 Criminal Behaviour and Mental Health, 10, 100–122 2000 © Whurr Publishers Ltd

Some benefits of dichotomization in psychiatric and criminological research

DAVID P. FARRINGTON[1] and ROLF LOEBER[2] [1]Institute of Criminology, University of Cambridge; [2]Western Psychiatric Institute and Clinic, University of Pittsburgh, Pittsburgh, PA USA and Free University, Amsterdam, The Netherlands

ABSTRACT

Background *The product-moment correlation* r *is widely used in criminology and psychiatry to measure strength of association. However, most criminological and psychiatric variables contravene its underlying assumptions.*

Aim *To compare statistical measures of association based on dichotomous variables with the use of* r.

Method *Explanatory variables for delinquency are investigated in the Pittsburgh Youth Study using a sample of 506 boys aged 13–14.*

Results *Dichotomization does not necessarily cause a decrease in measured strength of associations. Conclusions about the most important explanatory variables for delinquency were not greatly affected by using dichotomous as opposed to continuous variables, by different dichotomization splits, or by using logistic versus OLS multiple regression. Non-linear relationships, interaction effects and multiple risk factor individuals were easily studied using dichotomous data.*

Conclusions *Dichotomization produces meaningful findings that are easily understandable to a wide audience. Measures of association for dichotomous variables, such as the odds ratio, have many advantages and are often more realistic and meaningful measures of strength of relationship than the product-moment correlation* r.

Problems of analysis

This paper is primarily concerned with the problem of measuring associations between explanatory and outcome variables in criminological and psychiatric studies. In examples, we particularly focus on the outcome variable of delinquency and on individual, family and socioeconomic explanatory variables. Our arguments, however, apply more generally to many other types of studies in the social and behavioural sciences.

The most popular method of measuring associations is to use the product-moment correlation *r*. This paper is particularly addressed to criminologists and psychiatrists who use statistical techniques based on *r*. In order to interpret any particular value of *r* (in order to estimate its confidence intervals or test its statistical significance), certain assumptions about the underlying nature of the variables have to be satisfied. The most important of these are that both variables have to be normally distributed and measured on interval scales, and that they have to be linearly related to each other.

Unfortunately, most variables that are measured in criminology and psychiatry are not of this type. An equal-interval scale requires that differences between values mean the same at all points on the scale. For example, a difference in age of 6 months should mean the same at all ages. Some variables (e.g. test scores) are measured only on ordinal scales. Others (e.g. ratings) are measured in a small number of (nominal or ordinal) categories. Some variables are inherently dichotomous (e.g. sex or whether a person has been convicted or not). Psychiatry, medicine and law are often more interested in types or categories of individuals than in scale scores, because of their concern with treatment or judicial sanctions. For example, psychiatrists are more interested in diagnosing psychopathic disorder or schizophrenia than in knowing scores on symptom checklists for psychopathy or schizophrenia.

Some variables have highly skewed distributions; for example, there may be a large number of basically law-abiding individuals at one end of a distribution and a long tail of chronic offenders at the other. Some variables have non-linear relationships; for example, very young or very old mothers may be at most risk of producing delinquent children. Some relationships are dramatically affected by a few outliers; *r* is particularly vulnerable to these. Some variables might have interactive effects; for example, the relationship between broken families and delinquency may depend on whether the lost parent is a criminal or not.

This last statement draws attention to the fact that the problems reviewed here apply not only in studying associations between two variables but also in the many types of multivariate analyses based on *r* (e.g. factor analysis or least-squares multiple regression). A popular solution to these problems is to argue that *r* is so robust that violations of its underlying assumptions are essentially unimportant (Fowler, 1987). However, there are also statistical solutions.

Statistical solutions

Our aim is not to give a detailed exposition of statistical solutions to the problems raised above. Instead, the aim of this paper is to suggest that the dichotomization of variables, and the use of statistical techniques based on dichotomous variables, might in some circumstances be preferable to the use of statistical techniques based on *r*. We will, however, mention some common statistical solutions to the problems of *r* (for more details, see standard textbooks such as Cohen and Cohen, 1983.)

One obvious solution to the problem of the nature of variables is to create normally distributed variables that look as though they are measured on interval scales. An easy way of achieving this is to combine a number of variables into a scale. For example, Fergusson and Lynskey (1995) developed a composite scale of 'family adversity' from numerous measures of different aspects of family functioning and child-rearing practices. In general, irrespective of the average inter-item correlation, the more items included in the scale, the higher will be conventional measures of reliability such as alpha. Another method of creating normally distributed variables is to use appropriate statistical transformations. For example, where a variable is highly skewed, its logarithm might be used in analyses.

Non-linear relationships can be tackled in a variety of ways, for example by including higher order terms in regression equations or using more complex regression models (Greenland, 1995). They can be detected by inspecting scatterplots. It is possible to use generalized linear models designed for variables that are not normally distributed and not linearly related. Similarly, interaction effects can be studied by including multiplicative terms in regression equations. Outliers can be detected in a variety of ways (e.g. by inspecting scatterplots) and can be transformed or eliminated.

All of the large number of statistical solutions have disadvantages as well as advantages. For example, forcing variables to be normally distributed may distort their true nature (e.g. if delinquency truly is a skewed variable). Despite a high value of alpha, a score on a combined scale may not measure one homogeneous construct and may be less meaningful than the constituent variables. A major problem is that, the more complex they are, *statistical solutions make it difficult to communicate results to non-statisticians*. Policy makers, criminal justice practitioners, clinicians and researchers who are not statisticians need to understand key results in criminology and psychiatry and cannot necessarily be expected to take them on trust.

Dichotomization

The main advantage of dichotomization is that it greatly simplifies the presentation of results and produces meaningful findings that are easily understandable to a wide audience. For example, in the Pittsburgh Youth Study (see later), the finding that 43% of boys from broken homes became delinquent, compared with 21% of boys from intact families, is immediately understandable. The 2 x 2 table, combined with ideas of predictive accuracy, false positives and false negatives, sensitivity, specificity (etc.) is particularly attractive to practitioners. Also the dichotomization of explanatory variables facilitates a 'risk factor' approach to the explanation and prediction of delinquency and psychiatric disorders (Kraemer et al., 1997), and encourages the identification of individuals who are particularly vulnerable because they possess several risk factors, and who may be specially targeted in prevention efforts. Often, comorbidity is the rule rather than the exception (Rutter, 1997).

