Psychosocial factors and public health: a suitable case for treatment?

J Macleod, G Davey Smith

Adverse psychosocial exposure or “misery” is associated with physical disease. This association may not be causal. Rather it may reflect issues of reverse causation, reporting bias, and confounding by aspects of the material environment typically associated with misery. A non-causal relation will not form the basis of effective public health interventions. This may be why psychosocial interventions have, so far, showed little effect on objective physical health outcomes. This paper reviews evidence for the “psychosocial hypothesis” and suggests strategies for clarifying these issues. It concludes that, although misery is clearly a bad thing as it erodes people’s quality of life, there is little evidence that psychosocial factors cause physical disease. In the absence of better evidence, suggestions that psychosocial interventions are needed to improve population physical health, in both absolute and relative terms, seem premature.

“Psychosocial” factors such as stress, hostility, depression, hopelessness, and job control seem associated with physical health—particularly heart disease. Adverse risk profiles in terms of psychosocial factors seem to cluster with general social disadvantage. Because of this the “psychosocial hypothesis” proposes that psychosocial factors are an important cause of health inequalities. Such factors include many mental states, psychological traits, or aspects of the social environment with a negative connotation. In this article we consider “psychosocial factors” to be any exposure that may influence a physical health outcome through a psychological mechanism. We wish to critically examine the evidence that the association between psychosocial factors and physical health is causal. If this association is not causal, interventions targeting psychosocial exposures are unlikely to lead to improvements in population health.

Consensus has arisen that evidence robust enough to form the basis of health policy should normally derive from randomised controlled trials of interventions modifying the exposure in question and assessing the effect of such manipulation on an objective health outcome. There are situations where such an approach is either unfeasible, or is unlikely to be adopted for other reasons. The practical feasibility of taking an experimental approach to questions of causation in psychosocial epidemiology has been demonstrated, though most evidence on the effects of psychosocial exposures is observational. We will consider all kinds of evidence relevant to the question posed in our title.

PSYCHOSOCIAL VERSUS MATERIAL EXPLANATIONS: IMPLICATIONS FOR INTERVENTIONS

The question of whether or not psychosocial factors, in themselves, determine physical health outcomes, and social gradients in these outcomes, has often been considered in the context of alternative explanations. “Social class” can be defined in several ways. Arguably, the fundamental attribute differentiating different “class” categories in all classification schemes relates to differences in the power to access material resources. Material factors therefore seem an obvious candidate for consideration as an explanation of social health inequalities. Some commentators have dismissed the possibility of an important causal role for material factors based on an argument that among people who are not, conventionally, materially deprived, mechanisms through which such factors might directly influence health are unknown. Such arguments, relate to the question of plausibility. We disagree that material conditions are an implausible cause of physical disease and have discussed possible mechanisms elsewhere. However, our focus in this article is evidence for the psychosocial hypothesis rather than that supporting competing hypotheses. Plausible pathways from “stress” to disease exist. We also note, as has been argued elsewhere and as Bradford Hill himself remarked in his original essay on causation, plausibility is a rather weak criterion for establishing causality as it depends on prevailing biomedical understanding and, to some extent, fashion.

It has also been argued that the dichotomy between “psychosocial” and “material” causal theories is largely false, as most material phenomena have social meanings. Nevertheless, subscription to one or the other of these explanatory models has important implications for the type of interventions that might be considered worth evaluating. Marmot and Wilkinson, two leading proponents of the psychosocial hypothesis have put it thus:

“If, in the spirit of neo-materialism, you give every child access to a computer and every family a car, deal with air pollution, and provide a physically safe environment, is the problem solved? We believe not.”
The spirit of neo-materialism would not necessarily lead to a public health policy based around the distribution of cars and computers. Nevertheless, psychosocial theorists believe that focusing on the distribution of material resources is unlikely to improve health. In their view it is not relative poverty, in itself, that is health damaging, it is how poorer people feel about their poverty. Thus, logically, it is feelings that interventions informed by a belief in the psychosocial hypothesis should target. However, we doubt whether psychological interventions to increase the feelings of control or reduce the feelings of hopelessness (for example) that people have without changing the actual contingencies of their lives (such as whether they can afford to choose to run a car or buy their children a computer) will solve the problem either. This is not an ideological position; we are sceptical because we do not think available evidence suggests a directly causal relation between psychosocial factors and health.

