0145-6008/96/2007-1148$03.00/0
ALCOHOLISM: CLINICAL AND EXPERIMENTAL RESEARCH

Vol. 20, No. 7
October 1996

Application of Decision Theory to DUI Assessment

Barry Leshowitz and Jonathan M. Meyers

The application of decision theory to screening of driving under the influence (DUI) offenders is illustrated in an evaluation study that investigated the validity of a structured interview and a survey instrument. These findings are examined graphically using the relative operating characteristic (ROC). This graph relates the proportion of true-positive cases to the proportion of false-positive cases for various placements of the decision cut-off score. The ROC's two principal parameters, test sensitivity and examiner bias, are used to provide a complete quantitative description of the performance of two screening instruments. A major goal of this study is to show how a new measure of examiner bias, the "cost ratio," improves the evaluation of DUI screening programs.

Key Words: DUI Screening, Decision-Making, Relative Operating Characteristic, Base Rates, Examiner Bias.

IN THIS study, we will illustrate how decision theory can be used to improve decision-making in driving under the influence (DUI) screening programs. In this effort, we will analyze the results of a recent study conducted at the Center for Prevention Research (CPR) at the University of Kentucky. It compared the validity of a survey instrument, which has been used in several statewide screening programs, with that of a structured interview based on DSM-III-R criteria.[1,2] We will apply decision theory's principal outcome measure, the relative (or receiver) operating characteristic (ROC), to the results of this study. In addition, we derive a new quantitative measure of examiner bias called the "cost ratio." By examining projections of various decision outcomes associated with changes in the cost ratio, we show that this index of examiner bias should be helpful in monitoring the direction and scope of DUI screening programs.

According to decision theory, the task of the assessor is to evaluate the value of the decision variable or test result. In DUI screening, if the decision variable is deemed by the assessor to represent a "positive" event, the offender is judged to be "high risk" for future DUI offenses. If the decision variable is deemed to represent a "negative" event, the offender is judged to be "low risk" for future DUI offenses. These judgments generally determine the level of intervention (e.g., education or treatment) appropriate for the convicted DUI offender. This application of the decision-theory model is depicted graphically in the inset to Fig. 1. Administration of the screening instrument to convicted DUI offenders generates two equal-variance Gaussian distributions of the decision variable: one distribution depicts positive cases that are assumed to be at high risk for future DUI offenses, and the other, negative cases who are of low risk for future DUI offenses.[3,4] Because the test instrument's measurement precision is not perfect, the two distributions of test results overlap. As a result, mistakes in judgment are inevitable, because the same value of the decision variable can arise from either the positive or negative event.

To deal with ambiguous test results, decision theory assumes that the assessor adopts the following decision rule: if the value of the decision variable (test score) exceeds the operative cut-off, the offender is judged to be "high risk" for future DUI offenses; otherwise, the offender is judged to be "low risk" for future DUI offenses. Thus, the greater (or more strict) the cut-off, the more likely ambiguous cases will be classified as "low risk." The more lenient the cut-off, the more likely cases will be classified as "high risk." Decision-making performance is evaluated by plotting the outcomes of the decisions as an ROC, which is a graph plotting the proportion of true-positives against the proportion of false-positives for various cut-off scores. Statistics describing test sensitivity and examiner bias, which are estimated from the theoretical ROC fitted to data, provide a complete description of test performance.

The aim of this study is to provide an example of the application of decision theory to assessment of DUI offenders. Decision theory, which is sometimes called "signal-detection theory," has been extensively applied to a broad range of practical diagnostic systems in medicine and engineering.[5–7] More recently, this technique has been applied to a diversity of mental health issues, including psychiatric assessment,[8] attention deficit hyperactive disorder in children,[9] alcohol disorders,[10] and violent behavior.[11] In this study, we seek to extend decision theory's application and show how the ROC may be used to increase informed decision-making in DUI screening programs.

METHODS

Overview

In this study, we analyze data collected in a study conducted by the CPR at the University of Kentucky[1] that compared the concurrent validity of the Alcohol Scale of the Driver Risk Inventory (DRI)[12] with that of the Fayette County DUI Assessment Interview (CPR interview).[2]

Participants

A total of 333 DUI offenders from Fayette County, KY, was evaluated in the Kentucky study.

