

Risk factors and prevention

Risk factors play a central part in prediction and prevention. A critical issue in any discussion of risk factors is to ensure that the term risk factor, and associated terms such as correlate and marker, are defined in a precise, consistent manner.¹ The starting point is to understand that a *correlate* is a variable that is associated, either positively or negatively, with an outcome. The presence or absence of a correlate can be measured in each subject. The term *subject* need not be an individual but could be a family, classroom, school, or an entire community. The outcomes can be dimensional, but they are restricted in this discussion to binary ones. Binary outcomes are the most relevant in medical practice. The physician and patient are concerned about whether the patient has or does not have the illness or disorder. The correlate can be measured at the same time as the outcome and thus be a concomitant of it, or it can be measured after the outcome and be a consequence or result of it. A *risk factor* can be considered a type of correlate. It is associated with an increased probability of an outcome, usually an unpleasant one. It has, however, a major distinguishing characteristic from other correlates, which is it occurs before the outcome. This is 1 of the 2 defining characteristics of a risk factor. The measure of the risk factor is taken on each subject before the subject has the outcome of interest. The second defining characteristic of a risk factor is that it can be used to divide a population into high risk and low risk subgroups. The probability of the outcome must be shown to be greater in the high risk compared with the low risk group. Thus, the 2 defining characteristics of a risk factor are that it precedes the outcome, and when used it divides a population into high risk and low risk subgroups.

Types of risk factors

There are 3 different types of risk factors that must be distinguished from each other in planning prevention initiatives. The first type is a risk factor that cannot be shown to change, and this is termed a *fixed marker*. Examples of fixed markers are traits such as sex, ethnicity, and date of birth. It should be noted, however, that although sex is a fixed marker, the mechanisms by which sex has its effect on a particular outcome may qualify as risk factors. When a risk factor can be shown to change spontaneously within a subject, or be changed as a result of an intervention, then this is termed a *variable risk factor*. If the risk factor, when it is manipulated, does not change the risk of the outcome, then this is called a *variable marker*. If the risk factor can be shown to be manipulable and when manipulated changes the probability of the outcome, then this is termed a *causal risk factor*. *Cause* as used here does not imply that the variable is the only cause of the outcome, nor does it deal with the pathways through which the causal risk factor might have its effect. It means only that manipulating the variable results in a reduction in the incidence of the outcome. It is important then to distinguish among a non-correlate, a correlate that is a concomitant or consequence, a correlate that is a risk factor, and among risk factors those that are fixed markers, variable markers, and causal risk factors.

Four other points should be mentioned. Firstly, risk factors for a particular outcome cannot be assumed to be the same, or to be of the same strength, in different populations. They may indeed be population specific, and this becomes an important issue in

multisite studies. Secondly, risk factors for the onset of disorder may not be the same as those for remission or for relapse. More specifically, risk factors for the onset of disorder in the general population may not be the same as those for remission or for relapse in the subgroup of the population with the disorder, and may differ again from those involved in relapse of the disorder in the subgroup of those who remit. In the first instance, the risk factors would relate most closely to prevention, in the second to treatment, and in the third to maintenance. The third point about risk factors is that they can only be measured in populations where there is variability in the frequency of both the risk factor and the outcome.² For example, if everyone in a population consumes the same amount of salt in their diet, then the identification of salt as a risk factor for hypertension in this population will be impossible. Similarly, if all children in a population qualify for a diagnosis of conduct disorder, then, in this instance poor parenting practices could not be identified as a risk factor for conduct disorder. The fourth and last point is that the discussion has focused on risk factors for unwelcome outcomes. All of the comments thus far, however, apply equally well to factors, called *protective factors*, that increase the probability of welcome outcomes.

Using antisocial behaviour or conduct disorder in childhood as an example, one is able to distinguish among the different types of variables.³ Being a blue-eyed child, for example, is not a correlate of conduct disorder. Having difficulty with peer relationships is usually a concomitant of conduct disorder, and getting into trouble with the law can be seen as being a consequence of conduct disorder. Moving on to risk factors, male sex is a fixed marker for conduct disorder. The incidence of conduct disorder is higher in boys than girls. Income level can be manipulated but there is no strong evidence as yet (eg, from randomised controlled trials) that manipulating income level results in reduced risk of the onset of conduct disorder. Income then can be termed at this time as a variable marker for conduct disorder. Strong evidence exists, however, that not only can parenting practices be changed, but when they are improved, they result in a reduced incidence of conduct disorder.⁴ Poor parenting practices then fulfill the criteria of a causal risk factor for conduct disorder. It should be kept in mind that a characteristic that is currently a variable marker may, with new evidence indicating that manipulation of the marker results in a reduction in the incidence of the outcome, attain the status of a causal risk factor.

The importance of distinguishing among these different terms is evident in prevention studies. If the intervention is to result in a reduction in the incidence of the outcome of interest, it must focus on a causal risk factor. Changing the frequency of a variable risk marker may result in a number of benefits, but 1 of them will not be reducing the incidence of the outcome of interest. The same holds true for a correlate. Fixed markers, of course, cannot be changed but they can be used to identify high risk populations for prevention studies.

Potency of risk factors

A second critical issue in any discussion of risk factors is the potency or strength of the risk factor. The potency of a risk factor can be defined as the maximal discrepancy achievable using

the risk factor to dichotomise the population into high and low risk groups.¹ A detailed discussion of the potency of risk factors is available.⁵ The present write up focuses on several issues that are particularly relevant to clinicians, policy makers, and researchers.