The main disadvantage of dichotomization is that information is lost: all non-delinquents are treated as equivalent, for example, rather than everyone having a delinquency score. The main aim of this paper is to review some advantages and disadvantages of using dichotomous variables to measure associations in criminology and psychiatry. There are key questions to be addressed:

(1) Does dichotomization inevitably lead to a decrease in the measured strength of associations between variables?
(2) How important is the loss of information consequent upon dichotomization?
(3) Does dichotomization lead to different conclusions about the relative importance of explanatory variables?
(4) Does dichotomization make it easy to deal with non-linear relationships? How common are they?
(5) What is the effect of different dichotomization splits?
(6) Does dichotomization lead to different conclusions about which are independently important explanatory variables in multivariate analyses?
(7) Does dichotomization make it easy to detect interaction effects? How common are they?
(8) Does dichotomization help in studying types of individuals and in identifying individuals affected by multiple risk factors?

Decreased strength of association?

Cohen (1983) pointed out that, if two normally distributed, continuous variables have a product-moment correlation r, dichotomizing one of them at the median attenuates the measured (point-biserial) correlation to 80% of r. This reduction in the measured strength of association causes a reduction in statistical power equivalent to discarding about 38% of the cases. Dichotomizing both variables at the median attenuates the product-moment (phi) correlation to 64% of r, causing a reduction in statistical power equivalent to discarding about 60% of the cases. More extreme splits cause even greater reductions in r (and hence in the percentage of variance explained) and in statistical power. This is one of the main objections to dichotomization.

Measures of association other than the product-moment correlation have been devised to compensate for the effects of dichotomization. In particular, consider two normally distributed, linearly related, continuous variables with a product-moment correlation r. If one of these is dichotomized, the population biserial correlation between the continuous and dichotomous variables is still r. Essentially, the biserial correlation adjusts the measured (point-biserial) correlation between a continuous variable and a dichotomous variable to the 'true' (product-moment) correlation on the assumption that the underlying distribution of both variables is normal (Hunter and Schmidt, 1990).

Similarly, while dichotomizing two normally distributed, linearly related, continuous variables at the median reduces the product-moment correlation between them to 64% of r, the population tetrachoric correlation between these dichotomous variables is still r. Essentially, the tetrachoric correlation adjusts the measured (phi) correlation between two dichotomous variables to the 'true' (product-moment) correlation on the assumption that the underlying distributions of both variables are normal. Tetrachoric correlations are widely used in behaviour genetic research (e.g. Pickens et al., 1991; Rowe and Farrington, 1997). They can be calculated using the PRELIS program. They have higher standard errors than phi correlations, but only to the extent that they are greater than phi correlations. Thus, if a tetrachoric correlation is 56% (100/64) greater than the corresponding phi correlation, its standard error is also 56% greater (Hunter and Schmidt, 1990). Converting phi correlations to tetrachoric correlations in general yields exactly the same number of statistically significant results, although sample tetrachoric correlations can sometimes exceed not only r but also 1.0.

As already mentioned, most variables in criminology and psychiatry are not normally distributed and not measured on interval scales. However, even if they were, measures of strength of association would not inevitably be lower after dichotomization if, for example, the tetrachoric rather than the product-moment correlation was used.

Correlations versus odds ratios

While it has been suggested that dichotomized variables cause a decrease in measured strength of associations, our argument on the contrary is that product-moment correlations often give a false impression of weak relationships between variables. If variables are dichotomized and the odds ratio is used as a measure of strength of relationships, this often gives a more realistic impression of strong relationships between variables.

The odds ratio is perhaps the best measure of strength of association between dichotomous variables. It is easily understandable as the increase in the odds (risk) of an outcome associated with a risk factor. Partial odds ratios are important measures of association in the logistic model. Unlike correlation-based measures (Smith, 1996), the odds ratio is independent of the marginal distributions (e.g. prevalence) of explanatory and outcome variables and unaffected by the study design (cross-sectional, retrospective or prospective: see Fleiss, 1981). This was why Edwards (1963) recommended that only the odds ratio and statistics derived from it (from 'the cross-product ratio') should be used as measures of association in 2×2 tables. However, many other measures of association in 2×2 tables have been proposed (Swets, 1986).

These points are illustrated in the hypothetical Table 1. The cross-sectional table shows what is found by drawing a random sample of 900 persons from the population: 100 have risk factor X, of whom 40 (40%) are delinquent, while 800 do not possess the risk factor, of whom 80 (10%) are delinquent. The odds ratio is 6.0 (confidence interval 3.68–9.77) and the phi correlation is 0.28 ($p < 0.0001$). In a prospective study, an investigator might deliberately decide to follow up 450 persons with the risk factor X (of whom 40%, or 180, would be delinquent) and 450 persons not possessing the risk factor (of whom 10%, or 45, would be delinquent). In this design, the odds ratio is still 6.0, but the phi correlation now changes to 0.35. In a retrospective case-control study, an investigator might study 450 delinquents (of whom one-third, or 150, would possess the risk factor X) and 450 non-delinquents (of whom 60/780, or 34.6, would be expected to possess the risk factor X). In this case, the odds ratio is still 6.0, but the phi correlation is now 0.32. These examples show how the phi correlation varies with the design of the study.

Odds-ratio and correlation-based approaches often yield different conclusions about strength of relationships. For example, consider the top table in Figure 1. On the basis of a correlation of 0.28 and only 7.8% of the variance explained, it might be concluded that the risk factor X is a weak predictor of delinquency. On the basis of an odds ratio of 6, and the fact that X is associated with a sixfold increase in the odds (risk) of delinquency, it might be concluded that the risk factor X is a strong predictor of delinquency. Looking at the 2 x 2 table, and contrasting 40% delinquent with 10% delinquent, we believe that the second interpretation is nearer the truth than the first, and

Table 1: Odds ratios and phi correlations with different designs

(a) Cross-sectional

	ND	D	T
NX	720	80	800
X	60	40	100
T	780	120	900

Notes: OR = 6.0, CI 3.68–9.77; phi = 0.28, p <0.0001

(b) Prospective

	ND	D	T
NX	405	45	450
X	270	180	450
T	675	225	900

Notes: OR = 6.0, CI 4.12-8.76; phi = 0.35, p <0.0001

(c) Retrospective

	ND	D	T
NX	415.4	300	715.4
X	34.6	150	184.6
T	450	450	900

Notes: OR = 6.0, CI 3.97–9.15; phi = 0.32, p < 0.0001

Notes: (a–c) X = risk factor, NX = no risk; D = delinquent, ND = nondelinquent; T = total; OR = odds ratio; CI = confidence interval hence that the odds ratio is a more realistic measure of strength of association than the product-moment correlation.

Moving from hypothetical to real data, the top half of Table 2 shows a table from Rosenthal (1990, p. 775). He pointed out that a randomized experiment on the effects of aspirin on reducing heart attacks had been terminated prematurely because the results were so clear-cut that it was considered unethical to continue giving the control group a placebo. However, the tabulated results showed a phi correlation of 0.034, or that only 0.1% of the variance in heart attacks was explainable by the aspirin treatment.