From the Department of Psychology (B.L.), Arizona State University, Tempe, Arizona; and the Department of Psychology (J.M.M.), University of Virginia, Charlottesville, Virginia.

Received for publication December 27, 1995; accepted May 7, 1996

Reprint requests: Barry Leshowitz, Ph.D., Department of Psychology, Arizona State University, Tempe, AZ 85287-1104.

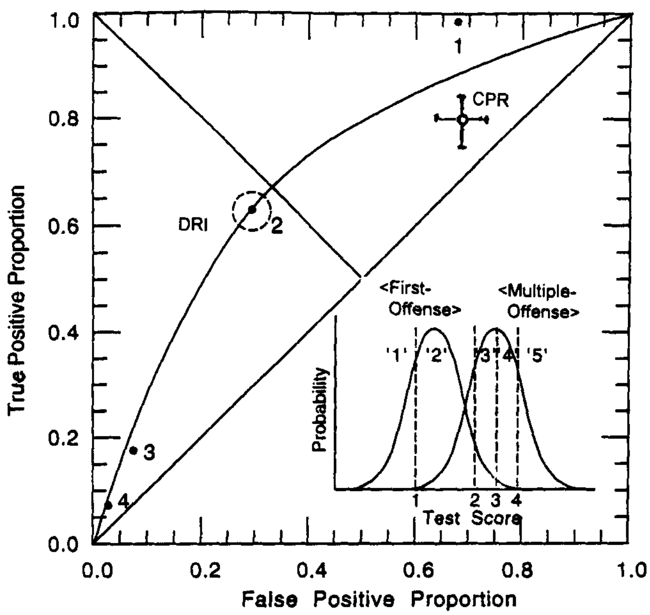

Fig. 1. Theoretical ROC (solid curve) fitted to the empirical rating-scale data for the DRI Alcohol Scale. Individual points are cumulative proportions measured from their respective criteria (for values, see Table 2). Circled point corresponds to the criterion used for collapsing the rating-scale data into a dichotomous variable. Inset depicts the underlying decision-theory model (not drawn to scale) for the DRI's 5-point rating procedure. Outcome of the CPR interview is shown with bars indicating ±1 SD, assuming binomial variance and 300 observations/point.

Materials

The DRI is a paper-and-pencil survey instrument consisting of five scales, one of which is devoted to assessing alcohol-related problems within a five-level risk classification system. The DRI's Alcohol Scale produces percentile scores ranging from 1 to 100 converted to 1 of 5 treatment recommendations, ranging from several hours of education to residential care. For the present analysis, the three extreme treatment recommendations calling for medical care indicated "high risk," whereas the two moderate treatment recommendations calling for educational intervention indicated "low risk."

The CPR interview was developed specifically for the Kentucky study by the Bluegrass Region Comprehensive Care and is based on the nine DSM-III-R criteria for alcoholic disease as set forth by the American Psychiatric Association.[13] Scores for the CPR interview are based on responses to the items selected from the nine clinical areas that define alcohol abuse in the DSM-III-R. Each area is scored by the assessor as either 0 or 1, with 1 being scored for any positive response to items concerning the area in question. For the present analysis, a total score of 3 or more indicated "high risk," whereas a score of <3 indicated "low risk." Illustrative items from both instruments are presented in Appendix A.

Procedure

After conviction of DUI, certified assessors administered the DRI survey and the CPR interview to each offender, and then made a treatment recommendation based on the combined results of the two instruments. In this investigation, we analyze the performance of the DRI and CPR instruments separately and will not consider the assessors' treatment recommendations.

Employing the methodology for classifying cases used in the original Kentucky study, we use previous DUI convictions as the criterion measure in our analysis of the concurrent validity of the DRI and CPR instruments. Offenders with no prior offenses are classified as "low risk" for future DUI offenses, whereas offenders with one or more prior offenses are classified as "high risk." The rationale for selection of this criterion measure is further elaborated in the "Discussion" section.

