Moreover, there may be genuine threshold effects for both the size of the income change that is required to have an effect on health status, and threshold effects for the length of time over which income changes apply. One study design that may overcome these obstacles is a study of lottery winners, where the winnings were paid out as sizeable amounts on a regular basis over some years (personal communication Dr Anthony Rodgers, July 2000, Clinical Trials Research Unit, Auckland, New Zealand). Third, the length of follow-up will have to be long. There are undoubtedly acute effects of substantial changes in income

(e.g. intake of alcohol, risk-taking behaviours), but it will take time for a change in income to fully get 'under the skin' via dietary and other mechanisms, and affect physiological markers of health (e.g. lipid levels), then morbidity (e.g. coronary heart disease incidence), and finally mortality. Moreover, before income can get under the skin, one may have to first accumulate wealth. Whilst the lag time from increased income to improved self-rated health outcomes (e.g. "do you rate your health as excellent/good/average/fair/poor") may be shorter, many researchers and policy analysts would be concerned about whether a 'real' change in health was being measured, or rather contamination from increased happiness due to financial support. In short, results for self-rated health outcomes only would not be convincing. Fourth, the results from such an income trial are rather artificial. Other than by government redistribution of income, income change in the real world is associated with other life changes such as personal investment in education, or the children leaving home thereby increasing disposable income - it may be these changes that affect health, or at least interaction of these changes with income that affects health. Finally, the results from exactly the same study design may vary between contexts. For example, income supplementation in New Zealand may have positive effects on health in New Zealand, but not in Finland; Finland has more entrenched social welfare support than New Zealand, suggesting that income (only) supplementation may be less critical in an egalitarian country like Finland. In epidemiological terms, this is equivalent to the prevalence of effect modifiers varying between what are otherwise two identical experiments.

Our purpose in conjecturing about a randomised trial of income is not to dismiss it as useless; to the contrary, we think it would be extremely useful. Our purpose is instead to highlight that:

- any such randomised study must be large, have a meaningful change in exposure, and be followed up for a long period of time
- thinking in a counterfactual sense (as approximated by the concrete example of a randomised study) suggests that the association of income with health will be less that that measured in observational studies
- thinking in policy terms (as approximated by a randomised study of income supplementation) we should not expect immediate gains to health from increasing personal incomes; the improvements in health are likely to be predominantly via pathways that take years, decades, or even generations. A possible rebuttal to this caution is that the deterioration in health in Eastern Europe after the collapse of the Soviet Union was rapid. However, time lags to poor health versus time lags to good health may be different - this area requires further research.

Conclusion

The counterfactual model is becoming a key model in epidemiological thinking, and raises substantial challenges to social epidemiology. We have presented a brief overview of a body of emerging work on the implications of the counterfactual model to social epidemiology - we encourage interested readers to view this interesting and thought provoking literature themselves, and maintain a watching brief for further developments. How useful the counterfactual model in social epidemiology will be remains to be seen. From a social science perspective, the counterfactual model remains atomistic. For example, analysing changes in income alone does not account for the moderating influences of other social policy and institutions.

Acknowledgments

Philippa Howden-Chapman for useful comments. Anthony Rodgers for discussion around the idea of a randomised income trial.

References

1 Black D, Morris J, Smith C, Townsend P. The Black report. In: Townsend P, Davidson N, Whitehead M, (1992), editors. Inequalities in Health. Hammondsworth: Penguin, 1980.
2 Susser M. Epidemiology in the United States after world war II: the evolution of technique. Epidemiol Rev 1985;7:147-177.
3 Susser M. Does risk-factor epidemiology put epidemiology at risk? Peering into the future. J Epidemiol Community Health 1998;52:608-611.
4 Pearce N. Traditional epidemiology, modern epidemiology, and public health. Am J Public Health 1996;86(2):678-683.
5 McMichael A. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. Am J Epidemiol 1999;149:887-897.
6 Lynch J, Kaplan G, Salonen J. Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. Soc Sci Med 1997;44:809-819.
7 Berkman L, Kawachi I. A historical framework for social epidemiology. In: Berkman L, Kawachi I, editors. Social Epidemiology. New York: Oxford University Press, 2000:3-12.
8 Susser M, Susser E. Choosing a future for Epidemiology: II. From black box to chinese boxes and eco-epidemiology. Am J Public Health 1996;86(5):674-677.
9 Susser M. Causal thinking in the health sciences: concepts and strategies in the health sciences. New York: Oxford University Press, 1973.
10 Rose. The strategy of preventive medicine. Oxford: Oxford University Press, 1992.
11 Koopman J, Longini I. The ecological effexts of individual exposures and nonlinear disease dynamics in populations. Am J Public Health 1994;84:836-842.
12 Koopman J. Comment: emerging objectives and methods in epidemiology. Am J Public Health 1996;86:630-632.
13 Berkman L, Kawachi I, editors. Social Epidemiology. New York: Oxford University Press, 2000.
14 Kaufman J, Cooper R. Seeking causal explanations in social epidemiology. Am J Epidemiol 1999;150:113-120.
15 Muntaner C. Invited commentary: social mechanisms, race, and social epidemiology. Am J Epidemiol 1999;150:121-126.
16 Cooper R, Kaufman J. Is there an absence of theory in social epidemiology? The authors respond to Muntaner. Am J Epidemiol 1999;150:127-128.
17 Krieger N, Davey Smith G. Re: Seeking causal explanations in social epidemiology [letter]. Am J Epidemiol 2000;151:831-832.
18 Kaufman J, Cooper R. Re: Seeking causal explanations in social epidemiology - the authors reply [letter]. Am J Epidemiol 2000;151:832-833.
19 Kaufman J, Poole C. Looking back on "Causal Thinking in the Health Sciences". Annu Rev Public Health 2000;21:101-119.

20 Turrell G, Oldenburg B, McGuffog I, Dent R. Socioeconomic determinants of health: towards a national research program and a policy and intervention agenda. Canberra: Centre for Public Health Research, School of Public Health, Queensland University of Technology (in association with the Health Inequalities Research Collaboration), 1999.
21 Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66:688-701.
22 Rosenbaum P. From association to causation in observational studies: the role of strongly ignorable treatment assignment. J Am Stats Assoc 1984;79:41-48.
23 Rothman K, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, 1998.
24 Greenland S, Robins J. Identifiability, exchageability, and epidemiological confounding. Int J Epidemiol 1986;15:412-418.
25 Greenland S. Randomization, statistics, and causal inference. Epidemiol 1990;1:421-429.
26 Davey Smith G, Phillips A. Confounding in epidemiological studies: why "independent" effects may not be all they seem. BMJ 1992;305:757-759.
27 Phillips A, Davey Smith G. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. Stat Med 1992;11:953-961.
28 Liu K. Measurement error and its impact on partial correlation and multiple linear regression analyses. Am J Epidemiol 1988;127:864-874.
29 Blakely T, Woodward A. Ecological effects in multi-level studies. J Epidemiol Community Health 2000;54:367-374.
30 Weed D. Beyond blackbox epidemiology. Am J Public Health 1998;88:12-14.
31 Kaufman J, Cooper R, McGee D. Socioeconomic status and health in Blacks and Whites: the problem of residual confounding and the resiliency of race. Epidemiol 1997;8:621-628.
32 Connor J, Rodgers A, Priest P. Randomised studies of income supplementation: a lost opportunity to assess health outcomes. J Epidemiol Community Health 1999;53:725-730.