HOW MIGHT PSYCHOSOCIAL FACTORS INFLUENCE HEALTH? CONTINGENT AND FUNDAMENTAL CAUSES

Our preceding remarks on plausibility notwithstanding, it is worth considering how “psychosocial adversity” might cause physical disease. It could foster unhealthy behaviour—smoking, for example—or it could lead directly to neuroendocrine perturbations that influence disease risk. At this point we should introduce a distinction between contingent and fundamental causes (elsewhere referred to as “probabilistic” and “sufficient” causes respectively). The relation of a contingent cause to a health outcome depends on the former’s link to a fundamental cause. Thus psychosocial adversity could cause increased disease risk in situations where psychosocial adversity was positively associated with smoking. Such associations are not at all automatic. In the 1950s most male doctors were smokers. In this situation, it could hardly be argued that smoking was determined by adverse psychosocial profiles among smokers, or indeed by social disadvantage. Then, as now, the determinants of such recreational drug use were complex. And, importantly, in places and at times when smoking was more prevalent among the socially advantaged the consequences of smoking showed a similar social distribution. For example a Dutch paper of 1948 noted a “disproportionate number of intellectuals amongst the lung cancer patients” The fact that smoking seems associated with lung cancer, whatever the social distribution of smoking, is good evidence both that this association is causal and that reducing smoking should reduce lung cancer risk.

RELATIVE IMPORTANCE OF PSYCHOSOCIAL AND MATERIAL INFLUENCES ON PHYSICAL HEALTH

In most populations, today, there seems a direct relation between material disadvantage and the various indices of adverse psychosocial exposure, which together could be characterised as “misery”. Consequently, both misery and material disadvantage seem associated with poor health. How secure are these associations? We have previously discussed the influence of reporting bias on associations between psychosocial exposures and illness and how such bias may have influenced the findings of published evidence—particularly that relating job control to angina. In simple terms, people who feel miserable (who report greater stress, lower job control, greater hostility, and so forth) may also feel sicker—without their being any objective evidence that they experience any greater amount of physical disease. We suspect that the general importance of this phenomenon has been underestimated in observational epidemiology. There has been extensive discussion of the cognitive processes involved in survey response and of the specific issue of “negative affectivity”. However, no reliable means have been found to avoid the problem in situations where substantially subjective exposures are related to substantially subjective outcomes. Indeed we recently demonstrated the shortcomings of the commonest strategies—adjustment for a measure of “complaint tendency” or exclusion of prevalent cases at the start of follow up studies. For these reasons it seems surprising that leading biomedical journals continue to publish reports of associations between substantially subjective exposures (such as most psychosocial exposures) and substantially subjective outcomes, generally without acknowledgement of the great difficulty attached to interpreting these relations as evidence of substantive effects. Thus part of the association between misery and health could, in fact, reflect an artefact (this is less of an issue with regard to material circumstances and health as material circumstances are more amenable to objective measurement).

However, there is good evidence that both material disadvantage and misery are associated with objective health outcomes. Such associations are unlikely to arise through reporting bias. It therefore seems that both material disadvantage and misery are genuinely associated with poorer health. But would both misery and material disadvantage affect physical health irrespective of their relation to each other? The association between material disadvantage and illness may arise, predominantly, because material disadvantage is miserable (the psychosocial hypothesis). Alternatively, the relation between misery and illness may, largely, reflect the fact that misery is now a marker for material disadvantage.

PSYCHOSOCIAL OR MATERIAL ADVERSITY AND OBJECTIVE DISEASE—which “CONFOUNDS” WHICH?

Associations between psychosocial exposures and physical health outcomes seem to persist (though are often attenuated) after adjustment for behavioural factors and adjustment for measures of material disadvantage. This is not necessarily strong evidence for their fundamental causal role because the results of statistical adjustment in multivariate models substantially depend on the precision of measurement of correlated covariates. The important correlated covariates in this context are the measures both of misery and of material disadvantage—in both cases measures are likely to provide considerably imprecise indices of the exposure of interest at the time of measurement. Add to this the dimension that it may be adverse exposure (either psychosocial or material) at another point of the life trajectory that is the key cause of increased disease risk and one can see that residual confounding is virtually inevitable. Consequently it is extremely difficult to declare that either psychosocial or material factors are “independent” causes of illness.