Table 1. Contingency Table Presenting Data Obtained in the Kentucky Study for the CPR Interview and the DRI Survey Instrument

	Offender status		
	Multiple offense	First offense	
High risk	True-positive	False-positive	"Positive bias"
CPR	79 (0.80)	161 (0.69)	240 (0.72)
DRI	62 (0.63)	71 (0.30)	133 (0.40)
Low risk	False-negative (miss)	True-negative	"Negative bias"
CPR	20 (0.20)	73 (0.31)	93 (0.28)
DRI	37 (0.37)	163 (0.70)	200 (0.60)
Total	99 (1.0)	234 (1.0)	333

Note: Entries denote the actual frequencies of each outcome. Corresponding proportions are presented in parentheses. Row totals indicate positive and negative bias.

RESULTS

Measurement of Overall Accuracy

Overall accuracy rates for the DRI and the CPR interview are displayed in a contingency table (Table 1). The contingency table crosses the frequencies of "known" cases of low risk and high risk with the frequencies of cases classified by the assessors. It can be seen in Table 1 that, using the CPR interview, the examiners correctly categorized 80% of the multiple-offense cases as high risk (true-positive rate). They correctly categorized first-offense cases as low risk at a rate of only 31% (true-negative rate). Summing the frequency of true positives and the frequency of true negatives divided by the total number of observations leads to an overall accuracy rate for the CPR interview of 46%. The DRI instrument achieved a much higher rate of accuracy. It correctly classified 63% of the multiple-offense cases as "high risk," and correctly categorized first-offense cases as "low risk" at a rate of 70%. Combining the two measures, the overall accuracy of the DRI was 68%.

An alternative approach to measuring test accuracy entails fitting decision theory's ROC graph to the data. The main virtue of decision theory's estimate of test accuracy is that it is independent of the cut-off selected by the examiner. As a result, it reflects the precision of the actual instrument. Figure 1 depicts the test results obtained for the DRI and CPR interview. The solid curve is the theoretical curve fitted to the DRI's rating-scale data. The Gaussian model shown in the inset, which is not drawn to scale, was used to generate the theoretical curve. Accuracy of the DRI can be estimated from the theoretical ROC by computing its displacement from the positive diagonal. This parameter of the ROC is the d' (or "d" prime) measure of test sensitivity. It corresponds to the separation between the two theoretical Gaussian distributions shown in the inset. For comparison, Fig. 1 also depicts the outcome of the CPR interview's binary-response procedure. (The theoretical ROC fitted to this single data point is not shown.) By inspection, d' or test sensitivity for the DRI Alcohol

Table 2. Effects of Placement of the Decision Cut-Off Score on False-Positives (FPs), True-Positives (TPs), Examiner Bias, Cost Ratio, and Projected Outcomes for Kentucky for the DRI Rating-Scale Survey and for the CPR Interview's Binary-Response Procedure

	Cut-off score				
	4	3	2	1*	CPR
FP (first-time offender)	0.03	0.08	0.31	0.68	0.69
TP (multiple offender)	0.07	0.17	0.63	1.00	0.80
Examiner bias (cut-off)†	(−)‡ 2.25	(−) 1.79	(−) 1.08	(+) 0.10	(+) 0.77
Cost ratio (C_{FP}/C_{FN})§	0.97	0.76	0.46	0.04	0.33
FP (projected)¶	945	2,520	13,950	21,420	21,735
Miss (projected)	12,555	11,205	4,995	0	2,700

* Entries in this column are approximations, because observed true-positive rate of 1.0 is untenable.

† Values of the cut-off.

‡ Direction of examiner bias in parentheses.

§ See Appendix B for an explanation of calculations. FN, false negative (miss).

¶ Projected frequency of false-positives and misses in Kentucky annually.

Scale is greater than that for the CPR interview. This finding is consistent with the earlier described difference in the overall accuracy of the two instruments.

Measurement of Examiner Bias

We now consider the effects of cut-off-score placement on the various decision outcomes for the DRI and CPR interview. Recall that the DRI incorporates a rating-scale procedure that converts percentile scores into levels of treatment. As shown in the model depicted in the inset to Fig. 1, the conversion requires the application of four simultaneously held cut-offs. In Fig. 1, the DRI's empirical rating-scale ROC comprises 4 points, with each successive point on the ROC corresponding to an increasingly more stringent cut-off score. The values of the plotted points are presented in Table 2. The solid ROC curve fitted to the DRI data depicts the theoretical relationship between placement of the cut-off score and rates of true-positives and false-negatives. As the cut-off becomes more stringent, both false-positives and true-positives decrease. The trade-off between increasing the desirable true-positive rate at the cost of increasing the undesirable false-positive rate depicted in the ROC is a consequence of the probabilistic nature of the decision-making situation. However, within the constraints provided by an imperfect test instrument, the examiners can seek to optimize their decisions. The topic of optimizing decision-making is considered in the "Discussion."

