

NBER WORKING PAPER SERIES

USING TARGET EFFICIENCY TO SELECT PROGRAM PARTICIPANTS
AND RISK-FACTOR MODELS: AN APPLICATION TO CHILD MENTAL
HEALTH INTERVENTIONS FOR PREVENTING FUTURE CRIME

David S. Salkever
Stephen Johnston
Mustafa C. Karakus
Nicholas Ialongo
Eric Slade

Working Paper 12377
<http://www.nber.org/papers/w12377>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2006

Partial support for this work was provided under National Institute of Mental Health Grants P30 MH066247-01 and MH042968. Valuable comments on an earlier version were provided by Marv Mandell, Dave Marcotte and Liz Stuart. The authors are responsible for any errors or inaccuracies. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2006 by David S. Salkever, Stephen Johnston, Mustafa C. Karakus, Nicholas Ialongo and Eric Slade. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Using Target Efficiency to Select Program Participants and Risk-Factor Models: An Application to Child Mental Health Interventions for Preventing Future Crime

David S. Salkever, Stephen Johnston, Mustafa C. Karakus, Nicholas Ialongo and Eric Slade

NBER Working Paper No. 12377

July 2006

JEL No. I12, D61

ABSTRACT

Statistical risk factor models are often proposed for screening high-risk children to participate in early intervention programs. Recent contributions to the program evaluation literature demonstrate the need for incorporating judgments about relative importance of false positives versus false negatives in screening. This paper formalizes these judgments as commensurable economic costs and benefits and applies them to demonstrate an approach to participant selection motivated by the standard cost-benefit criterion of maximizing expected net benefits. Implications of this approach are explored using data from a mental health prevention trial. We illustrate the response of expected net benefits to the choice of a selection risk level, the sensitivity of the optimal selection risk level to per participant cost/benefit magnitudes, and the use of the target-efficiency approach for choosing among alternative risk-factor models. Several strategies that directly incorporate expected net benefit maximization as a criterion in the model estimation process are also examined.

David S. Salkever
Department of Public Policy
University of Maryland, Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
and NBER
salkever@umbc.edu

Stephen Johnston
Department of Public Policy
University of Maryland, Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
sjohns21@umbc.edu

Mustafa C. Karakus
Westat
1650 Research Boulevard
Rockville, MD 20850
mustafakarakus@westat.com

Nicholas Ialongo
Department of Mental Health
Johns Hopkins Bloomberg School of Public Health
Hampton House, Room 809
624 North Broadway
Baltimore, MD 21205
nialongo@jhsph.edu

Eric P. Slade
Department of Veterans Affairs
VISN 5 MIRECC, Baltimore
University of Maryland, School of Medicine
737 West Lombard Street, Room 526
Baltimore, MD 21201
eslade@psych.umaryland.edu

Introduction

Statistical models of risk factors have often been proposed or used for identifying high-risk children as participants for intervention programs (1-3). These models include bivariate associations of individual risk factors with undesirable outcomes, as well as regression models that include multiple risk factors. Recently, Kraemer et al. (4) examined the application of a variety of purely statistical performance criteria for such models and have stressed the importance of incorporating expert judgments on the clinical and policy significance of the consequences of false positive and false negative classifications. In particular, they argue that such expert judgments are necessary in order to select the optimal statistical test for any specific intervention program. While the examples presented in Kraemer et al. (4) were limited to bivariate associations, Kiernan et al. (5) suggested that such judgments could also be accommodated in multiple risk factor models estimated by regression tree (recursive partitioning) methods. Another recent application of this method, presented in Berk et al. (6), combines recursive partitioning with expert judgments (by police officials) of relative costs of false negatives versus false positives in responding to domestic violence calls. A recent paper by Menditto et al. (7) tests the performance of a logistic regression model of risk of elopement by state psychiatric hospital patients in which false negatives are more costly than false positives.

Taking the Kraemer et al. (4), Berk et al. (6), and Menditto et al. (7) arguments as a point of departure, in this paper we replace expert judgments on relative significance or costs of prediction outcomes with quantitative estimates of expected economic costs and benefits of the intervention. We explore how these estimates, in combination with the basic economic principle of expected net benefit maximization, can be applied to the problem of selecting the optimal predictive test for an early intervention program to prevent adult crime. To signify the parallel

between this approach to targeting a program, and the generally accepted view in cost-benefit analysis that economic efficiency is achieved by program choices that maximize benefits minus costs, we shall refer to our procedure as a “target-efficiency” method of participant selection. (Note that the term was originally applied (8,9), for analogous reasons, to alternative formulae for redistributing income to poor families via taxes and income transfers.)

Using illustrative data, we examine the sensitivity of optimal rules for selecting high-risk participants to changes in expected economic cost and benefit figures, and we illustrate the implications of these figures for choosing among alternative risk factor models. We begin by describing the context for our analysis. Then we consider the problem of finding an optimal level of “selection risk” which we define as the high-risk selection threshold for assigning participants to the intervention. An empirical demonstration of a solution to this problem is then presented, along with an examination of the sensitivity of the optimal selection risk level to changes in expected intervention costs and benefits. This is followed by exploration of several extensions of our empirical example in which we apply our target-efficiency approach to the problem of choosing among alternative risk-factor models. The paper concludes with a summary and briefly discusses limitations of the proposed methods and priorities for further research.

I. The Context for the Analysis

Suppose we are considering an early-intervention crime-prevention program to be applied to a target group of potential participants made up of two subgroups, positives and negatives. The former are children who, in the absence of our intervention, would in fact become criminals as adults, and the latter are children who would not become adult criminals. We seek a test for targeting our intervention to the children who are at high risk of being positives. We can

conceptualize each test as consisting of two parts: 1) a process for predicting the risk of a positive outcome and 2) a selection risk level that differentiates high vs. low levels of this risk.

Of course, since we can not directly observe the adult outcomes of the children in the target group, any empirically-based process for predicting risk will require analysis of a retrospective “control” data set which contains both data on observable risk factors that are also available for our target group and data on outcomes. The “control” population described in this data set should be similar in observable risk factors to the target population and should not have been subjected to any intervention.

Based on this analysis of a retrospective “control” data set, we can evaluate the performance of any particular test with reference to the outcomes observed in that data set (plus other retrospective control data sets if available). As in Kraemer et al. (4), Berk et al. (6), and Menditto et al. (7), the focus of our exposition here is on the criteria used in that evaluation.