Confounding of psychosocial exposure-disease outcome relations by material factors related to social position is only
likely in the situation where adverse psychosocial exposure is closely correlated with material disadvantage and where disease risk is higher among the disadvantaged. Study of these relations in populations where this is not true can, therefore, give important clues regarding confounding. For example, we studied perceived psychological stress in a population where higher stress was associated with material advantage rather than disadvantage (such populations were probably not uncommon before the spread of “victim culture” among the working class). Higher stress was also associated with unhealthy behaviour (higher smoking and drinking and lower exercise), making it a promising candidate as a contingent cause of disease. Thus, in many ways, this population seemed an ideal setting to study the relative importance of material compared with psychosocial factors to health and the independence of both of these putative causes of disease. Higher stress (despite its association with unhealthy behaviour) was associated with lower risk of heart disease, whereas material disadvantage was associated with higher risk of heart disease. Most other classes of disease followed this pattern; strongly suggesting the independence and primacy of material factors in determining disease risk in this population. In other words these results suggested that material circumstances confound the association between psychosocial factors and health—rather than stress mediating the association between material disadvantage and health.

Today, it may be difficult to find other populations where psychosocial and material disadvantage are not closely linked and corroborations of our findings could be sought—alternative strategies are therefore needed. One approach entails the use of sensitivity analyses. These can help judgement of the plausibility of confounding by an unmeasured factor as an explanation for an observed effect. For example, consider a relation between hopelessness and disease apparent after adjustment for adult occupational class. This relation may actually be mediated through a third unmeasured aspect of the material environment related to both hopelessness and occupational class. However, the association of this hypothesised factor to study measures of hopelessness must be stronger than its association with study measures of occupational class—sensitivity analyses indicate how much stronger, for any given effect estimate. Given the relatively crude nature of most “social position” indicators (a dichotomous adult occupational class distinction is typical) it seems plausible that in this situation a psychosocial variable indexing some aspect of misery might have a closer association to true life course material circumstances than a simple adult occupational classification.

Sensitivity analyses entail comparatively simple computations; some have suggested more mathematically complex analytic solutions to the problem of drawing causal inferences from observational data. In particular, structural equation modelling (SEM), an elaboration of path analysis, has been advanced as something of a panacea in this regard. We are not disputing that SEM has theoretical strengths that may well lead to its wider use in observational epidemiology. However, we have a general scepticism regarding the extent to which analytic sophistication can resolve the issue of confounding in observational data. For example, consider the history of evidence relating to the effect of hormone replacement therapy (HRT) on cardiovascular risk. HRT was promoted, partly, on the premise that it would reduce risk of heart disease. Such an effect had been demonstrated in numerous prospective observational studies. A meta-analysis of these studies went so far as to suggest that the summary estimate of protective effect found (0.5 95% CI (0.43 to 0.56)) was, “Unlikely to be explained by confounding factors or selection”. This assessment was apparently based on the size of the effect, the rigour of the studies contributing to the analysis, and the analytical sophistication that had been brought to bear to minimise the influence of confounding. Subsequent observational studies involving even greater rigour and analytical sophistication came to similar conclusions. Yet when experimental evidence eventually became available it showed no reduced cardiovascular risk among women taking HRT, in fact a possible increased risk was suggested.

The observational evidence was so misleading because HRT use is markedly socially patterned such that associations between HRT use and health are confounded by other exposures related to social position. The influence of such confounding is so strong that residual confounding is almost inevitable—even after attempts to adjust for this through sophisticated analysis. These pitfalls appear at least as likely to apply to associations between psychosocial exposures and heart disease in observational studies. HRT use is a relatively straightforward exposure to measure precisely and objectively and is surely no more unusual in its social patterning than most psychosocial factors.

The HRT “cautionary tale” also illustrates another criterion that can help distinguish causal from confounded associations in observational epidemiology, specificity of association. In fact, based on this criterion, Pettiti and colleagues advised against taking the results of their own observational study into HRT use and heart disease at face value. They too found that HRT use seemed to approximately halve the risk of heart disease, however it showed a similar effect on mortality from accidents, homicide, and suicide. They suggested (in retrospect correctly) that this finding was highly suggestive of residual confounding and that experimental studies were needed to resolve the issue. The criterion of specificity could also usefully be applied to observed associations between psychosocial exposures and physical disease. The neuroendocrine mediated psychosocial hypothesis predicts a particular association with risk of cardiovascular disease. An observed association that seems specific to increased risk of cardiovascular disease is more likely to be causal than one that is apparent in relation to most of the classes of mortality that are associated with disadvantage.