DISCUSSION

Decision-making is optimized when placement of the cut-off maximizes the expected value of the decision. The effect of base rates on the location of the optimal cut-off is discussed first.[14] Inspection of the ROC in Fig. 1 and summary data in Tables 1 and 2 reveal that DRI's centered criterion (#2 in Fig. 1) is to the left of the negative diagonal. Selection of this location is consistent with the greater incidence of first-offense cases in the study's sample. In contrast, we observe that the CPR interviewers' single cut-off was biased "positively," as evidenced by its placement to the right of the negative diagonal. Biased toward making "high-risk" judgments, the CPR examiners achieved a seemingly acceptable true-positive rate, but only at the cost of a greatly inflated false-positive rate and a low overall accuracy rate.

If all that mattered in decision-making was the achievement of the highest overall accuracy rate, it can be shown that placement of the cut-off should be based entirely on the base rates of known cases in the population. For example, in the present study, given the 70% incidence of first-time (low-risk) offenders in the DUI population, strict application of the base-rate principle would require that *every* DUI case be blindly diagnosed as "low-risk." This type of examiner bias would achieve a "respectable" overall accuracy rate of 70%. From the broader perspective of the long-term interests of society, however, such an outcome would be unacceptable. Many multiple-offense individuals who are experiencing serious alcohol-related problems would be categorized as "low-risk." Because of this improper classification, these high-risk individuals would be assigned to an inappropriate education program when a more intensive treatment program would be warranted. The bottom-line implications of this hypothetical example are clear: Besides base-rate information, the costs and benefits of the decision outcomes should be explicitly considered in setting an optimal cut-off.

Insight into how to balance base rates and the costs and benefits may be gained by examining the decision's "cost ratio." This measure of examiner bias relates the perceived cost of a false alarm to that of a miss. As an illustration of cost-ratio analysis of examiner bias, the reader is directed to the empirical estimates of the cost ratio for the Kentucky study. In Table 2, we present calculations of the cost ratio for decisions rendered for the CPR and DRI instruments (see Appendix B for an explanation of these calculations). The cost ratio of 0.33 for the CPR (see the rightmost column) shows that examiners valued the cost of a false alarm to be one-third that of a miss. In contrast, examiners using the DRI instrument with cut-off #2 valued the cost of a false-positive to be about one-half that of a miss (false-positive/miss = 0.46).

To illustrate the practical impact of these cost ratios on administration of DUI screening programs, we have projected miss and false-positive rates for the state of Kentucky for the cost ratios presented in Table 2. These annualized projections for the two instruments assume a total of 45,000 DUI offenders in Kentucky.[1] For the CPR interview, we project that 2,700 high-risk DUI offenders would be "missed" in screening. Nearly 22,000 low-risk offenders would be "falsely" judged to be in need of intensive treatment. Application of the DRI with cut-off #2 would result in misses and false-positives of 4,995 and 13,950, respectively. The observed differences in the projected outcomes for the two instruments are caused by the precision of the two instruments and placement of the cut-off score.

Whether or not an observed cost ratio and associated decision outcomes are deemed acceptable depends on the values the community holds on rehabilitation of DUI offenders. For example, anchoring a great deal of decision-making in our legal justice system is the principle that, "it is better that ten guilty persons escape than that one innocent suffers."[15] This legal maxim translates into a decision rule that posits the cost of a false-positive to be 10 times that of a miss. Examiners in the Kentucky study did not come close to achieving a 10:1 cost ratio. Apparently believing that the safety of others is placed at great risk by the actions of DUI offenders, the examiners adopted cost ratios that were <1.