II. Finding an Optimal Level of Selection Risk

Within the context just described, let us first assume that a process for predicting the risk of positive outcomes based on observable risk factors has been arrived at, so that all that remains in designing our test is to choose the selection risk level at which assignment to the intervention warranted. As suggested above, the procedure for making this decision is based on applying any proposed selection risk level to empirical evaluation with a retrospective control data set. In the clinical diagnosis literature, the criteria used in this evaluation are typically the resulting sensitivity and specificity of the proposed test. Kraemer et al. (4) have discussed the relationship between these criteria and the purely statistical criteria that are usually applied in other disciplines (epidemiology, sociology, psychology). They have demonstrated that statistically equivalent tests can vary widely in terms of sensitivity and specificity; hence the need for

additional expert judgment of clinical and policy significance of different types of errors (false positives vs. false negatives). Menditto et al. (7) provide an example in which the risk level is chosen to minimize the number of false positive plus false negatives, and compared this prediction performance with those obtained by several lower selection risk levels that implicitly incorporate judgments that false negatives are more costly than false positives.

In this analysis, we explicitly represent these judgments in terms of 1) the expected economic costs of the intervention program and 2) the expected economic benefit of enrolling each positive child in the program. It is assumed that knowledge of these expected costs and benefits has already been obtained from previous evaluation studies. From this economic perspective, the intuition involved in finding the optimal level of high risk can be described in simple terms. If we choose a very high selection risk level, only a small number of children would be referred to the intervention. This has the advantage of keeping intervention costs low but it has the disadvantage of effectively treating only a small number of children who are positive (i.e., who will become adult criminals in the absence of treatment). Conversely, setting the selection risk level at a low threshold will result in higher costs but will also treat a larger number of children who are positive. A rule for trading off of these concerns is the principle of maximizing expected net benefit, which is a fundamental concept in the economic theory of cost-benefit analysis. Applying this rule, the optimal level of selection risk is defined as the level that balances these concerns of costs and benefits and results in the largest expected net benefit of the intervention (as evaluated with the control data).

A formal mathematical statement of the problem is straightforward. Let P and N be the number of positives and negatives in the control data set used for evaluating the test. Let X_i denote the vector of observable risk factors of the i th child in the control data and let β be a

vector of regression coefficients estimated from the control data set by regressing the risk factors X on the dichotomous outcome (1 = positive outcome, 0 = negative outcome). Given the form of the regression relationship (e.g., multiple probit, multiple logistic, or linear probability model), the estimated values for β , and the values in X_i , we can compute π_i , the predicted probability that the i th child in the control data set would have a positive outcome. Replicating this process for each child in the control population, we obtain a P -element vector of predicted probabilities for the positives in the control population (Π_P), and an N -element vector of predicted probabilities for the negatives (Π_N). Then for any proposed selection risk level of predicted probability, π^* , the i th child in the control data would be designated as high risk only if $\pi_i > \pi^*$. Thus, the number of positive children designated as high risk (i.e., “true positives”), P^* , would be equal to the number of elements of $\Pi_P > \pi^*$; N^* (i.e., the number of “false positives”) would be defined analogously.

Applying this risk level in a simulated intervention with the control population, the expected costs of the intervention would be $(P^*+N^*)C$, where C is expected intervention cost per child. The expected benefit would be BP^* , where B is the expected benefit per true positive child treated in the intervention. Thus the expected net benefit corresponding to π^* would be $BP^* - (P^*+N^*)C$. With known values for B and C , we can compute expected net benefit values for each possible level of selection risk (π^*) and thereby find the level of selection risk that maximizes expected benefit in the simulated intervention.¹ (Note that we do not assume that the intervention yields benefits for each true positive child. Instead, B represents the expected benefit for a randomly selected true positive child. For true negative children, we assume that the intervention yields no benefit.)

III. An Empirical Example

To demonstrate the process of finding the optimal level of selection risk (π^*), we use data from the control population of cohorts 1 and 2 of the Johns Hopkins Prevention Intervention Research Center's (JHU PIRC) Baltimore intervention trials. The cohorts were recruited in 1985 and 1986 from 43 first-grade classrooms in 19 elementary schools located in 5 socio-demographically distinct areas in eastern Baltimore City. The numbers of children in the two successive cohorts were 1,196 and 1,115. (For information on the project, the characteristics of the children, the interventions, and the data content and collection processes, see the project website <http://www.bpp.jhu.edu/Cohort3/methods.measures.young.adult.followup.word.htm> (accessed on December 6, 2005).) In our analysis, we only examine male students in the control group populations. (Females were excluded due to their very low rate of subsequent adult incarceration as shown in data collected by the JHU PIRC.)

Descriptive statistics on the data and variables used in our empirical analysis are shown in Table 1. In our initial analysis of this example, we will focus on the outcome, socio-demographic, and first-grade variables and on the 542 male control group members with data sufficiently complete to allow construction of these variables. (A subsequent section of the paper will look more closely at the third-grade and sixth-grade variables.) Note that the first row of Table 1 describes our outcome variable, a 0-1 dichotomy indicating whether a child was subsequently incarcerated in the adult correctional system. The table indicates that 16.6 percent of our study group had in fact been incarcerated at least once in the adult system as a young adult (by age 26).

Table 2 presents the results of maximum likelihood probit regressions of our outcome variable on the socio-demographic risk factors and first-grade school rating risk factors shown in

Table 1. In our initial discussion, we will focus on the results obtained with the linear model (shown in Columns 1 and 2).

In addition to these empirical results about risk factors, critical inputs for our analysis are the assumptions made about the dollar magnitudes of expected intervention cost per child included in the program (C) and the expected benefit resulting from enrolling each positive child who would be incarcerated in the future in the absence of the intervention program (B). Note that B can be viewed as the product of two figures, the change in the probability of incarceration as a result of the program and the benefit of avoiding incarceration and the crime that resulted in the incarceration.

For our illustrative example, figures for C and B were based on the results of the Seattle Social Development project as reported by Aos et al. (10). This project was a three-part intervention for teachers, parents and students in grades 1-6 that focused on schools in high crime urban areas. Cost-benefit evaluation of the project yielded an estimated cost per participant of \$4,355; we use this as our assigned value for C. They do not report a specific value for B but instead provide a figure for estimated net benefit per participant (including both positives and negatives). This figure is -\$456. (Note that it is restricted to benefits to taxpayers only; their corresponding figure that includes crime victim benefits is +\$14,619.) Converting this figure to a gross benefit per participant by subtracting out costs yields a figure of \$3,899. Since Aos et al. do not report an incarceration rate for the control group, we assume for purposes of our exposition that approximately 20 per cent of treated children in the program were positives (i.e., would have been incarcerated in the absence of the intervention), thus yielding a value for B of approximately \$20,000 for gross benefit per positive participant.ⁱⁱ

It is assumed here that the populations served by the Seattle project are similar to the Baltimore population in our example and that the estimates of C and B can therefore be applied in the example. Using the results from Table 2, Columns 1 and 2, we can compute a predicted risk-level (i.e., probability of incarceration) for each person in our Baltimore data set. Given these predicted risk-levels, we can use our values for C and B to simulate the expected program net benefit of alternative selection-risk levels. The results of this process are reported in Figure 1.