To permit such an assessment, psychosocial epidemiological studies reporting risks of cardiovascular mortality should also report risks of other mortality categories. Few studies report mortality in this detail. For those that do the results are interesting. For example, in the Kuopio Ischaemic Heart Disease Study (probably the largest and most detailed study of psychosocial factors and health undertaken), increased hopelessness predicted death from heart disease, though also from cancer and from accidents and violence. These results suggest that the relation between hopelessness and heart disease is unlikely to be directly causal but rather reflects confounding by some other factor. In this situation, intervening to reduce hopelessness—if this were possible—would be unlikely to reduce heart disease.

Genetics may provide another solution to the problem of confounding in some situations. If levels of an exposure are determined by a particular polymorphism in the population, and outcome is unlikely to be confounded by social or behavioural factors as genotype is effectively randomly assigned (so called “Mendelian randomisation”), this approach may seem of limited relevance in psychosocial epidemiology, however it can still provide important insights. For example, it has been suggested that plasma fibrinogen may lie on the causal pathway between psychosocial factors and cardiovascular risk. However, the status of fibrinogen as a cardiovascular risk factor remains controversial. Plasma fibrinogen concentrations are related to a polymorphism in the β-fibrinogen gene, presence of the “A” allele being associated with higher levels. Among controls of a recent large case-control study, fibrinogen increased by...
Similar findings were recently reported from the SHEEP association between fibrinogen and CHD risk are not causal. To 1.10. People whose genotype would have subjected them to causal then a similar per allele relative risk of CHD should be of CHD of 1.20 (95% CI 1.13 to 1.26). If this relation were truly a 0.12 g/l rise in fibrinogen was associated with a relative risk in SHEEP, it seems likely that they would be able to generate a similarly spurious relation between work stress and vascular risk in the same study. Furthermore, if these associations between fibrinogen and CHD risk are not causal. Similar findings were recently reported from the SHEEP study. \(^5\) Ironically, the SHEEP investigators were one group who had previously suggested a causal association between work stress and CHD risk, mediated by fibrinogen. \(^6\) Clearly this now seems unlikely and it is more plausible that the association between psychosocial factors and fibrinogen seen in SHEEP arose through confounding by the same factors that confounded the association between fibrinogen and cardiovascular risk in the same study. Furthermore, if these unmeasured confounding factors were able to generate a spurious and misleading relation between fibrinogen and CHD risk in SHEEP, it seems likely that they would be able to generate a similarly spurious relation between work stress and CHD risk.

THE EFFECTIVENESS OF PSYCHOSOCIAL INTERVENTIONS—EXPERIMENTAL EVIDENCE

A powerful strategy to minimise the possibility of confounding is random allocation of exposure level within an experimental study. In this way any confounding factors (measured or unmeasured) should be evenly distributed across the different levels of exposure such that any effect seen is truly that of the exposure. To permit this approach exposure level must be modifiable through an intervention amenable to random allocation. Arguably, this provides the strongest and most practically relevant evidence on causality because a positive treatment effect demonstrates both the existence of a causal relation and the effectiveness of an intervention based on this.

This approach may represent one way forward in psychosocial epidemiology. \(^7\) However, based on currently available evidence, we are sceptical that counselling people to “cheer up” or “relax” or “take more control” without changing their access to material resources will improve their physical health.

The number of experimental studies of the effects of psychosocial intervention on objective measures of disease is comparatively small. Most have assessed effects on prognosis among people with established heart disease. The factors determining prognosis in people with heart disease may not be the same, or may not have the same relative importance as those determining disease development. Nevertheless, this evidence is still useful as an indicator of the potential of psychosocial interventions to improve population cardiovascular health. In most published examples “stress reduction” interventions were delivered as part of an intervention package targeting multiple risk factors such as smoking, diet, and exercise and aiming to improve case management. Reviews of these studies have suggested small but significant effects on prognosis, but have not been able to disaggregate the effects of the psychosocial component of the intervention from that of other components. \(^8\) \(^9\)

A “pure” psychosocial intervention was assessed in the recently completed ENRICHD trial. \(^10\) Depression is perhaps the psychosocial factor with the strongest candidate for a causal relation with heart disease. \(^11\) Because of this, ENRICHD assessed the effect of treating depression on prognosis of established heart disease. The intervention was effective in reducing depression, but heart disease prognosis was the same among controls as in the intervention group. Indeed, lead investigators on ENRICHD have acknowledged that the association between depression and heart disease may not be causal and have emphasised that the principal justification for treating depression is improved quality of life, rather than reduction in mortality. \(^12\)

Patel and colleagues evaluated the effects of relaxation therapy among “high risk” subjects without clinical heart disease at study recruitment. \(^13\) One (of 81) control subjects died from heart disease compared with none (of 88) intervention subjects during four year follow up. The authors, not unreasonably, suggested that, “If the results of this study could be obtained in a larger study the financial and health care implications would be enormous.”

