Dichotomizing Continuous Outcome Variables: Dependence of the Magnitude of Association and Statistical Power on the Cutpoint

Author(s): David R. Ragland

Source: *Epidemiology*, Vol. 3, No. 5 (Sep., 1992), pp. 434-440
Published by: Lippincott Williams & Wilkins
Stable URL: http://www.jstor.org/stable/3702637

Dichotomizing Continuous Outcome Variables: Dependence of the Magnitude of Association and Statistical Power on the Cutpoint

David R. Ragland

Dichotomizing a continuous outcome variable casts that variable in traditional epidemiologic terms (that is, disease, no disease). One consequence is overall reduced statistical power. A more fundamental concern is that the magnitude of various measures of association (for example, prevalence ratio, odds ratio) and statistical power depend on the cutpoint used to dichotomize the variable. The phenomenon is illustrated with a hypothetical situation assuming a two-level predictor variable and a normally distributed outcome variable. As the cutpoint is increased from lower to higher values, the prevalence ratio increases steadily, the odds ratio is described by a U-shaped curve, and statistical power is described by an inverted U-shaped curve. Furthermore, the *extent* of these effects depends on the difference between the means of the continuous outcome variable for the two levels of the predictor variable. An empirical example is given using data on education and blood pressure (dichotomized to create a high blood pressure *vs* low blood pressure variable). Except at each end of the distribution, the results follow the hypothetical example. The observation has implications for public health and medical treatment; different cutpoints should be examined to determine the optimal cutpoint in terms of policy and/or treatment decisions. The observation described here also has implications for statistical interpretation; statements about the magnitude of association or statistical significance have limited meaning unless both the cutpoint and the distribution of the outcome variable are specified. (Epidemiology 1992;3:434–440)

Keywords: dichotomous outcomes, data analysis, measures of association, odds ratio, prevalence studies, prevalence ratio, statistical power, study design.

Disease has traditionally been studied as a binary outcome: there is either disease or no disease. With this binary concept, a natural summary statistic is the "prevalence," that is, the number of diseased individuals divided by the total (diseased plus nondiseased). Measures of association such as the prevalence ratio and the odds ratio are used to compare risks under different conditions of exposure. Extensive statistical methods have been developed for the analysis of binary outcomes.[1]

In many studies, however, the outcome variable is continuous (for example, blood pressure, serum cholesterol, glucose tolerance, birth weight). Correlation-based methods, including regression analysis, are available to deal with such outcomes.[2] In some cases, it may be necessary or desirable to dichotomize these variables. For example, blood pressure over a certain value is likely to be treated. A common solution is to create a dichotomized measure of blood pressure, that is, hypertension, in which an individual is categorized as "hypertensive" if at least one of several overlapping conditions is met, such as diastolic blood pressure >90 mmHg, systolic blood pressure >140 mmHg, or treatment for high blood pressure. In other cases, a continuous variable is dichotomized because a range of the variable has distinct clinical significance. Blood pressure and other continuous variables such as birth weight,[3] respiratory function,[4] blood glucose,[5] olfactory function,[6] and depression[7] are often transformed into dichotomous variables for analysis.

These motivations are appropriate, but there are potential problems in the analysis of a continuous outcome variable that has been dichotomized. One is that dichotomizing variables reduces the statistical power of the analysis.[8] A second potential problem is more fundamental: *the magnitudes of the calculated statistics are functions of both the cutpoint and the underlying distribution of the outcome variable.* This problem is illustrated first with a hypothetical example and then with an empirical example. Although both examples illustrate the problem for prevalence data, the problem occurs also for studies of incidence.

From the Department of Biomedical and Environmental Health Sciences, 140 Warren Hall, School of Public Health, University of California at Berkeley, Berkeley, CA 94720.

Submitted May 13, 1991; final version accepted February 21, 1992.

Hypothetical Example: Age and Blood Pressure

Suppose we want to study the relation between age and blood pressure. Dichotomizing the predictor variable into young and old (for simplicity), let population A be young and population B be old, with $\mu_A = 128$ mmHg, $\mu_B = 132$ mmHg, $\sigma_A = \sigma_B = 15$ mmHg, and with blood pressure in both populations normally distributed (Figure 1). If these two populations are compared with respect to a dichotomous blood pressure outcome, then each subject is assigned either "low blood pressure" (that is, nonhypertensive) or "high blood pressure" (that is, hypertensive). In each case, the expected "prevalence" of hypertension is equal to the area under the normal curve of blood pressure to the right of the cutpoint.

If the cutpoint used is 140, for example, the expected prevalence of hypertension is 0.21 for the young population (A) and 0.30 for the old population (B). Now we can compare the prevalence of hypertension for each group. Two measures of association commonly used are the prevalence ratio P_B/P_A and the prevalence odds ratio $P_B(1 - P_A)/P_A(1 - P_B)$. In this example, age is positively related to blood pressure, with the prevalence ratio = 1.43 and the odds ratio = 1.61.

If the cutpoint used is 160, the expected prevalence of hypertension is 0.017 for the young population and 0.031 for the old population, giving a prevalence ratio of 1.94 and an odds ratio of 1.97. Using a cutpoint of 160 mmHg yields considerably stronger measures of association than those calculated when 140 mmHg is used as the cutpoint.

Figure 2, *a–c*, shows what happens to the prevalence ratio, the odds ratio, and the statistical power as the cutpoint defining hypertension moves through the range of blood pressure values. Figure 2*a* shows the normal distributions of blood pressure in the two groups. The prevalence ratio steadily increases as the cutpoint increases (Figure 2*b*). The odds ratio behaves somewhat differently as the cutpoint moves through the range of blood pressure values (Figure 2*b*). When the cutpoint is low, the odds ratio is large. As the cutpoint increases, the odds ratio first decreases and then increases. The odds ratio reaches its minimum value at the midpoint between the means of the two distributions. It is apparent from this example that both the odds ratio and the prevalence ratio vary substantially, even within the middle range of the distributions.

Figure 2*c* shows how the statistical power of detecting a difference between two populations[1] varies, assuming $n_1 = n_2 = 500$, as the cutpoint used to define hypertension moves through the range of possible values. When the cutpoint is 140 mmHg, the power is 0.84; if the cutpoint is increased to 160 mmHg, the power is reduced to 0.24. The power function is symmetrical about the midpoint between the means of the distributions; it increases rapidly to a maximum at the midpoint and then declines.

The relationships between the cutpoint and the calculated measures of association, illustrated in Figure 2, *a–c*, are functions of the z-score difference between the means of the two distributions. Figure 3, *a–c*, shows how the relationship between the cutpoint and the measures of association and statistical power depend on the difference (D) between the two distribution means. The variations in the prevalence ratio (3*a*), the odds ratio (3*b*), and the statistical power (3*c*) increase as the *difference* between the means in the two populations increases.

Empirical Example: Education and Systolic Blood Pressure

Table 1 gives descriptive statistics for systolic blood pressure by education level group (high school only or some college *vs* college graduate) from the Western Collaborative Group Study. The means are only

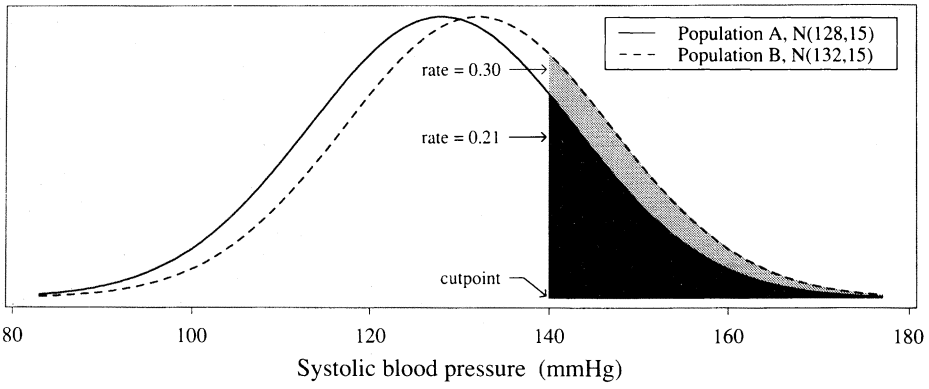

FIGURE 1.
Theoretical systolic blood pressure distribution of hypothetical populations A and B and corresponding hypertension prevalences generated by a cutpoint of 140 mmHg.

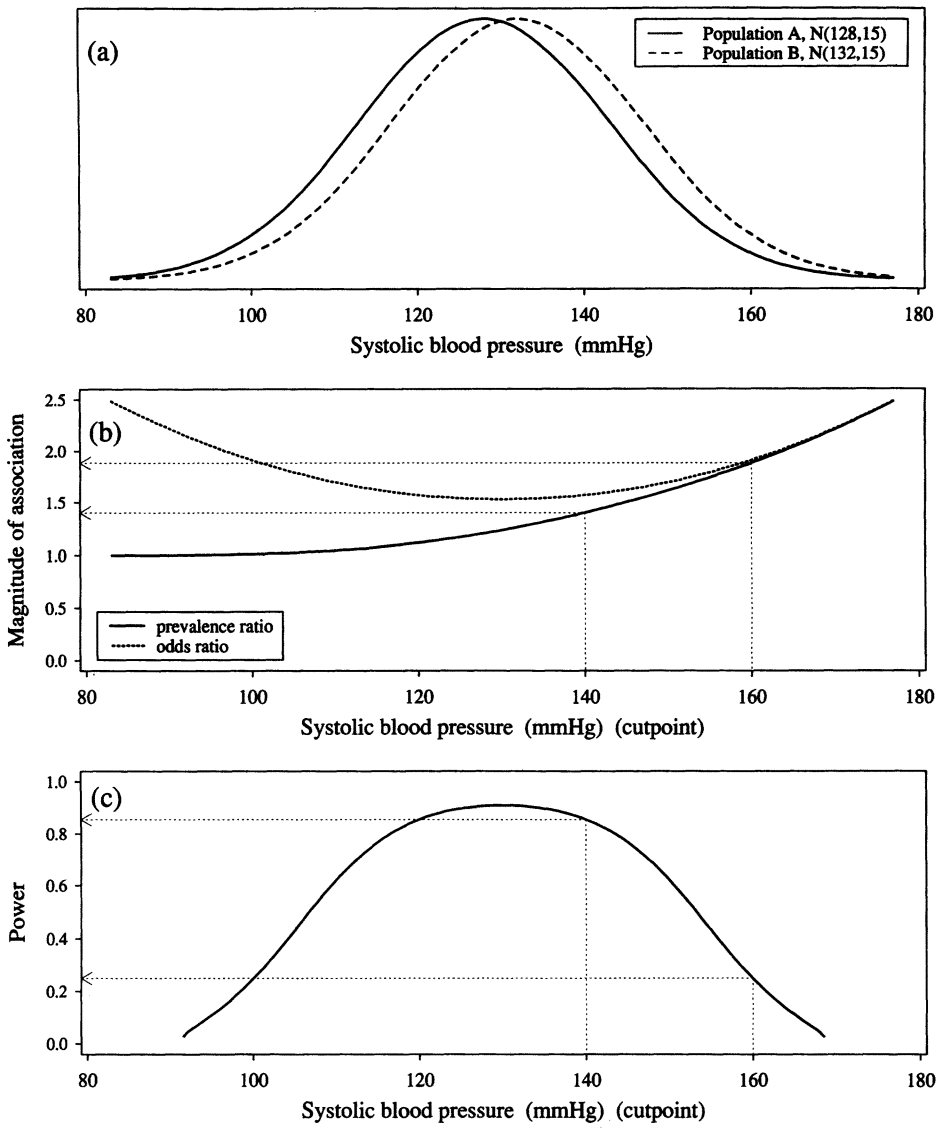

FIGURE 2.
***a*, hypothetical normal densities for systolic blood pressure in two populations. *b*, prevalence ratio and odds ratio by the cutpoint used to dichotomize blood pressure. *c*, statistical power (assuming $n_1 = n_2 = 500$) as a function of the cutpoint used to dichotomize blood pressure.**

slightly different (129.8 mmHg for the group with high school only or some college, and 127.7 for the group of college graduates), but because of the large study size, the difference produces a low *P*-value ($P = 0.0001$).

Figure 4*a* illustrates the systolic blood pressure distributions for the two education groups in the Western Collaborative Group Study. Figure 4, *b* and *c*, shows what happens when the cutpoint defining hypertension is placed at different points on the blood pressure axis. In Figure 4*b*, the prevalence ratio and odds ratio are shown for different cutpoints. As in the hypothetical example illustrated in Figure 2*a*, the prevalence ratio is 1 when the cutpoint is low, and then it steadily increases as the cutpoint increases. When the cutpoint is very high, then the numbers in the high group become too low, and the prevalence ratio is unstable. The odds ratio is high for low cutpoints, increases, and then follows the prevalence ratio. The odds ratio is higher than the prevalence ratio throughout the range of cutpoints.

Figure 4*c* shows the *P*-value comparing the two education level groups computed at each possible cutpoint value of systolic blood pressure. Reflecting the variation in statistical power, the *P*-value varies substantially with changes in the cutpoint.

In this example, cutpoints of 140 mmHg and 160 mmHg yield similar values for the prevalence ratio and odds ratio (Figure 4*b*), but other cutpoints yield quite

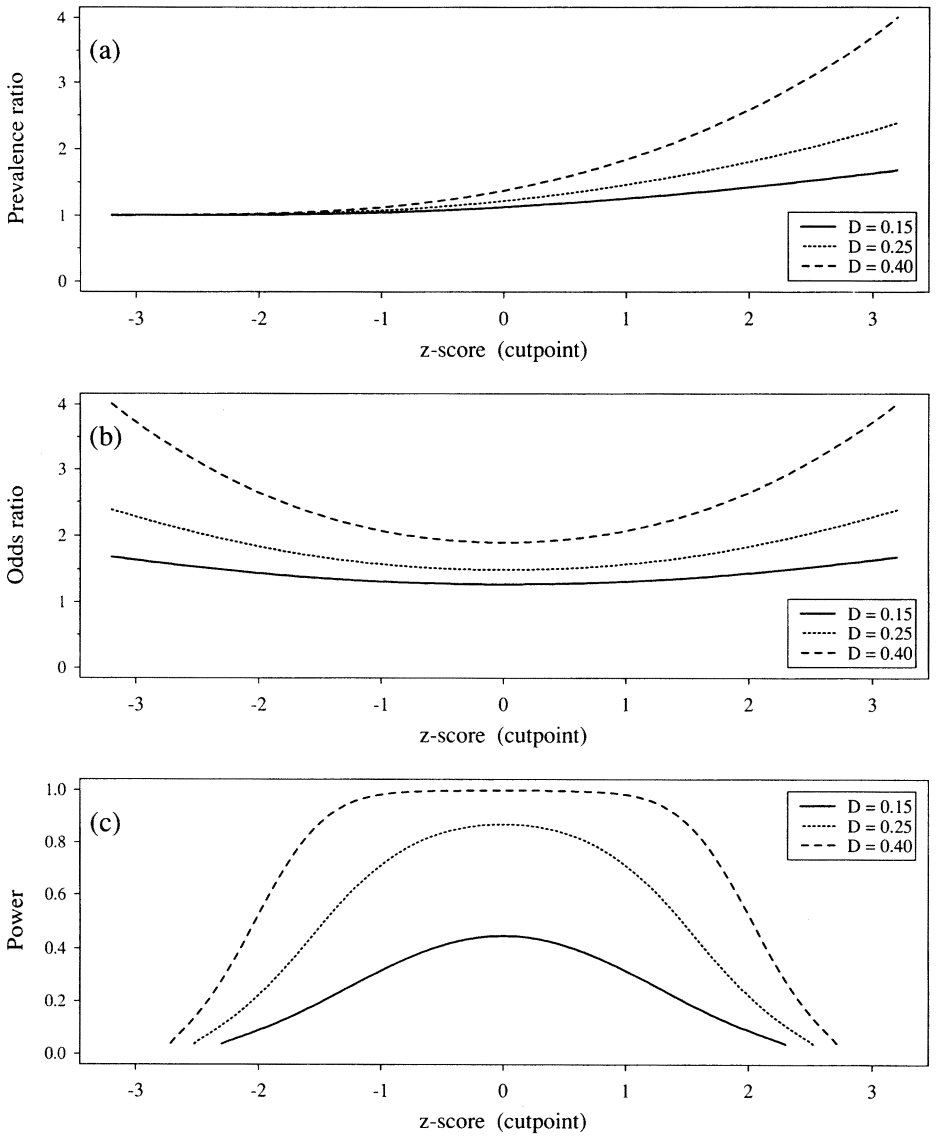

FIGURE 3.
***a*, prevalence ratio; *b*, odds ratio; and *c*, statistical power (assuming $n_1 = n_2 = 500$) as a function of the cutpoint used to dichotomize blood pressure.**

TABLE 1. Systolic Blood Pressure (mmHg) Mean and Standard Deviations for Two Education Groups (High School Only or Some College; College Graduate), and *t*-Test Value for the Difference between the Means*

Statistic	Education Group: High School Only or Some College	Education Group: College Graduate
Number	1,425	1,729
Mean blood pressure	129.8	127.7
Standard deviation	15.2	15.0
t-Test		
t-Value	3.97	
Degrees of freedom	3,152	
Significance level	<0.0001	

* From the Western Collaborative Group Study, 1960–1961.[9]

different values. The *P*-values are quite different for the two cutpoints (Figure 4c), $P < 0.01$ for 140 mmHg and $P = 0.10$ for 160 mmHg.

Discussion

The distribution of an outcome measure for each level of a predictor variable represents the entire set of information available for analysis. When the outcome variable is continuous, means can be generated for each level of the predictor variable and used to construct a summary measure of association. When the continuous outcome variable is dichotomized, some of the information contained in the underlying distribution is discarded. The information abstracted from this underlying distribution is not independent of how that information was abstracted; that is, there is not a

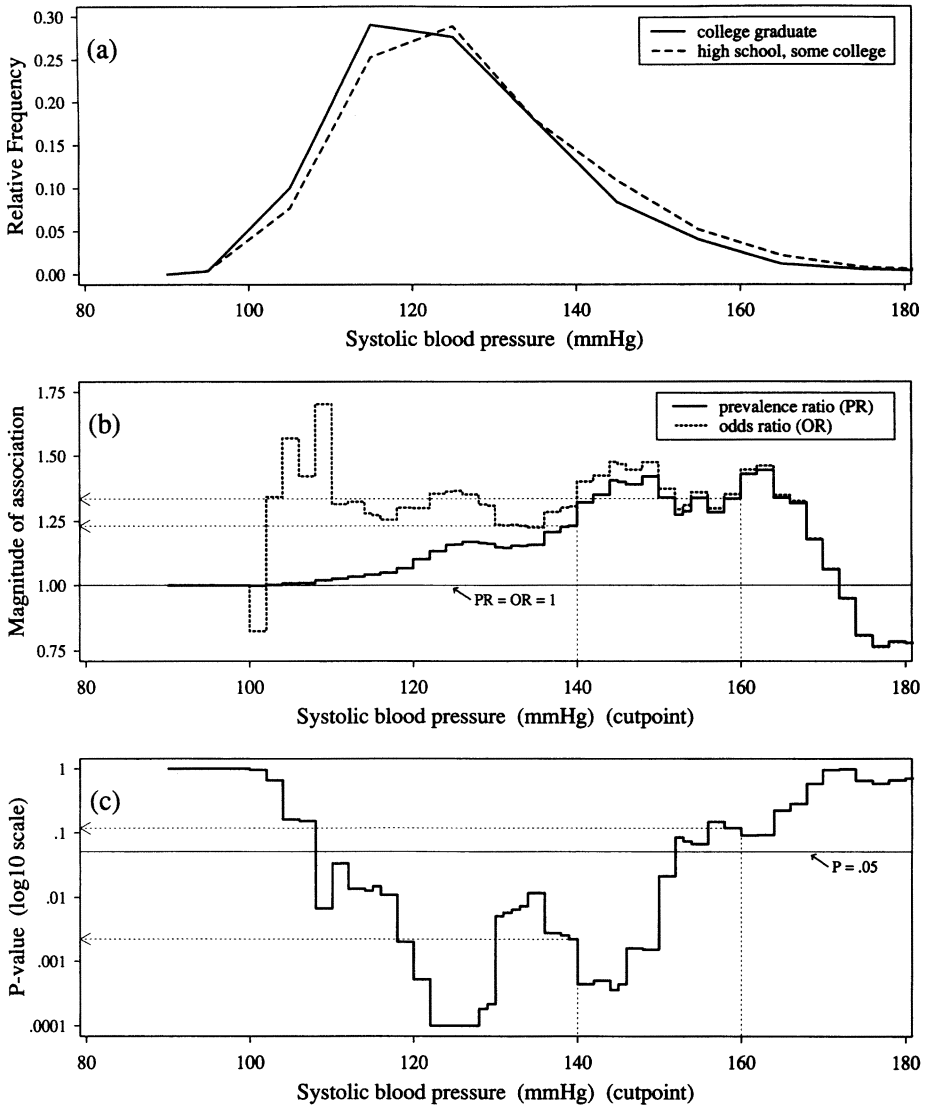

FIGURE 4.
***a*, distribution of systolic blood pressure for two education levels. *b*, prevalence ratio and odds ratio by the cutpoint used to dichotomize blood pressure. *c*, *P*-value for difference in populations as a function of the cutpoint used to dichotomize blood pressure.**

unique correspondence between the summary expression and the underlying relation.

The model used here for illustrating this point assumes a normal distribution with fixed variance. The empirical example used a variable that approximates these conditions. It can be shown that different assumptions about the shape and variance of the distributions can have quite different results but, in general, that the measured effect cannot be assumed to be constant, even approximately, over changes in the cutpoint.

It should be noted that the phenomenon described here applies only to outcome variables. Usually, changing cutpoints in a predictor variable will not produce systematic changes in summary measures of association.[10] It should also be noted that, for the sake of simplicity, the predictor variable is assumed to be dichotomous. The same basic phenomenon, however, will be seen when the predictor variable is continuous—for example, an odds ratio based on two different values of a continuous predictor variable will show a systematic variation in magnitude as the cutpoint for the outcome variable is shifted.

One additional assumption is that the exposure variable shifts the entire distribution of the outcome variable. Under a different assumption, for example, that exposure only made some noncases into cases, without a more general shift in the distribution, other patterns would be observed.

IMPLICATIONS

It is common practice in clinical and epidemiologic studies to dichotomize continuous outcome variables, in part because therapeutic and policy decisions are

usually based on thresholds and cutpoints rather than on continuous values. The "prevalence" based on such dichotomized outcomes is reported, and the "prevalences" among different groups are compared using various measures of association and statistical tests.

The implications for this practice depend on the goal of the analysis. If the goal is to determine public policy, or to plan treatment, then one may want to systematically examine different cutpoints to determine the optimal cutpoint for the specific purpose in mind. For example, when hypertension prevalences for black *vs* white populations are compared, the ratio of black to white hypertension is greater for more severe hypertension.[11] The implications of severe hypertension are much different than for a wider measure that also includes mild hypertension, and decisions about the importance of elevated blood pressure in different populations should be made considering the appropriate cutpoint.

On the other hand, if the intent of the analysis is to examine the magnitude of or the *P*-value for an association between two variables, then the implication is different. In this case, it seems inappropriate to explore until one finds that cutpoint that simultaneously optimizes both the magnitude of association and the *P*-value and then to report the results based only on this cutpoint. This caution always applies because different results can arise from statistical fluctuation. Beyond this fluctuation, however, the additional phenomenon described here indicates a systematic variation in results obtained from the same set of distributions.

In either case, whether the goal is to make policy decisions or to evaluate a measure of association, the examples show that within the same set of data, different cutpoints can lead to different measures of association. Some published studies in which a continuous outcome measure has been dichotomized might have yielded different conclusions if different cutpoints had been used. For example, using data from the empirical example given above, the conclusion of a report on the association between education and dichotomized systolic blood pressure would depend on the cutpoint chosen.

It follows also that comparisons across studies using dichotomized outcome variables have limited meaning without specifying the cutpoint and the underlying distributions for the studies compared. At the least, interstudy comparisons of measures of association should be based on similar cutpoints. Even in research areas where the cutpoint used across different studies is often the same, (for example, studies of depression,[7] studies of pulmonary function[4]), differences in the shape or position of the underlying distributions can introduce systematic variation into comparisons between and among studies. This phenomenon represents a source of variation among studies and has implications for reviews or meta-analyses in which systematic comparisons are made across studies.

Courses of Action

What should be done when it is necessary or desirable to dichotomize a continuous outcome variable? One possible solution is to utilize a cutpoint that is in some sense "standard," is viewed as a threshold for therapeutic intervention, or has some other clinical significance. For many continuous variables of interest in epidemiologic and medical studies, however, there is no single cutpoint on which all researchers and clinicians agree, and, in any case, differences in underlying distributions can still affect the calculated statistics. Another possible solution is to report results for each possible cutpoint, for example, as shown in Figure 4, *b* and *c*. Such a presentation permits reporting the "entire picture," although it has the drawback of being considerably more complicated.

The ideal solution would take into account the entire distribution for purposes of statistical testing but would allow measures of association for different cutpoints to be presented. One possibility is to adopt an approach widely used in evaluating sensitivity and specificity of diagnostic tests. Statistical and graphical techniques described in conjunction with the receiver operator characteristic are used in that context, which allow not only a graphical presentation of the association over different cutpoints but produce a single summary statistic for statistical testing.[12]

References

1. Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed. New York: Wiley and Sons, 1981.
2. Neter J, Wasserman W. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. Homewood, IL: Irwin, Inc., 1974.
3. Doyle LW, Kitchen WH, Ford GW, Rickards AL, Kelly EA, Callanan C, Raven J, Olinsky A. Outcome to 8 years of infants less than 1000 g birthweight: relationship with neonatal ventilator and oxygen therapy. J Paediatr Child Health 1991;27: 184–188.
4. Sedline S, Tager I, Speizer FE, Rosner B, Weiss ST. Longitudinal variability in airway responsiveness in a population-based sample of children and young adults: intrinsic and extrinsic contributing factors. Am Rev Respir Dis 1989;140:172–178.
5. Sugarman J, Percy C. Prevalence of diabetes in a Navajo Indian Community. Am J Public Health 1989;79:511–513.
6. Schwartz BS, Doty RL, Monroe C, Frye R, Barker S. Olfactory function in chemical workers exposed to acrylate and methacrylate vapors. Am J Public Health 1989;79:613–618.
7. Barnes GE, Currie RF, Segall A. Symptoms of depression in a Canadian urban sample. Can J Psychiatry 1988;33:386–393.

8. Selvin S. Two issues concerning the analysis of grouped data. Eur J Epidemiol 1987;3:284–287.
9. Rosenman RH, Brand RJ, Jenkins CD, Friedman M, Straus R, Wurm M. Coronary heart disease in the Western Collaborative Group Study: final follow-up experience of 8½ years. JAMA 1975;233:872–877.
10. Rothman K. Modern Epidemiology. 1st ed. Boston: Little, Brown, 1986.
11. Hypertension Detection and Follow-up Program Cooperative Group. Race, education and prevalence of hypertension. Am J Epidemiol 1977;106:351–361.
12. Swets JA, Pickett RM. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York: Academic Press, 1982.