As intuition would suggest, expected net benefit is strongly affected by the selection risk level, rising from -\$560,410 at a level of 0.0 (i.e., everyone is enrolled in the intervention) to +\$468,525 at a level of 0.25 and then declining to -\$4,355 at a level of 0.7. Thus, 0.25 is the optimal selection risk, though the variation in expected net benefit over the selection risk range of 0.24 to 0.28 is quite small. (The discontinuities in the net benefit function presumably arise because the number of positives in our data is fairly small and small changes in the selection risk level may result in discrete changes in the number of positives selected for program participation.)

Kraemer et al. (4) suggest that the optimal test procedure will depend on the relative weights (reflecting clinical and policy significance) of false positives and false negatives. The analogous observation in the current context is that the optimal selection rule will vary with the relative levels of C and B. We examined the sensitivity of the optimal selection risk level in our empirical example by allowing the level of B to vary holding C constant. (In our simple example, as in earlier studies (4-6), the ratio of B to C is sufficient to determine the optimal selection risk level.) The results of this exercise are shown in Figure 2. As expected the optimal selection risk level is negatively related to the ratio of B to C, dropping sharply from 0.9 when the B/C ratio increases above 1.65, and then declining more gradually as the B/C ratio continues to increase.

IV. Selecting Risk-Factor Models to Maximize Expected Net Benefit

Obviously the target-efficiency approach can be used not only for choosing the optimal selection risk level with any given risk-factor model, but also for choosing among alternative risk-factor models. If we confine ourselves to single risk factors and bivariate associations with outcome, as in Kraemer et al. (4), the choice involves comparing the maximum expected net benefit levels pertaining to each of the alternative candidate risk factors (e.g., teacher rating of educational progress vs. peer reports of fighting). In the case of choosing among alternative multiple risk-factor models, the choice involves comparing the maximum expected net benefit levels pertaining to each of the candidate models.

As a simple example, we consider choosing between the simple linear probit model in Columns 1 and 2 of Table 2 and the models in Columns 3-4 and Columns 5-6 that allow for nonlinearities in the relationships of some risk factors to the probit index. (We shall refer to the latter two respectively as the nonlinear and reduced non-linear models.) Typically, such a choice between models would be made with reference to likelihood-ratio statistics or other measures of goodness of fit of the various models. However, if the objective is to devise a method for selecting intervention program participants that maximizes expected net benefit, these usual statistical criteria are not pertinent.

The comparisons of the three models yield the following results:

	Max. Exp. Net Benefit	Optimal Selection Risk
Linear	\$468,525	0.25
Nonlinear	\$472,075	0.26
Reduced Nonlinear	\$483,200	0.22

While one might expect that adding the most additional parameters to the model (as in the nonlinear model) results in the highest maximum expected net benefit, we do not confirm this in our example. The reduced nonlinear model produces the largest maximum expected net benefit. Note however, that differences between the models in net benefit and in optimal selection risk levels are small. Indeed, the differences are so small that it is reasonable to consider whether these differences are simply due to random factors. One straightforward way to assess this possibility is to use bootstrap replications of the overall process with the control data set (including both the estimation of regression coefficients and the computation of the optimal selection risk level for each model). This process would generate a distribution of maximum expected net benefits corresponding to each model and standard tests for differences in distributions across models could be applied. One could also test for differences in the optimal selection risk levels across the models.ⁱⁱⁱ

The expected net benefit criterion can also be used to choose the timing of the intervention. In the current example, we can use the approach to address the question of whether the program should be implemented at first grade or at a later stage in children's development. For this purpose, we modify our multiple probit regression models to include data from the Baltimore data set on third-grade and sixth-grade teacher ratings of the children.

Results for the linear probit regressions that add third grade ratings are shown in Table 3, Columns 3 and 4 below. Results obtained when sixth grade ratings are also added are shown in Table 3, Columns 5 and 6. For comparison purposes, in columns 1 and 2 we repeat the first-grade-only results from Table 2, columns 1 and 2. A comparison of these three models can help us to understand how the expected net benefits of the intervention program might change if it were implemented in third grade or in sixth grade rather than in first grade. One might expect

that having additional information from teacher ratings in later grades would improve the accuracy of our models so that the third-grade or sixth-grade interventions could be targeted more efficiently. Note, however, that such a comparison should also reflect the changes in C and B that may occur if the implementation is delayed to a later grade. Such changes will be expected if the content of the program changes, if the effect of the program on students' future course is altered by the delayed implementation, and simply to reflect the basic economic fact that the present value of costs decline as implementation is delayed due to discounting. (Changes in discounting for benefits arising from crimes averted are not indicated, however, unless the timing of these benefits is altered by the delay.) In our discussion, however, we ignore possible changes in B and C between the models to focus purely on the gain in expected net benefits arising from superior predictive power of the third grade and sixth grade models.

The results in Table 3 indicate that the teacher rating of aggression in third grade (SCTAG3) has coefficients that are statistically significant but that the significance of the coefficient for first-grade peer rating of aggressive behavior (PERF) drops when this variable is included. (The loss of 92 cases due to missing data may also influence this result.) Most of the other results for specific variables do not change very much as the third-grade and sixth-grade variables are added in. (Note however that the number of missing cases increases in sixth grade and the number of cases in the regression drops to 366.)^{iv}

With respect to the results of principal interest, comparisons of the three models are shown in Table 4 below. A very large gain in expected net benefits is obtained by adding third-grade variables to the risk-factor model. Further addition of sixth-grade variables yields a slightly lower expected net benefit figure. The implication of the results, however, is that expected net benefits of the program may be increased substantially by implementing in third grade (rather

than first grade) and having additional information available on risk factors. (Changes in costs or effectiveness as a result of later implementation could either weaken or strengthen this conclusion.) Finally, for both of the new models we also find a lower level of the optimal selection risk compared to the result for the first-grade model.^v

V. Maximizing Expected Net Benefit as a Criterion for Model Estimation

At this point, the discerning reader may be wondering why we use a purely statistical criterion (e.g., maximization of a likelihood function) for obtaining coefficient estimates of our risk factor model, while we use an economic criterion (maximization of expected net benefits) for choosing the optimal level of selection risk. While the statistical criterion may imply desirable statistical properties (such as consistency or asymptotic unbiasedness of coefficient estimates), and produce results that are easily interpretable (e.g., estimates of the probability that a particular participant will in fact turn out to be a positive), the statistical criterion may also conflict with the economic criterion. This is likely to be the case since the statistical criterion treats false positives and false negatives symmetrically while the two types of errors may have very different implications for program net benefits.