Other measures of strength of association suggest that the decision to terminate was not as perverse as phi might indicate. The odds ratio was 1.83 (CI 1.43–2.35), showing that the odds of a heart attack were 83% greater in the placebo condition; 1.71% in the placebo condition had heart attacks, compared with 0.94% in the aspirin condition. The maximum possible phi correlation in this table is only 0.118. The tetrachoric correlation was 0.143 – more than four times as great as the phi correlation.

Generally, the phi correlation will decrease as the prevalences of the risk factor and delinquency deviate from 50%. The prevalence of heart attacks, for example, was only 1.3%. Also, the phi correlation will decrease the larger the difference between the prevalences of the risk factor and delinquency. However, in both instances, the odds ratio can remain constant. To bring out the implications of these points, consider a study of the correlates of violence for males and females. The prevalence of violence will, in general, be less for females. Imagine that the odds ratio for the comparison of an explanatory vari-

Table 2: Measuring association in illustrative tables

(a) Rosenthal (1990)

	NH	H	T	%H
A	10 933	104	11 037	0.94
NA	10 845	189	11 034	1.71
T	21 778	293	22 071	

Notes: OR = 1.83 (CI 1.43–2.35); phi = 0.034, p <0.0001

(b) Fergusson and Horwood (1995)

	NFO	FO	T	%FO
NCD	701	206	907	22.7
CD	3	26	29	89.7
T	704	232	936	

Notes: OR = 29.5 (CI 8.4–123.5); phi = 0.27, p <0.0001

Notes: (a–b) CD = conduct disorder; FO = frequent offender; A = aspirin; H = heart attack; N = not; T = total; OR = odds ratio; CI = confidence interval able X and violence is *the same* for males and females. In this case, r(phi) and the percentage of variance explained will in general be *less* for females. The use of r rather than the odds ratio might lead to the erroneous conclusion that the correlates of violence are stronger for males than for females. Hence the choice of measure of association can have major implications for the interpretation of scientific results and for decisions pertaining to interventions.

Loss of information

The differential sensitivity of measurement of explanatory variables may make it difficult to draw conclusions about the relative strength of their relationships with an outcome variable. If the product-moment correlation r is used in all cases (as is common), a dichotomously measured variable will appear to be less important than a continuously measured variable (in cases where the underlying variables are truly continuous and linearly related), even if the underlying strength of associations is the same, as Cohen (1983) pointed out. If explanatory or outcome variables are skewed or measured in a small number of categories, the maximum possible r can be considerably less than 1, as in the previous example.

Dichotomization, of course, equates the sensitivity of measurement of all variables (the number of different values). However, it has been argued that the benefits of dichotomization in this respect are outweighed by the disadvantage of losing information about differences between individuals on more continuous variables (neglecting problems of crude measurement and errors in measurement). Unfortunately, loss of information is inevitable in most psychiatric analyses, because the amount of information collected far exceeds the ability of researchers to analyse it and report the results.

A typical study – even a typical one-hour interview – may involve the measurement of hundreds of variables. Inevitably, researchers have to reduce these variables to a small manageable number for analytic purposes, and inevitably information is lost in this process. Often, the sample size and the analytic technique severely limit the number of variables that can be included. It may be hard to determine whether the loss of information consequent upon dichotomization is greater or less than the loss of information consequent upon combining variables or including only a very small subset of variables in a typical analysis (excluding numerous measured variables that might possibly be important). Studying only small subsets of variables encourages numerous piecemeal analyses of large datasets.

Dichotomized variables do not contain inherently less information than scales; it all depends on the relative number of variables of each type and on the accuracy of measurement. For example, four dichotomous variables could contain the same amount of information as one scale with 16 categories, but people may make more errors in classifying individuals into 16 categories on one scale than on four dichotomous variables. Categorical variables can always be converted into a number of dichotomous ('dummy') variables. For example, a four-category outcome variable could be converted into three dichotomous variables, leading to three separate logistic regressions (Zhao and Kolonel, 1992).

In psychiatric research, the dichotomization of variables is relevant to discussions about categorical versus dimensional approaches. Dichotomous diagnoses are often based on the subjective combination of several variables. A key issue is whether psychiatrists should recognize qualitative differences between normality and abnormality (diagnoses), or instead use symptom scores (Caron and Rutter, 1991). In the outstandingly important Christchurch Health and Development Study, Fergusson and Horwood (1995) addressed this issue by comparing the predictive efficiency of categorically and dimensionally scored measures of conduct disorder, oppositional defiant disorder, and attention deficit-hyperactivity disorder. Because dimensionally scored variables were stronger predictors of outcomes than dichotomous measures based on a diagnostic classification, they suggested that the use of (dichotomous) diagnostic categories produced variables with low predictive validity. However, they also pointed out that researchers should investigate the extent to which measurement choices affected results and interpretations, and they have systematically compared results obtained with dichotomous and continuous variables in other analyses (Fergusson et al., 1993, 1997).

The bottom half of Table 2 shows an illustrative table from Fergusson and Horwood (1995, p. 480), relating the presence or absence of conduct disorder to the presence or absence of frequent offending. Note that 26 of 29 conduct-disordered children became frequent offenders. The phi correlation calculated from this table was 0.27. The authors concluded that, since this was less than the (point-biserial) correlation between the conduct disorder symptom score and frequent offending (0.45), the continuous scale had higher predictive validity than the diagnostic category.

Arguably, the phi correlation is not the best measure of association in this table. The great disparity between the row and column marginal totals means that the maximum possible value of phi in this table is not 1 but 0.31. The tetrachoric correlation is an enormous 0.92 (using the correction formula of Hunter and Schmidt, 1990), and the odds ratio is an enormous 29.5 (CI 8.4–123.5). All these figures, and indeed the percentages in the right-hand column (90% versus 23%), indicate a very strong relationship indeed between the conduct disorder diagnosis and frequent offending, casting doubt on Fergusson and Horwood's (1995) conclusions.

The Pittsburgh Youth Study

Data collected in the Pittsburgh Youth Study will be used to investigate the relative importance of explanatory variables. While our illustrations focus on delinquency, our points are widely applicable to much psychiatric, behavioural and social science data. The Pittsburgh Youth Study is a prospective longitudinal survey of the development of delinquency, substance use and mental health problems in three samples of about 500 Pittsburgh boys, totalling 1517 boys (Loeber et al., 1998).

Efforts were made in the Pittsburgh Youth Study to measure a wide range of variables that were alleged to be causes or correlates of delinquency, substance use and mental health problems. In order to maximize the validity of all variables, information from different sources was combined as far as possible, as was information from the screening and first follow-up assessments. By various means, the long list of potential risk factors was reduced to 40 explanatory variables that were thought to cover the key explanatory constructs that were measured.

