EDITORIAL

Social epidemiology: towards a better understanding of the field

Stanislav V Kasl and Beth A Jones

In the February 2001 issue of the International Journal of Epidemiology the section on ‘Point-Counterpoint’ offered a one-page statement by Zielhuis and Kiemenej1 (Z&L) that was entitled ‘Social epidemiology? No way.’ Had this been submitted to the New England Journal of Medicine, such a statement would have been met with quiet editorial approval. But in this journal, with its marvellous tradition of publishing a broad range of epidemiological studies that investigate the role of a large array of biomedical, social, and psychological influences on health status, this opinion piece was met with considerable opposition, if not hostility.2–6 Zielhuis and Kiemenej argued that the essence of epidemiology resides in its biomedical approach to aetiology. Here, the behavioural and social sciences have no contribution to make, though they can participate in descriptive epidemiological studies (‘frequency research’). This extremely narrow, reductionistic perspective on the work of epidemiologists precludes the study of the more distal determinants of health; those represented by the various societal, social, and psychological processes. In addition, this view completely uncouples epidemiology from any kind of a broader public health orientation.

The additional argument put forward by Z&L was that psychosocial epidemiologists would need ‘to obtain a thorough training in biomedical and social sciences’ to make an effective contribution. The unstated inference is that there are no such dually trained scientists. While this may be true, it is not clear why an interdisciplinary research team is not an adequate substitute. Ultimately, the failure of the Z&L statement to be persuasive is more pragmatic than scientific: researchers do not like to be told by putative colleagues what to think, what labels to use, and, above all, what not to study. Most of the debate needs to be a methodological one, not an ideological one. While psychosocial aetiology of organic disease is not an oxymoron, it nevertheless represents a methodological and conceptual challenge in which, often enough, the difficulties and limitations of biomedical epidemiology and the social/behavioural sciences interact synergistically. So if the message of the Z&L statement is, not that social epidemiology is a forbidden word or an impossible discipline, but that it is a very difficult research domain, then one must readily agree with them.

A recent issue of Psychosomatic Medicine contains ‘The Great Debate’: Can psychosocial interventions improve clinical outcomes in organic disease?7 Both sides agreed to comment on the same set of articles—those chosen to give the psychosocial perspective its best shot—and many of these studies were observational designs representing the core of social epidemiology. The ‘negative’ speakers commenting on the evidence were the current and the former editors of the New England Journal of Medicine. Their criticisms, interestingly enough, were not unique to the domain of social epidemiology studies, but were part of the traditional list of methodological weaknesses, such as: (1) data dredging; (2) failure to account for confounders; and (3) abandoning the ‘intent-to-treat’ data analysis guidelines for randomized clinical trials. Thus it remains unclear if research in social epidemiology has some unique challenges, or if it simply has more ample weaknesses of the traditional kind. Interestingly, Drs Sellman and Angell did some ideological boundary-setting of their own: for example, they refused to be interested in psychosocial effects on physical health which were mediated by lifestyle habits or adherence.

This issue of the International Journal of Epidemiology contains a number of articles that fall within the domain of social epidemiology. Presumably, they represent submitted (not solicited) papers that were accepted and are assembled together in one issue, but are not meant to define the field in any way. They represent a nice range of illustrative studies, but they do not constitute a comprehensive picture of the field. The boundaries of social epidemiology—the study of social and psychological determinants of health—are neither sharp nor unchanging. Perhaps the recent volume, Social Epidemiology, edited by Berkman and Kawachi,8 can be used to illustrate some of the important research themes. Among others, it contains chapters on: (1) socioeconomic position, income inequality, and discrimination; (2) working conditions and job loss; (3) social integration and social cohesion; (4) depression and other affective states; (5) social context of health behaviours and psychosocial interventions; and (6) ecological and multilevel approaches to social determinants.

In the commentary that follows, we seek to achieve a better understanding of the methodological and conceptual issues (difficulties) that arise when one is examining psychosocial influences on health within the context of a biomedical aetiology of disease. In other words, we believe that there are different or new challenges in social epidemiological research that go beyond those encountered in classical (biomedical) epidemiological approaches. First, we will identify and briefly describe some of these challenges. Then we will make specific comments on some of the articles with a view toward providing concrete illustrations of these points. We realize, of course, that our views may be somewhat idiosyncratic and that other investigators and scholars in this field would express these
issues differently. Above all, we believe that social epidemiology is a true integration of the biomedical and the social/behavioural disciplines, rather than a purely social/behavioural approach to the study of health issues on a population level. Given that, the primary concern is to account for the role (mechanisms) of psychosocial risk factors in biomedical aetiology of disease. Often, this involves keeping track of four inter-related issues: (1) study design; (2) measurement of exposure/risk factors; (3) data analysis strategies (in the service of clarifying aetiological pathways); and (4) aetiological theory about how exposure affects outcomes.