These observations suggest that it may be interesting to explore alternative strategies for estimating the coefficients of parametric risk-factor models. As a matter of interpretation, the results of these estimates will produce an index value for each individual (which we shall call the “risk index”) but this value is no longer interpretable as an estimate of the probability of turning out to be a positive.

One other advantage of using a purely statistical criterion to estimate coefficients of a parametric risk factor model is that the computation of these estimates can usually rely on

standard techniques for finding maximum values of continuous functions. Computational procedures may be more challenging when we depart from this criterion.

To examine the possible economic gains of extending our target-efficiency approach to include estimates of coefficients in parametric multiple risk-factor models, we illustrate two possible methods. First, we consider a relatively simple modification of our previous analysis. Instead of estimating a probit multiple risk-factor model with an unweighted likelihood function as our statistical criterion, we estimate this probit model with a weighted likelihood function that gives greater weight to those cases that were in fact positives than to those case that were in fact negatives. The logic here is that because of the relative size of C and B, it is more important to do a good job of identifying the positives as true positives than it is to identify the negatives as true negatives; therefore the former cases get greater weight in the likelihood function. How much greater should this weight be? We use our target-efficiency criterion to answer that question; in other words, we search over alternative values for this weight to find the value that (in combination with the optimal selection level of the risk index based on the coefficient estimates corresponding to that weight) produces the largest expected net benefit.

Applying this method to our linear probit model with first-grade data, we allowed the weight for positives (relative to negatives) to range from 1 to 4 and computed the maximum expected net benefit corresponding to each value of the weight. The results of these calculations are summarized in Figure 3. It is clear that the varying the size of the weight over the range 1.0 to 4.0 does not result in very substantial changes in the level of net benefits. For example, moving from a weight of 1.0 to the weight of 1.3 (at which expected net benefits are maximized) only corresponds to a 3.4% increase in expected net benefits (from \$468,525 to \$483,365).

The second method that we illustrate is a straightforward but cumbersome search process. We search over all relevant sets of possible values for the coefficients of the risk-factor model, computing the optimal selection level of the risk index for each set of coefficients and computing the maximum expected net benefit level corresponding to that optimal selection risk level, until we find that set of coefficients with the largest maximum expected net benefit value.

Since our economic criterion is not a well-behaved, continuous function of the coefficient values, the computational burden of this approach is potentially much greater than that in our previous illustration. As we are undertaking these computations for illustrative and expository purposes, we have made several assumptions that simplify the process considerably. One important assumption in our example is that only two risk factors are assumed to be relevant predictors: the peer rating of fighting in first grade and the teacher rating of educational progress in first grade. Second, since we are estimating a parametric risk-factor model, we assume a specific functional form for the risk-index function; in particular, we assume a probit functional form with a risk-index that is linear in the two risk factors. Thus, three coefficients will characterize our risk-index function (an intercept and a coefficient for each of the two risk factors).

Even with these simplifications, the development of a reasonably efficient computer program for conducting the search process is not a simple matter and we did not attempt this for the current paper. Instead, to illustrate the potential usefulness of this method we have confined our calculations to only 36 alternative sets of coefficient values for the risk-factor model. For comparison purposes, one set of coefficient values are the maximum-likelihood estimates obtained from a probit regression of the outcome on our two risk factors. We then allowed each coefficient to take on values that were 50%, 100%, 150% and 200% of those maximum-

likelihood estimates. Once we identified what appeared to be a global maximum, we then undertook a local search around this apparent maximum to locate the maximum more precisely. The result of this process is reported in Column 1 of Table 5. Column 2 of this table reports the results from the standard maximum-likelihood estimation for comparison purposes.

Comparing the results in Column 1 versus Column 2, we see that the maximum expected net benefit value obtained by the search process (\$190,940) is almost 44% larger than the value obtained from the maximum likelihood probit regression (\$132,735). This suggests that using the target efficiency criterion for selecting coefficient values in the risk-factor model may produce substantial gains over the values obtained using the purely statistical criterion of likelihood-function maximization.

We do note, however, that our two-factor probit model leaves much room for improvement; the expected net benefit is far below the level (reported above) that we obtained when we included a number of additional variables in the probit regression (\$468,525). Thus, the relative gain from the target-efficiency approach to coefficient estimation could be considerably smaller with a model that included more risk factors. In addition, the computational burden of the search process becomes much greater when additional variables are added to the risk-factor model.

A further aspect of the risk-factor analysis that could be reconsidered in the light of the target-efficiency criterion is the matter of functional form. For conceptual reasons or reasons of convenience, an empirical analysis that seeks to assess the importance of various risk factors typically involves the use of a parametric functional form, such as the linear probit model that we have employed throughout our discussion. If, however, our goal is to maximize expected net benefits, it may be preferable to use a nonparametric or semiparametric approach instead. One

promising nonparametric approach is the recursive partitioning method applied by Kiernan et al. (5) and Berk et al. (6), that uses a criterion (for selecting optimal partitions) that corresponds to the weights derived from our values for B and C.^{vi}

With the ROC4 software used by Kiernan et al. (5), we determined the optimal split of our 542 control subjects into three groups based on the values for the same two risk factors as in our probit analysis.^{vii} The groups were defined as follows: Group 1 – $PERF < 0.19$, Group 2 – $PERF \geq 0.19$ and $TOCGB1 < 4$, Group 3 - $PERF \geq 0.19$ and $TOCGB1 \geq 4$. The corresponding percentages of true positives were 8.4, 17.6 and 31.0 respectively. In this type of analysis, choosing the optimal risk cutoff corresponds to choosing the groups to enter the intervention that will maximize expected net benefits. In the present case, our expected net benefits were maximized, by choosing only Group 3; the corresponding net benefit figure is shown in Table 5, Column 4 to be \$216,595. This is 13.4 per cent larger than the optimal value from our two-factor probit model shown in Table 5 above. For comparison purposes, in Column 3 of Table 5 we show the result obtained when the partitioning criterion is based on a weight of 0.5 (equivalent to assuming Type 1 and Type 2 errors are of equal importance). The expected net benefit in this case is nearly as large as when the weight is 0.8.^{viii}

This example suggests that nonparametric recursive partitioning provides a modest gain over our parametric search procedure presumably by relaxing the functional form constraint of the parametric approach. The non-parametric result also appears to be less sensitive to the choice of weights.