The illustrative examples in this paper focus on delinquency in the oldest sample of 506 boys, who were aged about 13–14 at the time of the first follow-up assessment. The measure of delinquency seriousness was based on information from boys, mothers and teachers about offences committed during a boy's lifetime up to the follow-up interview. Serious delinquents (36.6%) were those who had committed burglary, car theft, robbery, attacking to hurt, forced sex, or selling drugs; moderate delinquents (26.5%) had committed joy-riding, theft over US$5, carrying weapons, or gang fights; minor delinquents (19.4%) had committed shoplifting, theft of US$5 or less, vandalism, or minor fraud; and non-delinquents (17.6%) had committed no delinquency or only minor stealing at home. We chose to use delinquency categories rather than a combined (continuous) scale (e.g. total number of delinquent acts) because these categories were meaningful and defensible, based on seriousness ratings in large-scale surveys.

As is typically the case in psychiatric and behavioural/social science research, the 40 key explanatory variables in the Pittsburgh Youth Study were of many different types. Not one was measured on an equal-interval scale, normally distributed, and with meaningful values. Two (labelled I in Table 3) were measured on equal-interval scales with meaningful values, but they had skewed distributions with long tails (age of the boy at follow-up, age of the mother at the boy's birth). Six others (labelled II) were measured on continuous scales with reasonably normal distributions (e.g. socioeconomic status), and seven variables (labelled III) were measured on continuous scales with skewed distributions (e.g. poor housing) or with bunching at certain values (e.g. unemployment of the father). Eleven variables (labelled IV) had a fairly large number of values (10 or more) but not enough to qualify as a truly continuous scale (e.g. education of the mother, measured in years of schooling), and seven variables (labelled V) were measured in a smaller number of categories (e.g. family size). The remaining seven variables (labelled VI) were measured as dichotomies.

For some purposes, each of the 40 explanatory variables was dichotomized to contrast the 'worst' quarter of boys (approximately) with the remainder. The one-quarter/three-quarters split was chosen to identify a deviant minority and also to identify approximately the same fraction of boys at risk as those

Table 3: Measured strength of relationships between explanatory variables and delinquency

	PHI	OR	TET
Child:			
IV. Lack of guilt (PT)	0.27	3.5	0.44
I. Old for grade (P)	0.19	2.3	0.31
IV. HIA problems (PT)	0.19	2.7	0.34
III. High ADHD score (P)	0.12	1.8	0.21
II. Low achievement (PBT)	0.21	2.6	0.34
II. Low achievement (CAT)	0.10	1.6	0.17
IV. Depressed mood (B)	0.11	1.7	0.20
Family:			
IV. Poor supervision (PB)	0.20	2.6	0.34
II. Poor communication (PB)	0.08	1.5	0.14
IV. No set time home (PB)	0.09	1.5	0.16
V. Physical punishment (PB)	0.13	1.9	0.23
IV. Boy not involved (PB)	0.13	1.8	0.22
VI. Parent substance use (P)	0.08	1.4	0.13
Macro:			
II. Low socioeconomic status (P)	0.09	1.5	0.16
VI. Family on welfare (P)	0.21	2.4	0.32
III. Poor housing (P)	0.11	1.7	0.18
IV. Small house (P)	0.11	1.7	0.19
IV. Poorly educated mother (P)	0.14	1.9	0.23
I. Young mother (P)	0.16	2.1	0.27
VI. Broken family (P)	0.21	2.8	0.36
VI. African-American (P)	0.19	2.3	0.30
III. Bad neighbourhood (P)	0.15	2.0	0.25
VI. Bad neighbourhood (C)	0.17	2.1	0.28

Notes: B = boy; C = census; CAT = California Achievement Test; P = parent; T = teacher. All figures significant at $p < 0.05$, one-tailed. OR = odds ratio. TET = tetrachoric correlation. Categories: I–VI: see text.

who were serious delinquents. Equating the fraction at risk to the fraction of delinquents means that the (phi) correlation has a maximum value of 1. Also, our reasoning was that identifying half of the boys as deviant would identify too many 'normal' boys as deviant. Conversely, identifying only 10% of the boys as deviant would identify too few cases for statistical analyses (only 50 before missing data). Of course, where variables were inherently dichotomous, there was no choice in identifying the risk category. The effect of different dichotomization splits will be discussed later.

Relative importance of explanatory variables

Since the measured strength of association depends on the index used, it follows that a comparison of relative strengths of associations between explanatory variables and an outcome may produce different results depending on the index. Table 3 shows three statistics derived from relating the 40 dichotomized explanatory variables to delinquency (dichotomized to compare serious delinquents with the remainder) in the Pittsburgh Youth Study: phi, the odds ratio and the tetrachoric correlation. Only significant results are shown in the table; the number of significant variables (23 out of 40) was far greater than chance expectation. To the best of our knowledge, the only comparable study showing three measures of strength of association (phi, the odds ratio and relative improvement over chance) is by Cohen (1996).

Variables were divided into child, family and macro (socioeconomic, demographic and neighbourhood). Table 3 shows the source of the variable (e.g. parent, boy or teacher). Lack of guilt had the strongest relationship with delinquency. Of the family variables, poor parental supervision was the strongest predictor. Other important predictors of delinquency were hyperactivity-impulsivity-attention deficit (HIA) problems, the family on welfare and living in a bad neighbourhood.

According to phi, relationships with delinquency were weak. Over all 40 explanatory variables, the average phi was 0.096, which meant that, on average, each predictor accounted for just under 1% of the variance in serious delinquency. The strongest predictor -- lack of guilt -- accounted for only 7% of the variance. Tetrachoric correlations were higher (averaging 0.161). However, odds ratios give a different impression. Lack of guilt was associated with more than a tripling of the odds of delinquency, and 10 other explanatory variables at least doubled the odds of delinquency. Assuming that an odds ratio of 2 or greater indicates a strong relationship (Cohen, 1996), 11 variables were strongly related to delinquency.

None of the phi correlations reached the 0.30 level; Fleiss (1981, p. 60) considered that values below this indicated 'no more than trivial association'. However, Rosenthal (1990) pointed out that a phi correlation of 0.3, explaining 'only' 10% of the variance, could reflect the difference between treatment success rates of 30% and 60% (see also Farrington and Loeber, 1989, for a mathematical exposition of the relationship between phi and percentage differences). By any realistic standard, this would be a substantial effect.