Cost ratios may also be used to help personnel to set optimal cut-off scores in DUI screening programs. If one can estimate, a priori, the cost of a miss and a false-positive along with the base rate of alcoholic disease in the population, then one can easily compute the optimal test cut-off score in Appendix B. For example, the cost of a false-positive might be equated to the cost of an average treatment program. Unfortunately, estimating the cost of a miss is much more arbitrary. The virtue of the cost-ratio approach lies not in its ability to prescribe a specific cut-off, but in its emphasis on the need to translate community values into operational criteria that control the direction and scope of DUI screening programs.

Two methodological limitations of the aforementioned analysis are worthy of mentioning. The first is the lack of independence between the predictor and criterion measures in the Kentucky study. Both the DRI scale and the CPR interview incorporate items relating to the offender's record of previous DUI convictions. However, this confounding does not seem to have affected this study's findings. In a recent analysis of two statewide adoptions of the DRI, it was found that deleting previous offense status from computations of DRI scores affected $<2\%$ of the determinations of high and low risk (Davignon, personal communication with Behavior Data Systems, November, 20, 1995). This variance is well within the error bars depicted in Fig. 1. Similarly, the near-chance accuracy achieved with the CPR interview would even be lower were questions on previous DUI arrests removed from the interview.

A second methodological concern relates to the specification of the criterion measure. After the procedure in the Kentucky study, we used court records in the analysis to classify "known" cases (rather than statements made to the interviewer). For the reasons outlined herein, we believe that the present concurrent validity design constitutes an acceptable methodology. In the field of psychological assessment, Lanyon and Goodstein[16] observe that validity studies often predict to a *concurrent* event or *postdict* to an earlier event. They also point out that these approaches are especially useful when the criterion (a) uses a different modality than the predictor and (b) cannot be measured directly without considerable effort. Both situations characterize the present evaluation study. In medical investigations of the effectiveness of drugs and devices, concurrent designs are often used in studies in preparation for large-scale clinical trials.[17] Recently, Lapham et al.[18] addressed many of these psychometric issues and recommended recidivism as the standard of reference in DUI validity studies. We agree that a prospective validity study is the method of choice, but wish to point out that useful information can often be obtained in studies that use the "next-best" method to the gold standard, namely a concurrent-validity design. Finally, we emphasize that the principal goal of this study was not to present validity data in support of a particular screening instrument. Our aim was to illustrate how decision theory may be applied to the evaluation of DUI screening programs.

In this study, we have attempted to show that decision theory is a helpful analytical tool for assessing DUI offenders. First, the ROC's sensitivity parameter (d') allows for an "apples against apples" comparison of the overall accuracy of alternative screening instruments. Second, the ROC provides a graphic representation of examiner bias. For rating-scale methods, this parameter can be readily adjusted to reflect changes in base rates and the costs and benefits of decision outcomes. Toward this end, we introduced a new measure of bias, called the "cost ratio," that relates the perceived cost of a false-positive to that of a miss. It is in its quantitative specification of examiner bias that decision theory offers the potential of increasing informed decision-making in screening DUI offenders.

ACKNOWLEDGMENTS

Peter Killeen, John Reich, David MacKinnon, Craig Nagoshi, and Terri Asher provided assistance in commenting on this manuscript, and their valuable contributions are hereby acknowledged.

APPENDIX A

The following illustrative items represent a partial summary of the DRI administered in the Kentucky study (DRI Copyright © 1987):

Section 1 (True/False)

1. I have a drinking (beer, wine, or liquor) or drinking-related problem.

2. There have been times, while driving, when I have not paid proper attention to what I was doing.

3. Within the past 3 years, I have been involved in a motor vehicle accident that was my fault.

79. I consider myself to be a "low-risk" driver, because I do not have accidents or get tickets.

80. There have been times when I have felt jealous or resentful of others.

The following illustrative items represent a partial summary of the Fayette County DUI Assessment Interview (CPR interview) administered in the Kentucky study:

DSM-III-R Assessment Criteria

1. Loss of Control (yes/no/undetermined, assessed over one's lifetime)
   (D1) Drink/use more than intended
   (D2) Persistent, strong urge to drink/use

2. Disruption of Functioning/Chemical Lifestyle
   (D3) Estimate hours per day under the influence and/or recovering from use
   (D4) Under the influence at school or work
   (D5) Social, job, or recreational activities and obligations given up or reduced
   (D6) Family arguments about drug use and/or related behavior

3. Tolerance? Physical Involvement
   (D7) Able to consume large quantities without significant effect
   (D8) Have you had . . . (physical symptoms) after stopping drinking/using for several hours
   (D9) Do you drink or use to start your day?

APPENDIX B

Computation of the optimal criterion (C) for any decision problem takes into account the following quantities: base rates (P) of the two conditions; and the benefits (B) and costs (C) of correct (true-negatives and true-positives) and incorrect decisions (false-positives and false-negatives or misses):

$$C = \frac{P\,(\text{neg})}{P\,(\text{pos})} \times \frac{(B_{TN} - C_{FP})}{(B_{TP} - C_{FN})}.$$

To compute the cost ratio, C_{FP}/C_{FN}, we assume that the perceived benefits (B) are much smaller than the perceived costs (C). This simplifying assumption seems warranted when it is realized that, in the screening of DUI offenders, the judicial system clearly emphasizes the need to reduce mistakes. Substituting into the aforementioned equation the base rates and the cut-off score corresponding to the observed true-positive and false-positive rates, we solve the equation for the cost ratio, C_{FP}/C_{FN}.

REFERENCES

1. Guarnieri M, Clayton RA: Assessment of DUI offenders in Fayette County, Kentucky: An evaluation using random assignment. Paper presented at the 10th annual Southeastern DUI/DWI offender systems conference. Myrtle Beach, SC, 1994
2. Clayton RR, Huffman M, Wester S, Caywood M: Final Report Assessment of DUI Offenders in Fayette County: An Evaluation Using Random Assignment. Lexington, University of Kentucky, Center for Prevention Research, 1994
3. Green DM, Swets JA: Signal Detection Theory and Psychophysics. New York, Wiley, 1966 (reprinted 1988, Los Altos, CA, Peninsula)
4. Leshowitz B: Receiver operating characteristics and psychometric functions determined under simple- and pedestal-detection conditions. J Acoust Soc Am 45:1474–1484, 1969
5. Swets JA: Measuring the accuracy of diagnostic systems. Science 240:1285–1293, 1988
6. Swets JA: The science of choosing the right decision threshold in high-stakes diagnostics. Am Psychol 47:522–532, 1992
7. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR: Variability in radiologists' interpretations of mammograms. N Engl J Med 331:1493–1499, 1994
8. Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman GL: Performance of screening and diagnostic tests. Arch Gen Psychiatry 44:550–555, 1987
9. Chen WJ, Faraone SV, Biederman J, Tsuang MT: Diagnostic accuracy of the Child Behavior Checklist Scales: A receiver operating characteristic analysis. J Consult Clin Psychol 62:1017–1025, 1994
10. Ross HW, Gavin DR, Skinner HA: Diagnostic validity of the MAST and the Alcohol Dependence Scale in the assessment of DSM-III alcohol disorder. J Stud Alcohol 51:506–513, 1990
11. Rice ME, Harris GO: Violent recidivism: Assessing predictive validity. J Consult Clin Psychol 63:737–748, 1995
12. Behavior Data Systems: The Driver Risk Inventory. Phoenix, Behavior Data Systems, 1987
13. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 3. Washington, D.C., American Psychiatric Association, 1980
14. Kahneman D, Slovic P, Tversky A: Judgment under Uncertainty: Heuristics and Biases. New York, Cambridge University Press, 1982
15. Blackstone W: Commentaries on the Laws of England, vol 4, p 27, London, Dawsons of Pall Mall, 1966
16. Lanyon RI, Goodstein LD: Personality Assessment. New York, John Wiley & Sons, 1982
17. Budoff MJ, Georgiou D, Brody A, Agatson AS, Kennedy J, Wolfkiel C, Stanford W, Shields P, Lewis RJ, Janowitz WR, Rich S, Brundage BH: Ultrafast computed tomography as a diagnostic modality in the detection of coronary artery disease. Circulation 93:898–903, 1996
18. Lapham SC, Skipper BJ, Owen JP, Kleyboecker DT, Thompson B, Simpson G: Alcohol abuse screening instruments: Normative test data collected from a first DWI offender screening program. J Stud Alcohol 56:51–59, 1995