The recursive partitioning method is computationally simpler than our parametric search method. In the latter, the range of possible coefficient values is not bounded and the number of possible points in the search is infinite; in the former the process involves a finite number of

possible partitions. Both methods may involve substantial computational time when implemented on a typical personal computer with large numbers of risk factors included in the model.

One other limitation of the recursive partitioning model is that it assumes the expected net benefit function is simply the sum of the expected net benefits over all the subjects assigned to treatment. (The parametric search model could use any function of the numbers of false and true positives as a criterion.) This limitation may be problematic, for example, if the cost of implementing the intervention is subject to economies or diseconomies of scale or discontinuities based on the total number of treated participants (e.g., constraints on the size of a single classroom). Program expected effectiveness per participant might also depend on the total number of participants. (For example, a larger group size may be less effective for each participant.)

Comparing our results in Table 5 with the earlier results in Table 2, our examples also suggest the possibility that either of the methods used in Table 5 would yield substantially larger expected net benefits as additional parameters and (especially) risk factors are added to the models. Naturally this extension would greatly increase the computational burden. The earlier results also suggest that when additional risk factors are considered, allowing for flexibility in the functional form of the risk-factor model may become less important.^{ix}

Finally, note that use of the target-efficiency criterion for finding an optimal multiple risk-factor model does not make statistical considerations irrelevant. One could ultimately devise an extremely complex model that achieved the maximum possible expected net benefits by perfectly sorting out positive from negative cases. Such a model would probably use a complicated functional form and contain many predictor variables so that the number of

coefficient values to be estimated (or data partitions to be formed) became very large relative to the number of subjects in the control data set. While such over fitting of the data might maximize our target efficiency criterion, it would also produce estimates that were highly sensitive to random influences in the process that generated the control data set and would probably generalize very poorly to additional control data sets or to the population for which the intervention is intended. The solution to this problem, as noted earlier, is to apply bootstrapping techniques to the entire process for searching out the optimal risk-factor model and risk-index values. Presumably this process would reveal extremely wide confidence intervals for the optimal point estimates when the model used is highly complex and the number of coefficients (or partitions) is very large.^x

VI. Summary and Conclusions

For the education or mental health professional responsible for implementing an early intervention crime-prevention program, selecting the children to participate in the program is an important issue that can strongly influence the net benefits realized from the program. If highly accurate risk-factor models were available for selecting students, and these models generated very large differences, between negative and positive children, in the predicted probability of a positive outcome (e.g., incarceration as an adult), the problem of selecting program participants would be fairly straightforward. It is rarely the case, however, that the available risk-factor models are highly accurate, and thus the rules for selecting participants on the basis of these models becomes a matter of concern.

Kraemer et al. (1) have argued persuasively for the inclusion of expert judgments about the clinical and/or policy significance of false positives and false negatives in devising rules for assigning participants to treatments. Empirical application of this idea in a multiple risk-factor

context has been demonstrated by Berk et al. (6) and Menditto et al. (7). Building on this previous work, this paper has illustrated that when these judgments are expressed as commensurable economic costs and benefits, we can use a target-efficiency approach for participant assignment that is motivated by the standard cost-benefit criterion of maximizing expected net benefits. We have explored the potentially important implications of this approach in the context of an empirical example using data from the control groups of cohorts 1 and 2 of the JHU PIRC trials and estimates of per subject costs and benefits from Aos et al. (10). Our example illustrates that expected net benefits are quite sensitive to the risk level chosen, with nearly five-fold variation as the selection risk level ranges between 0.1 and 0.5. Not surprisingly, the optimal selection risk level is also somewhat sensitive to the relative magnitudes of per participant costs vs. benefits.

We also illustrate the use of the target-efficiency approach for choosing among alternative multiple risk-factor models. In our example, the strongest differences between alternative models (in their ability to maximize expected net benefits) appears to arise from the inclusion or exclusion of predictors that are strongly related to the outcome we seek to prevent. This finding may point toward the inclusion of some risk factors in the selection model that may not meet the usual strict criteria of statistical significance. This does, however, raise concerns about over fitting that could be assessed by computing bootstrapped confidence intervals around the maximum expected net benefit figures for each competing model.

In the final section of our analysis, we examine the possible gains of incorporating expected net benefit measures directly into the estimation process in either parametric or non-parametric approaches. To keep computational tasks manageable in this non-statistical approach, we restrict our attention to models involving only two risk factors. In the parametric case,

substantial gains in expected net benefits are observed (relative to the statistical likelihood-maximization approach) but the relative gain is much smaller in the particular non-parametric method applied (i.e., the recursive partition approach). We note that in both cases the reduced number of risk factors results in substantial declines in maximum expected net benefits relative to the more complete models examined in earlier sections of our analysis. While this might argue for expanding the number of risk factors included in these non-statistical models, we note that computational problems or over fitting problems (especially in the non-parametric case) may present us with trade-offs in choosing a risk-factor modeling strategy for maximizing target efficiency.

While we have made several strong simplifying assumptions here, it may be worthwhile to extend our test of the target efficiency approach to more complex situations. Examples are cases where outcomes are polychotomous rather than binary (e.g., no crime, non-violent crime, violent crime) and cases where the expected costs and/or benefits are not simply equal to the sum of individual-participant-level expected costs or benefits (e.g., because of program scale effects).

It must also be emphasized that the results presented here are drawn entirely from a single data set drawn from a particular geographic location. Evidence from many more settings and examples are clearly needed to reach an informed judgment about the usefulness of the target-efficiency approach to program participant and risk-factor model selection. This is especially so because the application of results from a population of “controls” to a new population of potential program participants is based on stringent assumptions about the similarities between the two populations. In particular, these populations need to be similar not just in the means of the variable values in our analysis but also in the joint distributions of these variables. Of course, the similarities of the “control” and potential participant populations could

be substantially enhanced by selecting from the control group, using propensity scores or other matching methods, to produce a sample that matches your population of potential participants.