While the three measures of strength of association had different interpretations, the order of importance of the explanatory variables was almost identical on each. This was quantified by calculating 'overall' product-moment correlations between measures over 40 variables. Thus, the phis correlated 0.998 with the tetrachoric correlations and 0.996 with the logarithms of the odds ratios (LORs), and the LORs correlated 0.999 with the tetrachoric correlations. Clearly, conclusions about relative strengths of association would be identical whatever measure was used. The remainder of this paper focuses on the phi correlation and the odds ratio as commonly used measures with very different interpretations.

It is not surprising that conclusions drawn from phi correlations about the relative importance of explanatory variables are very similar to those drawn from odds ratios. Asymptotically (and certainly with an *n* of 500), the phi correlation is mathematically related to the logarithm of the odds ratio (Agresti, 1990, p. 54). If the cell entries in a 2 x 2 table are (a, b, c, d), the asymptotic standard error of the logarithm of the odds ratio (LOR) is:

$$\mathrm{SE(LOR)} = \mathrm{SQRT}\left[\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right] \tag{1}$$

Hence, on the null hypothesis, (LOR/SE) squared is distributed as Z squared. But, Z squared = chi-squared = *n* x phi squared. Hence, phi and LOR are mathematically related.

Dichotomous versus continuous variables

In order to investigate whether there might be differences between results obtained with dichotomous and continuous variables, four sets of product-moment correlations (*r*s) were calculated between explanatory variables and delinquency:

(a) dichotomous predictors (one-quarter/three-quarters), dichotomous delinquency (serious/remainder);
(b) dichotomous predictors, continuous (four-category) delinquency;
(c) continuous predictors, dichotomous delinquency;
(d) continuous predictors, continuous (four-category) delinquency.

Since seven of the explanatory variables were originally coded dichotomously, 33 *r*s were calculated in each condition, and overall correlations were then calculated between the 33 *r*s.

Dichotomizing the outcome variable had little effect. For dichotomous explanatory variables, the overall correlation between dichotomous delinquency *r*s and continuous delinquency *r*s was 0.940. For continuous explanatory variables, the overall correlation between dichotomous and continuous delinquency *r*s was 0.960. Dichotomizing the explanatory variables had only slightly more effect. For dichotomous delinquency, the overall correlation between dichotomous and continuous predictor *r*s was 0.855. Similarly, for continuous delinquency, the overall correlation between dichotomous and continuous predictor *r*s was 0.865.

It would be unrealistic to assume that the four-point delinquency scale was an equal-interval scale, as required by conventional analytic strategies such as least-squares multiple regression. In any case, our interest was mainly in characteristics of the seriously delinquent boys, suggesting that they should be compared with the remainder. Therefore, the only realistic choice was between dichotomous and continuous explanatory variables, with dichotomous delinquency. As already noted, the results with these two types of variables were highly correlated (0.855). The average value of *r* was slightly greater with continuous explanatory variables (0.110, as opposed to 0.086 with dichotomous explanatory variables). However, the high overall correlation between the two approaches meant that they produced very similar results with regard to the relative importance of explanatory variables.

Generally, absolute values of *r*s were very similar using the tetrachoric correlation with dichotomous explanatory variables and dichotomous delinquency or using the product-moment correlation with continuous explanatory variables and continuous delinquency. The average correlations over 33 explanatory variables were 0.143 for continuous data and 0.146 for dichotomous data. Therefore, dichotomization does not inevitably lead to lower measured relationships or to lower statistical power, providing that appropriate measures of association are used (such as the tetrachoric correlation).

Non-linear relationships with delinquency

According to Sonuga-Burke (1998), non-linear relationships between causal factors and outcomes provide the strongest evidence for categorical models of disorder. Many statistical techniques assume linear relationships between variables, but it was not difficult to discover non-linear relationships with delinquency. To illustrate this, the 20 explanatory variables that were measured on the most continuous scales were divided into four quartiles and related to serious delinquency (dichotomized). Five were not significantly related to delinquency; Table 4 summarizes relationships for the remaining 15.

In particular, the age of the mother at the boy's birth was non-linearly related to delinquency. The mothers in the youngest quartile (age 19.4 or less) were associated with the highest prevalence of delinquent sons (50%), while the moderately old mothers (age 23.0–26.7) were associated with the lowest prevalence (24%). The significance of non-linear trends was measured using the Cochran-Armitage Linear Trend test (Agresti, 1990, pp. 100–102). This partitions the overall 4 x 2 chi-squared into linear and non-linear components. The non-linear component is distributed as chi-squared with 2 degrees of freedom. Age of the mother at the boy's birth was significantly non-linearly related to delinquency (chi-squared = 9.1, p = 0.01).

The other variable whose non-linearity was statistically significant on this test was positive parenting, which measured how far the mother praised her son for prosocial behaviour (chi-squared = 5.9, p = 0.05). This was an example of a protective factor with no opposing risk factor, since the prevalence of delinquency was low in the most positive category (26%) but similar in the other three categories (about 40%). In most cases, protective and risk effects were opposite sides of the same coin, since the prevalence of delinquency was typically low at one end of a scale and high at the other. Using four categories rather than a continuous score makes it possible to disentangle protective and risk effects.

The age of the boy was also non-linearly related to delinquency, but not quite significantly; 48% of the oldest boys (age 14.0 or older) were delinquent,

Table 4: Percentage delinquent in quartiles

	Quartiles			
	Good	Good average	Bad average	Bad
Child:				
Lack of guilt	10	20	42	57
Old for grade	27	28	31	48
High ADHD score	22	32	46	47
Low achievement (PBT)	17	33	41	55
Low achievement (CAT)	18	32	45	42
Depressed mood	31	32	36	47
Family:				
Poor supervision	23	29	39	53
Poor communication	23	35	45	43
Boy not involved	30	36	32	47
Positive parenting	26	42	40	39
Disagree on discipline	20	36	28	31
Macro:				
Low socioeconomic status	27	35	40	43
Poor housing	30	31	41	45
Young mother	34	24	35	50
Bad neighbourhood (P)	31	33	31	48

Notes: PBT = parent, boy, teacher; CAT = California Achievement Test

compared with 27–31% of those in the other three categories (chi-squared = 5.2, p = 0.07). This non-linear trend helps in interpreting why age is related to delinquency. If age reflected maturation, the percentage of delinquent boys in each age quartile should have increased linearly. The non-linear trend suggests that older age is a proxy for low school achievement; the oldest boys were those who had failed a grade and consequently had been held down. This is why we called this variable 'old for grade'.

Different dichotomization splits

One obvious problem with dichotomization is to decide where to dichotomize. It is important to investigate how far results are affected by different dichotomization splits. Therefore, we compared the chosen one-quarter/three-quarters split with a less extreme one (half/half) and with a more extreme one (one-sixth/five-sixths). Of the 40 explanatory variables, these three splits could be achieved for only the 20 most continuous variables.