It seems significant that almost 20 years later nobody has reported replication of these findings in a larger study.

OTHER PSYCHOSOCIAL INTERVENTION STRATEGIES

Some leading proponents of the psychosocial hypothesis have emphasised that, in relation to the particular case of work stress, “mass psychotherapy” is not the intervention strategy they would propose. \(^14\) Instead they propose modification of the structure of work to give workers more control and thus reduce their stress experience. This approach has been most rigorously evaluated in Scandinavia with results that some have interpreted as promising. \(^15\) However, a recent systematic review concluded that interventions to reduce work stress focused on organisations and principally intended to increase “job control” were generally ineffective. \(^16\) Worker autonomy is a phenomenon that is intrinsically linked to workplace hierarchies of power, accountability, and reward. In other contexts it is used as an indicator of general social position, for this reason. \(^17\) Changing worker autonomy, in anything but extremely superficial ways seems impossible, without fundamental changes in the structure of work, how it is organised and how it is remunerated. Perhaps this is why experimental studies so far have seen such modest results. Rolling out of such interventions—delivered by committed and enthusiastic investigators in the setting of sympathetic organisations—into the world of work generally seems unlikely to be an effective public health strategy.

PSYCHOSOCIAL OR MATERIAL SOLUTIONS

This lack of evidence supporting the psychosocial hypothesis is not, of itself, strong evidence for an alternative material explanation. Indeed, though we feel that observational evidence strongly supports the primacy of material circumstances in determining health inequalities, there is comparatively little experimental evidence on the effectiveness of material interventions. \(^18\) Given the political sensitivity of
wealth redistribution as a concept this is perhaps not surpris-
ing. Ingenious strategies to provide experimental evidence on the health effects of material factors have been suggested. The proposed follow up of National Lottery winners is not completely unproblematic in relation to practical and scientific considerations, but it would be an interesting exercise. It has been suggested that policy informs the evidence base more than vice versa and we suspect that, realistically, this is unlikely to change. Psychosocial solutions may seem attractive to some policy makers—because they permit location of responsibility for health at the level of the individual and their unhealthy feelings. Unlike material solutions, psychosocial solutions do not necessitate fundamental social change—some have suggested that this is a point in their favour.

Fundamental social change may seem unlikely to many people and insistence on a public health strategy that demands it could be dismissed as self indulgent or, at best, unrealistic. We support realistic strategies to improve population health in both relative and absolute terms. We also think that another consideration is important. The legitimacy of psychological distress as a health problem is not dependent on the strength of the evidence for a causal relation between psychological distress and somatic disease. Indeed in many wealthy countries with long—and lengthening—life expectancies, the burden of such distress may be of more importance than much somatic disease. Whether, in addition to being an important cause of physical disease, particularly heart disease. Our interpretation of this evidence is that currently, it provides little support for a direct causal relation and no basis to propose psychosocial interventions as a public health strategy to improve health in general or to reduce health inequalities. We look forward to better evidence that might change our mind. Similarly we hope that the adoption of more redistributive fiscal policies may permit us greater insights into the direct effects of material factors on physical health.

ACKNOWLEDGEMENTS

Thanks to Nancy Kreiger and John Lynch for comments on a draft of this paper.

References


66 Walker I. If only lottery winnings were randomly assigned. BMJ 2002;324:979.

THE JECH GALLERY

Neurocysticercosis: households, pigs, and risks

Neurocysticercosis is a major cause of neurological disease in most developing countries and is an emerging disease in industrialised areas. My intention is not to provide a detailed explanation of this problem, but to give an image to the paragraph of a recent review on the topic: “ . . . Many rural households keep pigs in small numbers; the animals constitute an important source not only of meat but also of immediate income. The rearing of free-ranging animals requires little investment for the rural poor. In the absence of a sanitary infrastructure, people use open areas and fields for defecation. Free-ranging pigs thus have access to human faeces, which perpetuates transmission of the parasite from human being to pig. Rural pork producers are not motivated to pass pork through meat inspection because of the threat of condemnation. Furthermore, local culinary habits facilitate the consumption of raw or partly cooked meat. These factors lead to transmission of the cysticercus from pig to human being in endemic areas.”

J Jaime Miranda
International Health and Medical Education Centre, University College London, London N19 5LW, UK;
j.miranda@ucl.ac.uk

Reference