Table 1. Variable definitions and descriptive statistics

Name	Definition	n	Mean	Std Dev.
<i>Outcome Variable</i>				
<i>PRISON</i>	= 1 if individual ever incarcerated	542	0.1660	0.3724
<i>Socio-Demographic Variables</i>				
<i>WHITE</i>	= 1 if white; else = 0	542	0.3726	0.4839
<i>CGLTHS</i>	= 1 if individual's caregiver education level is less than high school	542	0.3025	0.4598
<i>CGHS</i>	= 1 if individual's caregiver education level is high school	542	0.3542	0.4787
<i>MCGEDUCN</i>	= 1 if individual's caregiver education level is unknown	542	0.1033	0.3046
<i>EMPLDCG</i>	= 1 if individual's caregiver employed	542	0.4612	0.4989
<i>MEMPLDCG</i>	= 1 if individual's caregiver employment status unknown	542	0.1771	0.3821
<i>First Grade Variables</i>				
<i>PERF</i>	Percentage of peers that nominated individual for starting fights, grade 1 fall	542	0.2617	0.1969
<i>SCTCP1^{ab}</i>	Mean teacher rated attention/concentration level, grade 1 fall	542	3.2111	1.3343
<i>TOCGB1^{ab}</i>	Teacher's global rating of how individual is progressing as a student, grade 1 fall	542	3.1051	1.2804
<i>Third Grade Variables</i>				
<i>SCTAG3^{ac}</i>	Mean teacher rated aggressive disruptive behavior, grade 3 spring	450	2.2868	1.1548
<i>SCTCP3^{ac}</i>	Mean teacher rated attention concentration problems, grade 3 spring	450	3.2216	1.1420
<i>TOCGB3^{ac}</i>	Teacher's global rating of how individual is progressing as a student, grade 3 spring	450	3.0377	1.3644
<i>Sixth Grade variables</i>				
<i>SCTAG6^{ad}</i>	Mean teacher rated aggressive disruptive behavior, grade 6 spring	366	2.2928	1.0966
<i>SCTCP6^{ad}</i>	Mean teacher rated attention concentration problems, grade 6 spring	366	3.6244	1.2606
<i>TOCGB6^{ad}</i>	Teacher's global rating of how individual is progressing as a student, grade 6 spring	366	3.1967	1.2406

^a (1 = Almost Never ... 6 = Almost Always)

^b = Grade 1 spring score if missing

^c = Grade 4 score if missing, grade 4 = interpolated value if missing

^d = Grade 7 score if missing

Table 2. Probit Models with First-Grade Risk Factors for Probability of Incarceration

Variable	1 st Grade Linear		1 st Grade Nonlinear		1 st Grade Nonlinear Reduced	
	Coefficient	P > z	Coefficient	P > z	Coefficient	P > z
Column #	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>
<i>Demographic Variables</i>						
<i>WHITE</i>	- 0.911	0.000	- 0.967	0.000	- 0.950	0.000
<i>CGLTHS</i>	0.815	0.001	0.784	0.001	0.833	0.000
<i>CGHS</i>	0.601	0.009	0.595	0.010	0.636	0.006
<i>MCGEDUCN</i>	- 0.422	0.237	- 0.323	0.378	--	--
<i>EMPLDCG</i>	- 0.229	0.156	- 0.258	0.115	- 0.245	0.132
<i>MEMPLDCG</i>	0.501	0.095	0.362	0.245	0.246	0.376
<i>First Grade Variables</i>						
<i>PERF</i>	1.029	0.005	3.001	0.051	4.160	0.001
<i>PERF</i> ²	--	--	- 4.707	0.005	- 4.103	0.008
<i>SCTCPI</i>	- 0.171	0.077	- 0.592	0.105	- 0.446	0.115
<i>SCTCPI</i> ²	--	--	0.115	0.215	0.036	0.342
<i>TOCGB1</i>	0.329	0.001	0.677	0.083	0.344	0.001
<i>TOCGB1</i> ²	--	--	- 0.005	0.955	--	--
<i>PERF</i> x	--	--	- 0.004	0.994	--	--
<i>SCTCPI</i> x	--	--	0.429	0.412	--	--
<i>TOCGB1</i> x	--	--	- 0.116	0.450	--	--
<i>CONSTANT</i>	- 1.970	0.000	- 2.063	0.000	- 1.965	0.000
<i>N</i>	542		542		542	

Table 3: First vs. First and Third vs. First, Third and Sixth Grade Probit Models for Probability of Incarceration

Variable	1 st Grade Model		1 st & 3 rd Grade Model		1 st , 3 rd , & 6 th Grade Model	
	Coefficient	P > z	Coefficient	P > z	Coefficient	P > z
<i>Demographic Variables</i>						
<i>WHITE</i>	- 0.911	0.000	- 0.751	0.000	- 0.654	0.005
<i>CGLTHS</i>	0.815	0.001	0.857	0.001	0.995	0.001
<i>CGHS</i>	0.601	0.009	0.726	0.006	0.835	0.005
<i>MCGEDUCN</i>	- 0.422	0.237	- 0.446	0.472	--	--
<i>EMPLDCG</i>	- 0.229	0.156	- 0.154	0.359	- 0.080	0.667
<i>MEMPLDCG</i>	0.501	0.095	1.518	0.003	--	--
<i>First Grade Variables</i>						
<i>PERF</i>	1.029	0.005	0.484	0.268	0.403	0.414
<i>SCTCP1</i>	- 0.171	0.077	- 0.230	0.037	- 0.196	0.114
<i>TOCGB1</i>	0.329	0.001	0.314	0.006	0.299	0.015
<i>Third Grade Variables</i>						
<i>SCTAG3</i>	--	--	0.219	0.009	0.200	0.031
<i>SCTCP3</i>	--	--	0.051	0.535	0.013	0.330
<i>TOCGB3</i>	--	--	0.018	0.811	- 0.085	0.888
<i>Sixth Grade Variables</i>						
<i>SCTAG6</i>	--	--	--	--	0.095	0.303
<i>SCTCP6</i>	--	--	--	--	0.140	0.242
<i>TOCGB6</i>	--	--	--	--	0.044	0.689
<i>CONSTANT</i>	- 1.970	0.000	- 2.502	0.000	- 3.090	0.000
<i>N</i>	542		450		366	

<u>Linear Probit Model</u>	<u>Max. Exp. Net Ben.</u>	<u>Optimal Selection Risk</u>
First-Grade Only	\$468,525	0.25
First and Third Grades	\$666,684*	0.18
First, Third and Sixth Grades	\$643,181**	0.14
* Based on N = 450 prorated to N = 542.		
** Based on N = 366 prorated to N = 542.		

	(1)	(2)	(3)	(4)
	Target-Effic. Probit	ML Probit	Partitioning w. 0.5 Weight	Partitioning w. 0.8 Weight
Constant	-0.8908	-1.7816		
Coeff. PERF	0.7294	0.7294		
Coeff. TOCGB1	0.0926	0.1853		
Selection "Risk"	0.37	0.30	0.32	0.31
Expected Net Benefit	\$190,940	\$132,735	\$214,015	\$216,595

Figure 1: Relationship of Expected Net Benefits to Selection Risk-Level

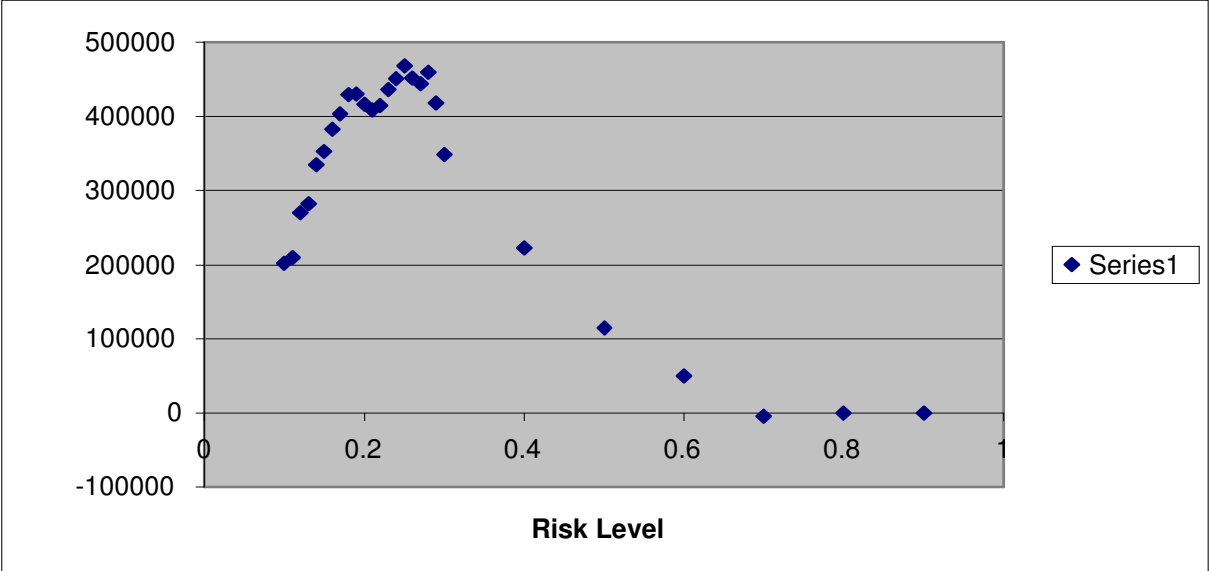


Figure 2: Optimal Selection Risk Levels for Varying B/C Ratios

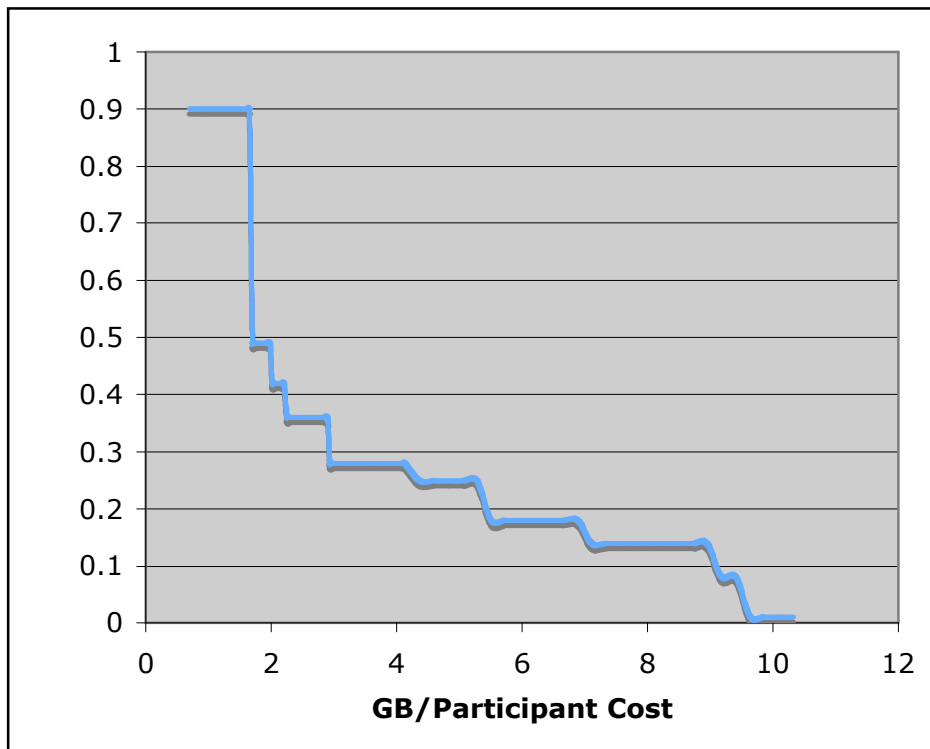
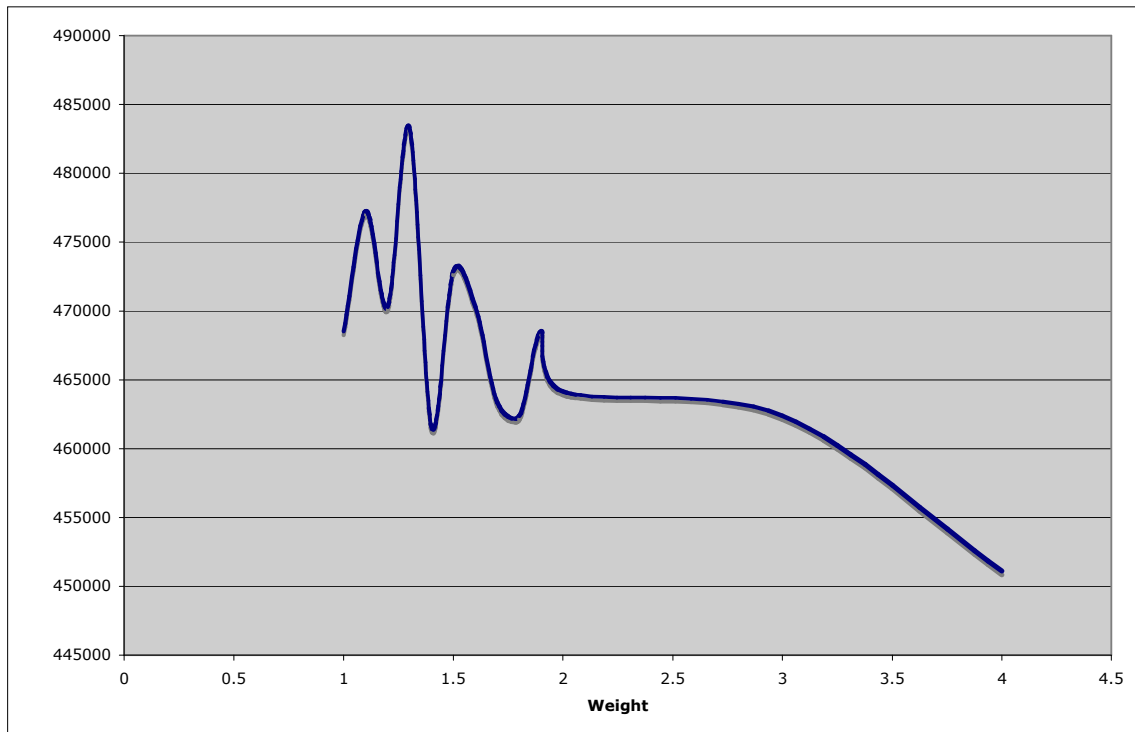