Each of the dichotomous variables was related to the dichotomized delinquency variable, and phi correlations were calculated. For example, lack of guilt (split half/half) had a phi correlation of 0.35 with delinquency, whereas lack of guilt (split one-quarter/three-quarters) correlated 0.27, lack of guilt (split one-sixth/five-sixths) correlated 0.21, and lack of guilt (continuous) correlated 0.36. All four correlations were significant at $p = 0.0001$. This ordering of correlations is to be expected if the underlying relationship is linear.

With non-linear relationships, the ordering of correlations was reversed. For example, age of the mother at the boy's birth (split half/half) correlated 0.15 ($p = 0.0006$) with delinquency, whereas the one-quarter/three-quarters split variable correlated 0.16 ($p = 0.0002$), the one-sixth/five-sixths split variable correlated 0.18 ($p < 0.0001$), and the continuous variable correlated 0.11 ($p = 0.008$). Over all 20 variables, the average correlation was 0.092 for the one-sixth/five-sixths split, 0.104 for the one-quarter/three-quarters split, 0.108 for the half/half split, and 0.135 for the continuous variables, suggesting (as indeed Table 4 shows) that most of these variables were tolerably linearly related to delinquency.

More importantly, the one-quarter/three-quarters split variables had an overall correlation (over 20 *r*s) of 0.955 with the one-sixth/five-sixths split variables, 0.862 with the half/half split variables, and 0.870 with the continuous variables. These high overall correlations show that the order of importance of the variables in their relationships with delinquency was not changed much by the different dichotomization splits. The advantage of the one-quarter/three-quarters split is that it detects variables that are linearly or non-linearly related to delinquency. Non-linear relationships are detected less well by the half/half split (and by the continuous variables), while linear relationships are detected less well by the one-sixths/five-sixths split.

The continuous variables had the greatest overall correlation with the half/half split variables (0.962). This very high overall correlation shows that the order of importance of the variables in their relationships with delinquency was essentially unchanged by dichotomizing at the median. This suggested that dichotomization *per se* is likely to have little effect on conclusions about the relative importance of explanatory variables.

The effect of different dichotomization splits on the dependent variable of delinquency was also investigated. Phi correlations were calculated between the 20 explanatory variables (dichotomized at the median) and: (a) delinquency, dichotomized as usual at serious offences versus the remainder; and (b) delinquency, dichotomized at serious and moderate offences versus the remainder. The overall correlation between the two sets of *r*s was 0.886. Therefore, it seems likely that different dichotomization splits on the dependent variable would also have little effect on conclusions about the relative importance of explanatory variables.

Independently important variables

The commonly used ordinary least-squares (OLS) multiple regression may be problematic with the types of variables measured in criminological and psychiatric research; it may often be implausible to argue that a one-point change in an explanatory variable causes a *b*-point change in the outcome variable across the whole range of values of the explanatory variable. This requires equal-interval scales and linear relationships. Dichotomization avoids this problem. Logistic regression analysis is the most suitable technique with dichotomous data, and the (partial) odds ratio emerges as a key measure of the strength of influence of a variable after controlling for other variables (Fleiss et al., 1986). In logistic regression of dichotomous variables, the weighting of the independent variable in the equation essentially indicates (after an exponential transformation) the increase in the odds (risk) of the outcome associated with a risk factor (independently of all other measured variables).

Loeber et al. (1998, p. 278) systematically compared logistic regression with dichotomous explanatory variables and multiple regression with continuous explanatory variables in the Pittsburgh Youth Study. In general, the two methods identified the same explanatory variables as independently important. Where there were marked differences between the results, they were often understandable. For example, a young mother was more important as a dichotomous variable than as a continuous variable because of the non-linear relationship between the age of the mother and delinquency (Table 4). Being a teenage mother at the time of the boy's birth was a risk factor, but the (continuous) age of the mother was not linearly related to delinquency. Therefore, we considered the dichotomous result to be more valid and relevant.

Detecting interaction effects

An advantage of dichotomous data is that interaction effects can easily be studied. It is often important to know whether the effect of a risk factor varies according to the values of another variable. For example, individuals from poor families may not become delinquent if they have two loving parents (in which case two loving parents might be viewed as a protective factor counteracting the risk factor of poverty), or the effects of a risk factor (e.g. teenage mothers) may be different in good and bad neighbourhoods. With continuous data, it is so difficult to investigate and interpret interaction effects that they are rarely studied and even more rarely found in non-experimental projects (McClelland and Judd, 1993).

The implicit assumption in conventional analytic techniques is that all individuals are homogeneous, and that all explanatory variables have the same effect on the outcome variable for all individuals. On this reasoning, what differs between individuals is the values of explanatory variables, not the way in which explanatory variables add together (e.g. the weightings of explanatory variables in a multiple regression equation) to influence the outcome variable. However, there may be different types of individuals on which risk factors have different effects (e.g. there may be one set of weightings for type 1 individuals and a different set for type 2 individuals).

It is important to investigate possible interactions between explanatory variables. With continuous data, only multiplicative interactions are typically studied, thus making it difficult to distinguish between different kinds of interaction effects. Some of the possible types of interaction effects are as follows (Farrington, 1994, 1997):

(a) The joint occurrence of two risk factors is associated with a disproportionately higher prevalence of an outcome such as delinquency than would be obtained if the two factors were merely additive (called an enhancing effect). For example, assume that the two risk factors are both coded L = low and H = high. An example of this effect would be the following percentages of delinquents: 10(LL), 20(LH), 20(HL), 80(HH).
(b) The joint absence of two risk factors is associated with a disproportionately lower prevalence of delinquency than if the two factors were merely additive (called a suppressing effect). An example of this effect would be the following percentages of delinquents: 10(LL), 40(LH), 40(HL), 50(HH).
(c) The presence of one risk factor predicts delinquency only in the presence of the other, or alternatively the absence of one risk factor prevents the other from having an effect (called an interactive protective effect). An example of this effect would be the following percentages of delinquents: 10(LL), 10(LH), 10(HL), 50(HH).

Loeber et al. (1998, pp. 118–120) searched for interaction effects between all 40 explanatory variables in predicting delinquency, but only found 47 that were significant at $p = 0.05$, compared with the chance expectation of 39 out of 780 ($40 \times 39/2$). Overall, the yield of significant interactions was rather meagre in comparison with the large number of analyses that were carried out. In light of much theorizing about the importance of interaction effects, perhaps their absence was more noteworthy than their presence. One disadvantage of dichotomization is that interaction effects may be artefactually created by particular dichotomization cut-off points (Maxwell and Delaney, 1993; Veiel, 1988). Nevertheless, different types of interaction effects could at least be distinguished, investigated and communicated easily using dichotomous variables.

Multiple problem individuals

Conventional analytic strategies assume that all individuals differ in degree rather than in kind, but often the main interest may be in a small group of highly deviant, multiple-problem individuals who seem qualitatively different from the remainder and who account for a large fraction of the total problems in the population. For example, in delinquency research, it has often been demonstrated that about 5% of a birth cohort accounts for at least half of all the crimes committed by that birth cohort (e.g. Wolfgang et al., 1972). Often, there is more interest in these 'chronic offenders' (partly because of their importance as targets for intervention efforts) than in the total range of variation in offending.

Dichotomous data make it easy to identify individuals affected by multiple risk factors. Magnusson and Bergman (1988) argued that many of the relationships between single risk factors and single outcomes in their research were attributable to a relatively small number of multiple risk factor, multiple-problem individuals. When they deleted the 13% of their sample with multiple risk factors, some of the significant relationships between single risk factors and single outcomes disappeared (e.g. between adolescent aggressiveness, adult criminality and adult alcohol abuse). Hence, the effect of a single risk factor might vary according to whether or not it forms part of a larger cluster of risk factors in any individual.

The use of dichotomous data encourages a focus on types of individuals rather than on variables, as Richters (1997) has advocated. Conventional statistical techniques can specify the proportion of variance in an outcome that is attributable to a particular variable, but not the proportion of individuals whose delinquency might be attributable to that variable. In contrast, with dichotomous data, it is easy to specify what proportion of individuals possess a particular risk factor or combination of risk factors, and what are the particular risk factors affecting any given individual. From the point of view of the explanation, prevention and treatment of criminality and psychiatric disorders, it is important to focus on types of individuals rather than to assume homogeneity.

The multivariate results showing the independently important explanatory variables for delinquency do not reveal what percentage of boys are delinquent in different risk categories (i.e. possessing different numbers of risk factors). Therefore, each boy was scored according to the number of independently important explanatory variables that he possessed (out of the most important 11 discovered in the regression analyses). Table 5 divides the risk scores into five categories, each containing about 100 boys. The percentage who were delinquent increased markedly with the risk score, from 10% of boys with 0–1 risk factors to 70% of boys with six or more risk factors (the prime targets for intervention).

The best method of summarizing this type of risk-score data is to calculate an ROC curve, plotting the probability of a true positive or sensitivity (the fraction of delinquents identified at each possible cut-off point) against the probability of a false positive or (1-specificity) (the fraction of non-delinquents identified at each cut-off point). Fergusson et al. (1977) pioneered the use of this method in criminology. The area under the ROC curve is a useful measure of association that is independent of particular cut-off points and hence of prevalence and selection ratio (Mossman, 1994; Rice and Harris, 1995). It has a simple and meaningful interpretation: it is the probability that

Table 5: Percentage delinquent versus risk score

Score	% delinquent	*n*
0–1	10.2	118
2	26.1	88
3	36.0	86
4–5	45.1	122
6+	69.6	92
Total	36.6	506

a randomly selected delinquent will have a higher score than a randomly selected non-delinquent.

We calculated the best-fitting ROC curve for the relationship between the (0–11) risk score and delinquency, using the ROCFIT software of Charles E. Metz. The area under the curve was 0.746 (SD = 0.022), which was significantly greater than the chance expectation of 0.5. The average area under ROC curves in the Mossman (1994) meta-analysis of predictions of violence was 0.78.

Conclusions

Because of the great variety of different types of variables in typical criminological or psychiatric studies, and because very few variables are normally distributed and measured on interval scales, common statistical techniques based on product-moment correlations are problematic. Many researchers use *r* (or techniques based on *r*) to summarize all relationships irrespective of the nature of explanatory and outcome variables. However, skewed distributions of variables, outliers, non-linear relationships and interactions between variables all present problems.

There are various statistical methods of overcoming these problems, including creating or transforming into normally distributed continuous variables, including higher-order or multiplicative terms, and inspecting scatterplots. Ideally, each variable should be studied one by one and measures of association tailored accordingly, but this is difficult with a large number of variables. Also, statistical solutions may make it difficult to communicate the results of research effectively to policy makers, clinicians and criminal justice practitioners.

The main advantage of dichotomization is that it greatly simplifies the presentation of results and produces meaningful findings that are easily understandable to a wide audience. The main disadvantage of dichotomization is that information is lost, because cases below the threshold are treated as equivalent (as are cases above the threshold). To the extent that a risk factor is linked to extreme cases, this may not matter. For example, low intelligence may predict delinquency, but gradations of intelligence above the cut-off point may not be linearly related to delinquency. Loss of information is inevitable in many analyses, and the key issue is whether the information lost is relevant to the research question being addressed. It seems likely that the disadvantages of dichotomization would be more severe if variables were measured on interval scales and normally distributed and less severe if they were measured in a small number of categories and skewed.

We have argued that, in some cases, the advantages of dichotomization might outweigh its disadvantages, and we have illustrated our points using data from the Pittsburgh Youth Study. Our results may be influenced by the choice of a meaningful four-category delinquency measure rather than a less meaningful but more continuous scale. We would emphasize that our results are illustrative rather than exhaustive. We cannot claim that our conclusions would apply in all analyses; additional tests based on more empirical and simulated data, and mathematical discussions, are highly desirable. How far these results can be generalized to other psychiatric or behavioural/social science data sets and variables is an empirical question. It is also desirable to study the sensitivity of conclusions to different methods of measurement. We should also point out that we have not attempted an exposition of modern statistical methods of analysing dichotomous data, such as latent class analysis (Fergusson et al., 1991).

Our conclusions are as follows. First, dichotomization does not cause a decrease in the measured strength of association if appropriate measures are used (e.g. tetrachoric correlations). Second, a typical criminological or psychiatric study includes many different kinds of variables, including dichotomous, categorical and skewed continuous distributions. Truly continuous, equal-interval, normally distributed variables are rare. Dichotomization equates the sensitivity of measurement of all variables and makes it possible to compare the predictive strengths of explanatory variables. Third, dichotomization encourages a 'risk factor' approach, which helps in targeting intervention efforts. Fourth, dichotomization makes it possible to use the odds ratio, which is a more meaningful, interpretable and realistic measure of strength of association than the product-moment correlation and the percentage of variance explained.

Fifth, a number of variables are non-linearly related to an outcome such as delinquency, with a large increase in the prevalence of delinquency in the 'worst' category or a large decrease in the 'best' category. Dichotomization that identifies extreme categories reveals these phenomena, whereas product-moment correlations between continuous variables may conceal them. Often, the study of extreme cases is particularly important for explanation, prevention and treatment. Sixth, results are not greatly affected by particular dichotomization splits, providing that they are not too extreme. Seventh, dichotomization (compared with continuous data) does not greatly affect the order of importance of explanatory variables. Because of the mathematical relationship between the phi correlation and the logarithm of the odds ratio, correlations and odds ratios produce much the same ordering of importance of explanatory variables. Eighth, results are similar with logistic and OLS multiple regression. Ninth, different kinds of interaction effects can be easily and systematically studied with dichotomous variables. Tenth, multiple risk factor individuals can be easily and systematically studied with dichotomous variables.

In summary, dichotomization greatly simplifies the presentation of results, yields findings that are easily understandable by a wide audience, and shows no sign in our analyses of producing misleading conclusions. While it has been heavily criticized in the past, we suggest that it could be used more often in psychiatric and criminological research.

Acknowledgements

The authors are especially grateful to David Fergusson for his insightful comments on an earlier version of this paper.

References

Agresti A (1990) *Categorical Data Analysis*. New York: Wiley.

Caron C, Rutter M (1991) Comorbidity in child psychopathology: concepts, issues and research strategies. *Journal of Child Psychology and Psychiatry* 32: 1063–1080.

Cohen J (1983) The costs of dichotomization. *Applied Psychological Measurement* 7: 249–253.

Cohen J, Cohen P (1983) *Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum.

Cohen P (1996) Childhood risks for young adult symptoms of personality disorder: Method and substance. *Multivariate Behavioural Research* 31: 121–148.

Edwards AWF (1963) The measure of association in a 2 x 2 table. *Journal of the Royal Statistical Society (Series A)* 126: 109–114.

Farrington DP (1994) Interactions between individual and contextual factors in the development of offending. In Silbereisen RK, Todt E, eds. *Adolescence in Context*. New York: Springer-Verlag, pp. 366–389.

Farrington D P (1997) Key issues in studying the biosocial bases of violence. In Raine A, Brennan PA, Farrington DP, Mednick SA, eds. *Biosocial Bases of Violence*. New York: Plenum, pp. 293–300.

Farrington DP, Loeber R (1989) Relative improvement over chance (RIOC) and phi as measures of predictive efficiency and strength of association in 2 x 2 tables. *Journal of Quantitative Criminology* 5: 201–213.

Fergusson DM, Fifield JK, Slater SW (1977) Signal detectability theory and the evaluation of prediction tables. *Journal of Research in Crime and Delinquency* 14: 237–246.

Fergusson DM, Horwood LJ (1995) Predictive validity of categorically and dimensionally scored measures of disruptive childhood behaviours. *Journal of the American Academy of Child and Adolescent Psychiatry* 34: 477–487.

Fergusson DM, Horwood LJ, Lloyd M (1991) A latent class model of child offending. *Criminal Behaviour and Mental Health* 1: 90–106.

Fergusson DM, Horwood LJ, Lynskey MT (1993) The effects of conduct disorder and attention deficit in middle childhood on offending and scholastic ability at age 13. *Journal of Child Psychology and Psychiatry* 34: 899–916.

Fergusson DM, Lynskey MT (1995) Antisocial behaviour, unintentional and intentional injuries during adolescence. *Criminal Behaviour and Mental Health* 5: 312–329.

Fergusson DM, Lynskey MT, Horwood LJ (1997) Attentional difficulties in middle childhood and psychosocial outcomes in young adulthood. *Journal of Child Psychology and Psychiatry* 38: 633–644.

Fleiss JL (1981) *Statistical Methods for Rates and Proportions* 2nd edn. New York: Wiley.

Fleiss JL, Williams JBW, Dubro AF (1986) The logistic regression analysis of psychiatric data. *Journal of Psychiatric Research* 20: 145–209.

Fowler RL (1987) Power and robustness in product-moment correlation. *Applied Psychological Measurement* 11: 419–428.

Greenland S (1995) Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* 6: 356–365.

Hunter JE, Schmidt FL (1990) Dichotomization of continuous variables: the implications for meta-analysis. *Journal of Applied Psychology* 75: 334–349.

Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, Kupfer DJ (1997) Coming to terms with the terms of risk. *Archives of General Psychiatry* 54: 337–343.

Loeber R, Farrington DP, Stouthamer-Loeber M, van Kammen WB (1998) *Antisocial Behaviour and Mental Health Problems: Explanatory Factors in Childhood and Adolescence.* Mahwah, NJ: Lawrence Erlbaum.

Magnusson D, Bergman LR (1988) Individual and variable based approaches to longitudinal research on early risk factors. In Rutter M, ed. *Studies of Psychosocial Risk.* Cambridge: Cambridge University Press, pp. 45–61.

Maxwell SE, Delaney HD (1993) Bivariate median splits and spurious statistical significance. *Psychological Bulletin* 113: 181–190.

McClelland GH, Judd CM (1993) Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin* 114: 376–390.

Mossman D (1994) Assessing predictors of violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology* 62: 783–792.

Pickens RW, Svikis DS, McGue M, Lykken DT, Heston LL, Clayton PJ (1991) Heterogeneity in the inheritance of alcoholism: a study of male and female twins. *Archives of General Psychiatry* 48: 19–28.

Rice MF, Harris GT (1995) Violent recidivism: assessing predictive validity. *Journal of Consulting and Clinical Psychology* 63: 73–78.

Richters JE (1997) The Hubble hypothesis and the developmentalist's dilemma. *Development and Psychopathology* 9: 193–229.

Rosenthal R (1990) How are we doing in soft psychology? *American Psychologist* 45: 775–777.

Rowe DC, Farrington DP (1997) The familial transmission of criminal convictions. *Criminology* 35: 177–201.

Rutter M (1997) Comorbidity: concepts, claims and choices. *Criminal Behaviour and Mental Health* 7: 265–285.

Smith WR (1996) The effects of base rate and cutoff point choice on commonly used measures of association and accuracy in recidivism research. *Journal of Quantitative Criminology* 12: 83–111.

Sonuga-Burke EJS (1998) Categorical models of childhood disorder: a conceptual and empirical analysis. *Journal of Child Psychology and Psychiatry* 39: 115–133.

Swets JA (1986) Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin* 99: 100–117.

Veiel HOF (1988) Base-rates, cut-points, and interaction effects: the problem with dichotomized continuous variables. *Psychological Medicine* 18: 703–710.

Wolfgang ME, Figlio RM, Sellin T (1972) *Delinquency in a Birth Cohort.* Chicago: University of Chicago Press.

Zhao LP, Kolonel LN (1992) Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *American Journal of Epidemiology* 136: 464–474.

Address correspondence to: Professor David Farrington PhD FBA, Institute of Criminology, University of Cambridge, Cambridge, UK.