In classical epidemiology, our risk factors often come with their own biological plausibility, i.e. mechanisms of disease causation. For example, the way hyperuricaemia leads to gout is reasonably straightforward. On the other hand, the effect of social isolation on mortality is very unclear with respect to mechanisms involved. This ambiguity puts greater demands on our designs. Elsewhere,9 we have discussed this in terms of the need to utilize a disease development paradigm: we need designs that will allow us to locate the influence of the social risk factor at a specific point in disease development. For coronary heart disease (CHD), the stages include: (1) asymptomatic status, risk factor(s) absent; (2) asymptomatic status, risk factor(s) present; (3) subclinical disease susceptible to detection; (4) initial symptom experience; (5) initial disease event (diagnostic criteria met); (6) course of disease; and (7) mortality (case fatality). Ideally, the design will allow us to determine which transition point(s) is (are) affected by the social variable.

In prospective designs in classical epidemiology, the biological risk factors at baseline are entered into multivariate models predicting outcomes obtained during follow-up. We seldom agonize about the aetiological process prior to baseline that might have produced the particular pattern of associations among the predictors. On the other hand, with a mix of psychosocial and biological predictors at baseline, disentangling causal priorities among them is often important to guide the analysis. But since the baseline data are cross-sectional, we are often unable to do so. Thus, if unemployment status is the baseline exposure in a mortality study, an association of unemployment with poor health status at baseline is ambiguous: poor health is a confounder if it led to the exposure but a mediator if it is already a consequence of the exposure. This again puts greater demands on our research designs and leads to the recommendation9 that we should utilize the natural experiment paradigm whenever possible: this is a doubly prospective design (before disease and before exposure) which allows us to study the causal priorities among the variables which lead up to the exposure.

Biomedical risk factors are often ‘silent’ variables: people are unaware of them (e.g. low high density lipoprotein levels) or their impact is reasonably narrow (e.g. taking medication for high blood pressure). In contrast, psychosocial exposures can often be very ‘reactive,’ such as being in a demanding job or becoming unemployed: people respond and cope and adapt. This means that early on the exposure may have a different meaning than later on. This leads to complications such as: (1) length of exposure may lose meaning; (2) cohort follow-up may need to start at the beginning of exposure to detect effects; (3) the cohort may need to be monitored post-baseline for exposure-linked changes in behaviours, attitudes, perceptions, and so on.

It should be nearly self-evident that measurement of exposures in psychosocial epidemiology represents special challenges. For example, the distinction between ‘objective’ and ‘subjective’ measures would almost never come up in classical epidemiology but it is a major issue for certain research themes, such as the health effects of psychosocial work exposures.10 For work exposures, the two sets of measures are not well correlated and tend to yield different results. For other research themes, such as the influence of social support on health, objective measures are seldom even attempted, in part because it is not clear what the source of ‘objective’ data should be.

A persistent problem of measuring psychosocial exposures is that constructs, such as Type A behaviour, have been operationalized in different ways and do not show the same relationships to health outcomes. Often, we do not understand why; if we did, we might then argue that failure to replicate results really illustrates a high specificity of findings, but in most instances that would be an excessively optimistic interpretation. This (undesirably) high sensitivity of psychosocial measures to minor variations in operationalization (as well as to study designs, settings, type of respondents, and so on) remains a major obstacle to being able to argue that there is a convergence of findings.

At this point we wish to illustrate some of the points raised above by a selective commentary on a few of the articles published in this issue.

The job loss study from New Zealand11 is a fine illustration of the strategy of looking for natural experiments that can provide us with an unbiased exposure (unlikely self-selection) and pre-exposure data. This becomes a strong, convincing design. Unfortunately, the data collection was minimal, based primarily on available records. Most sorely missed are any data on the re-employment experience of the cohort that was made redundant, and, to a lesser extent, the later job losses among those who initially continued working. Thus it becomes impossible to interpret the results, which suggest very weak effects, in comparison to the broad set of unemployment studies.12 In addition, no information is given on the economic impact of the unemployment experience; in general, financial compensation tends to attenuate the impact of job loss, especially among lower skill blue collar workers.12 It might also be noted that the one significant finding, the apparent high rate of self-harm (hospitalization or death) among those losing a job, is not as much a function of an increase from baseline, but a (puzzling) decrease among those who continued working in a presumably more insecure economy. This is somewhat reminiscent of results from a panel study of alcohol abuse and unemployment:13 the big effect was the reduced risk for alcohol abuse among those who were employed, but the community rate of unemployment was going up. In psychosocial epidemiology, sometimes it takes us a while to understand our exposure variables!

The study of job strain and CHD in the Nurses’ Health Study14 contributes negative results to a rather smallish set of studies that are not in agreement on the role of job strain among women as a CHD risk factor. The authors point out that since this was a study of a single occupation (thus making the title somewhat misleading with its reference to US women), there could have been an insufficient variability in the measures of the predictors. While this should have been easy to check, in our opinion this is not the important issue. Since the psychological demands and job control measure are a mix of
descriptions of the work environment which may be shared by
other workers (quasi-objective) and of idiosyncratic perceptions
and reactions (subjective), then using the measures in a single
occupation means that the quasi-objective component is
reduced to near zero and the idiosyncratic component takes
over almost completely. A related point is that the Karasek
measures were developed for use across many occupations, but
in a study of a single occupation, probably more targeted and
precise measures of these concepts could be developed. It is
possible that in the health professions (especially MD’s) the
great amount of job control and discretion becomes a source of
psychological demands; thus in certain specific occupations,
more precise measurement may be necessary. So while blood
pressure remains blood pressure in classical epidemiology,
sometimes our measures of psychosocial exposures change in
meaning depending on contextual issues.