There are many different measures of association arising out of the 2x2 table. The definitions tend to vary by the field of investigation. For example, epidemiology commonly uses odds ratio and attributable risk, whereas psychology and sociology lean towards phi coefficient and gamma, respectively. Each of these measures, and others, usually carries with it implicit trade offs between false positives and false negatives. A major contribution of the article by Kraemer *et al*⁵ is to make these trade offs explicit. It turns out then that no single measure of the potency of a risk factor will always be the right one, and no measure will always be the wrong one.

It is important to be reminded that the strength of association between a risk factor and an outcome can be attenuated because of the unreliability of the measure of either variable. It may be, for example, that the potency of a causal risk factor is found to be less than that of a correlate, and the reason could be that it is more difficult to measure reliably the causal risk factor than it is to measure reliably the correlate.

To show that a variable is a risk factor for an outcome, it is not only necessary to show that it precedes the outcome but also that there is a statistically significant relation between the risk factor and the outcome. However, statistical significance is not sufficient to show that a risk factor is of clinical or policy importance. Statistical significance is dependent on sample size, and, with a large sample, statistical significance is easier to show, but the strength of the relation between a risk factor and an outcome in this case may be of no practical use. Consider, for example, data from the Dunedin Multidisciplinary Health and Development Study.⁶ This study is a longitudinal investigation of a complete cohort born between 1 April, 1972 and 31 March, 1973 in Dunedin, a city on New Zealand's South Island. The original sample size was 1073 and attrition rates have been low through multiple assessments in childhood and adolescence. It was reported that a statistically significant relation existed between the risk factor based on data at ages 3 and 5 years and the outcome of stable and pervasive antisocial behaviour at age 11 years. The positive predictive value was 15% and the sensitivity 64%, however. Thus, of all children identified as having the risk factor at age 3 years, only 15% ended up having the outcome at age 11 years. Furthermore, of the children qualifying for the outcome at age 11 years, over one third (36%) were identified as not having the risk factor at ages 3 and 5 years. So, although the strength of the relation between the risk factor and the outcome is statistically significant, the results indicated that the risk factor was of no practical importance for identifying a high risk group for the outcome in question.

A related issue in determining the clinical or policy relevance of the potency of a risk factor is a consideration of the relative importance of false positives and false negatives. What, for example, is the potential harm of falsely identifying a person at high risk for an outcome? In Huntington's disease, for instance, one would imagine that a false positive identification for this outcome could adversely affect the quality of life of that individual. Conversely, a false negative error could also be dam-

aging for people at risk for Huntington's disease in that they would be robbed of the opportunity to plan rationally for their future and that of their family. Thus, in any determination of the potency of a risk factor, it is always necessary to decide on the relative importance of false positives and false negatives. The threshold arrived at for the presence of the risk factor will relate directly to the results of the measure of potency.

Returning to the topic of risk factors and prevention studies, it is not only essential to focus prevention efforts on altering causal risk factors, but it is also important to know how the potency of the causal risk factor will play itself out in real life consequences. An example of a relevant measure in this regard is attributable risk. It provides an estimate of morbidity attributable to a risk factor.⁷ It will estimate the maximum that the outcome could be reduced if the risk factor were eliminated. In prevention studies, it is important to select causal risk factors with a high attributable risk so that success in reducing or eliminating the effects of the causal risk factor will result in a clinically meaningful reduction in the outcome of interest.

Conclusion

In summary, the study of risk factors will benefit from a clear definition of terms where there is an unambiguous distinction among correlates, fixed markers, variable markers, and causal risk factors. Measures of the potency of risk factors are many, and showing that the relation between a risk factor and an outcome is statistically significant does not guarantee that this relation will be of value in the clinical and policy domains. A major consideration in determining the potency of a risk factor is the relative importance of false positives and false negatives and the resulting implications for setting a threshold for the presence of the risk factor. Lastly, preventive studies should focus not only on altering causal risk factors but centre also on those causal risk factors for which the elimination of their deleterious effects can be expected to produce a clinically meaningful reduction in the outcome of interest.

DAVID R OFFORD, MD

Director, Canadian Centre for Studies of Children at Risk
McMaster University
Hamilton, Ontario Canada

HELENA CHMURA KRAEMER, PhD

Professor of Biostatistics in Psychiatry
Stanford University
Stanford, California, USA

- 1 Kraemer HC, Kazdin AE, Offord DR, et al. Coming to terms with the terms of risk. *Arch Gen Psychiatry* 1997;54:337-43.
- 2 Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985;14:32-8.
- 3 Offord DR. Conduct disorder: risk factors and prevention In: Shaffer D, Philips I, Enzer NB, editors. *Prevention of mental disorders, alcohol and other drug use*. Rockville, MD: The Office, 1989: 273-397.
- 4 Tremblay RE, Pagani-Kurtz L, Masse LC, et al. A bimodal preventive intervention for disruptive kindergarten boys: its impact through mid-adolescence. *J Consult Clin Psychol* 1995;63:560-8.
- 5 Kraemer HC, Kazdin AE, Offord DR, et al. Measuring the potency risk factors for clinical or policy significance. *Psychological Methods* 1999;4:257-71.
- 6 White JL, Moffitt TE, Earls F. How early can we tell? Predictors of childhood conduct disorder and adolescent delinquency. *Criminology* 1990;28:507-33.
- 7 Lipman EL, Offord DR, Boyle MH. What if we could eliminate child poverty? The theoretical effect on child psychosocial morbidity. *Social Psychiatry and Psychiatric Epidemiology* 1996; 31:303-7.