Figure 3: Maximum Expected Net Benefit as a Function of the Relative Weight for Positive Cases



REFERENCES

1. Hill, Laura G., Coie, John D., Lochman, John E., Greenberg, Mark T. Effectiveness of Early Screening for Externalizing Problems: Issues of Screening Accuracy and Utility. Journal of Consulting and Clinical Psychology 72:5 (Oct 2004): 809-820
2. Lochman, John E., Conduct Problems Prevention Research Group. Screening of child behavior problems for prevention programs at school entry. Journal of Consulting and Clinical Psychology 63:4 (Aug 1995): 549-559
3. Loeber, R., Dishion, T. Early predictors of male delinquency: A review. Psychological Bulletin 94:1 (Jul 1983): 68-99.
4. Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, and Kupfer DJ, Measuring the potency of risk factors for clinical or policy significance. Psychological Methods 4(3):257-271, 1999.
5. Kiernan M, Kraemer HC, Winkleby MA, King AC, and Taylor CCB. Do logistic regression and signal detection identify different subgroups at risk? Implications for the design of tailored interventions. Psychological Methods 6,1(2001):35-48.
6. Berk R, He Y, and Sorenson SB. Developing a practical forecasting screener for domestic violence incidents. Evaluation Review 29:4 (August 2005):358-383.
7. Menditto AA, Linhorst DM, Coleman JC and Beck NC. The use of logistic regression methods to enhance risk assessment and decision making by mental health administrators. Journal of Behavioral Health Services and Research 33:2 (April 2006): 213-224.
8. Garfinkel I. and Haveman R. Earnings capacity and the target efficiency of alternative transfer programs. American Economic Review 64(2):196-204 (1974).

9. Creedy J. Comparing tax and transfer systems: Poverty, inequality and target efficiency. *Economica* 63(250), Supplement: Economic Policy and Income Distribution, S163-174 (1996).
10. Aos S, Phipps P, Barnoski R and Lieb R. The Comparative Costs and Benefits of Programs to Reduce Crime (Version 4.0). Washington State Institute for Public Policy, May 2001.
11. Breiman L, Random forests. *Machine Learning* 45:5-32 (2001).

ENDNOTES

ⁱ For purposes of exposition, we ignore uncertainty about the precise levels of B and C.

Uncertainty in these levels can be incorporated into the procedures described below via bootstrap replications of these procedures for randomly selected values of B and C from any specified bivariate distribution consistent with prior belief or evidence.

ⁱⁱ Recall that we are not assuming that every positive child was prevented from incarceration by the program; thus, B is the average gross benefit per true positive.

ⁱⁱⁱ As suggested above, uncertainty about values of B and C could also be incorporated into the model selection process via bootstrapping.

^{iv} Since the comparison presented here involves models with first-grade data versus models with first-grade and third-grade data, additional costs for collecting the third-grade data should also be incorporated though they have been ignored here. An alternative comparison could also be carried of models based only on first-grade data versus models based only on third grade data. Presumably any differentials in data collection costs between these two models would be much smaller. Similar comments apply to comparisons involving models using sixth-grade data.

^v Prorating the third-grade and sixth grade model results to the N for all children on whom we have first-grade data assumes purely random attrition between the first-grade and later years' surveys. We tested the sensitivity of our comparison to this assumption by using only the 450

individuals who were included in the estimation of the first and third grade model, re-estimating the first grade model and then computed the optimal selection risk and maximum expected net benefit level for each of the two models. The third-grade model yielded a maximum expected net benefit that was 20.3 % larger (\$553,520 vs. \$459,980) and a lower optimal selection risk (0.18 vs. 0.27). Thus the gain in expected net benefit from using the first and third grade model (relative to the first grade model) was smaller in relative terms but still substantial.

^{vi} For our values of $C = \$3,899$ and $B = \$20,000$, the cost for a false negative is \$16,101 and the cost for a false positive is \$3,899. The ratio of these values implies a criterion function (weighted Kappa) with a weight in 0.805. To accommodate the available software used in the Kiernan analysis, this was simplified to a weight of 0.8. (Note that an unweighted Kappa criterion function would use a weight of 0.5.)

^{vii} Limiting the number of groups to three is analogous to limiting the two-factor probit model to three identified parameter estimates by only using linear terms for the two risk factors. The limit of 3 groups was applied by only considering the first two splits in the partition “tree” that resulted in maximum expected net benefits.

^{viii} In this case the groups were defined as follows: Group 1 – $\text{TOCGB1} < 4$, Group 2 - $\text{TOCGB1} \geq 4$ and $\text{PERF} < 0.21$, and Group 3 - $\text{TOCGB1} \geq 4$ and $\text{PERF} \geq 21$. The corresponding percentages of true positives were 12.7, 13.8 and 31.8 respectively. As before, expected net benefit was maximized when only Group 3 was selected for the program. (Note however that in the case the ROC4 software produced four groups (i.e., a second-level split in each side of the “tree”) so that we had two alternative sets of 3 groups to choose from.)

^{ix} In the specific example that they analyze, Berk et al. (6) report substantially better predictive performance of recursive partitioning (compared to logistic regression). Their measure of

predictive performance is based on the 0.5 selection risk level rather than an optimal selection risk level as developed here.

^x Berk et al. (6) stress the importance of the over fitting problem in the context of recursive partitioning and demonstrate the use of “random forests” (11), an extension of standard bootstrapping methods, to address the problem.