The Rosvall et al. study also deals with job strain but the
design is cross-sectional and the outcome is preclinical
atherosclerosis. The results they obtained are rather puzzling:
they do not fit the predictions from the job strain model and
whether they reflect ‘a more complex pattern of interaction’
abstract) which will stand up to replication is difficult to
predict. The authors quote an earlier paper in which one of us
was making the recommendation regarding the perspective of
the disease development schema, i.e. the study of psychosocial
influences at different stages of the disease process. It is nice to
be quoted, but the issue here is whether the recommendation
must be implemented in longitudinal designs to get its full
benefit. In a cross-sectional design in which cases with history
of clinical disease have been removed, an association between a
psychosocial variable and high plaque prevalence may be
difficult to interpret: (1) it may mean that the variable increases
the risk of progression from mere presence of risk factors to
development of subclinical disease; or (2) it may mean that the
variable is protective of the progression from subclinical disease
to clinical events, thus leading to a greater accumulation of
those with plaques. So while subclinical disease is a valuable
outcome in psychosocial studies because it is a relatively ‘silent’
risk factor, thus avoiding the problem of reverse causation
(disease influencing the risk factor), the full benefits of the
strategy come in longitudinal designs, such as the study of
hostility and subclinical progression.

The Sykes et al. study poses the question of whether
psychosocial variables could help explain the ‘French paradox,’
the combination of a diet high in saturated fats and lower CHD
mortality. Since neither dietary nor alcohol consumption is
included in analysis, this makes it a very truncated look at the
French paradox. Furthermore, no attempt is made to spell out
the paradox in psychosocial terms so that a relevant set of
variables would be used. Instead, the included variables are the
traditional psychosocial risk factors from general cardiovascular
epidemiology. The fact that none of them predicts hard out-
comes is difficult to understand but relates to the common
dilemma that minor variations in measurement may lead to
rather different results. In any case, the observed negative
profile of the risk factors for the French subjects becomes some-
what irrelevant to the study question posed, given the variables
did not relate to CHD.

The study of self-assessed health (SAH) and mortality repre-
sents an interesting paradox in psychosocial epidemiology: the
basic finding of the association with mortality (after controlling
for numerous variables, above all other indicators of health
status) is an uncommonly robust one, yet we still cannot deter-
mine of this is an interesting finding or not. The danger that an
unmeasured health status confounder will render this finding
trivial leads to the open-ended goal of controlling for ‘enough’
potential confounders. What would confirm it as an interesting
finding would be our ability to ‘explain’ (at least statistically)
the association of SAH with mortality by another psychosocial
construct. The present study attempts to do that but with no
theoretical justification for the particular choice of the
psychosocial variables. It is likely that the relevant construct, if
there is one, is some specific optimism/pessimism about health;
stressors or neuroticism or social support are only remotely
promising and some have been tried before. The one large gap
in our knowledge is that we do not know much about the
behavioural and psychological consequences of positive and
negative SAH. That is, nearly all studies control only for baseline
covariates and we need follow-up monitoring of the cohort to
see if there are intermediate consequences that are the mediators
of the link to mortality. In addition, we need to pay closer
attention to age as a moderator, since the meaning of SAH is
likely to be different among young healthy people and older
people with chronic conditions and disability. Parenthetically,
we might also note that this robust evidence about SAH and
mortality is difficult to translate into intervention strategies.

In this commentary we have tried to identify a number of
current and methodological challenges that we feel are
common in studies in psychosocial epidemiology. Then we used
some of the articles assembled in this issue to illustrate these
challenges. Ultimately, we hope this will lead to a better under-
standing of psychosocial epidemiology by those who are not
active participants in this research domain.

References

1 Ziefhuis GA, Kiemenej LALM. Social Epidemiology? No way. Int J
Epidemiol 2001; 30:43–44.
2 Krieger N. Commentary: Society, biology and the logic of social
Epidemiol 2001; 30:46–47.
4 McPherson K. Commentary: Epidemiology? Keep it broad and deep.
6 Siegrist J. Commentary: Social epidemiology—a promising field. Int J
Epidemiol 2001; 30:50.
7 THE GREAT DEBATE: Resolved: Psychosocial interventions can
64:549–70.
9 Kasl SV, Jones BA. Some methodological considerations in the study
of psychosocial influences on health. In: Vingerhoets A (ed.). Assessment in
10 Kasl SV, Jones BA. An epidemiological perspective on research design,
measurement, and surveillance strategies. In: Quick JC, Tetrick LE
(eds). Handbook of Occupational Health Psychology. Washington, DC: