RUNNING HEAD: ANTISOCIAL PREDICTORS OF VIOLENCE

Antisocial Predictors of Violence:

A Meta-Analysis

James H. Derzon

Hamilton Fish Institute,

George Washington University

October 10, 2000

Abstract = 170

Text = 5,951

Figures = 5

December 30, 2002

http://hamfish.org/pub/apv.pdf

Abstract

Youthful antisocial behavior is often viewed as a precursor to later violent and threatening behavior. Olweus reported aggressive reaction patterns in males that over time approached the stability of intelligence. While Olweus did not examine violent behavior directly, his study is often cited as evidence for the stability of violence. To examine the evidence for this assertion, this study synthesized the evidence from 82 reports of 58 prospective studies that followed individuals over some period of the life span. After correcting effect sizes for exogenous study features, the grand mean correlation of antisocial and substance misusing behaviors with later crimes against persons was estimated to be $\underline{r} = .33$, a far cry from the stability of intelligence. Because these predictors are often used to select people into intervention, this study estimated the conditional error rates associated with identification for preventive intervention. Overall, selection failed to identify 66 percent of those who displayed later violence, while, on average, 60 percent of those engaging in antisocial or substance-using behavior were not later violent.

(N = 170)

ANTISOCIAL BEHAVIOR AND THE PREDICTION OF VIOLENCE

Schools, and society in general, have become increasingly wary of youth transgressions (e.g., Bennett, 1993). To deter youth from committing offenses and to punish those youth who do offend, increasingly stricter codes of conduct have been implemented (Antonucci, 1994). Whether explicit or implied, many of these policies are predicated on the idea that once youth begin to engage in antisocial activities they are at increased risk to continue to behave inappropriately (Dwyer, Osher, & Warger, 1998; Farrington et al. 1990). Using meta-analytic techniques, this paper examines the evidence for this belief and several other aspects of the predictive relationship of antisocial and substance-using behaviors on the display of violent or threatening behavior. In addition to summarizing evidence on these relationships, this paper assesses the concurrent and predictive relationships of these predictors with violence and specifies the accuracies and errors associated with identifying persons for intervention or punishment to reduce violence based on previous displays of these antisocial behaviors.

According to Gottfredson and Hirschi (1987), aggression measured over intervals of a year or two may correlate at levels approaching the reliability of the measures. In other words, they believe that aggression is a stable trait of individuals within the limits of measurement. The most highly cited evidence for this observation comes from a study by Olweus in a 1979 assessment of the stability of aggressive reaction patterns. Using the criteria that a study had to be longitudinal in design and that the aggressive behavior or reaction pattern had to be observed or inferred by individuals other than the subjects themselves, Olweus (1979) located 14 publications of 16 studies on the stability of male aggressive reaction patterns. From these documents Olweus coded 24 stability coefficients based on samples of between 32 and 410 males with measurement intervals ranging from 0.5 to 21 years.

Based on these data, Olweus showed average raw and disattenuated correlations of .63 and .79, respectively, using one estimate per study and coefficients of .55 and .68, using the entire set of 24 estimates. Olweus also demonstrated that the size of the correlation, when corrected for attenuation, tended to decrease linearly as the interval between the two times of measurement increased. He compared these findings to the regression line for Stanford-Binet intelligence test data. A later study, assessing the stability of aggressive female reaction patterns (based on six studies contributing 21 stability coefficients), showed essentially the same results (Olweus, 1981, Figure 6). Olweus concluded there was much greater longitudinal and cross-situational stability (i.e., trait-like behavior) in aggression than would be expected given Mischel's earlier conclusion "that the concept of personality traits as broad response dispositions is . . . untenable" (1978, p. 146, cited in Olweus, 1980). The conclusion from these studies, in other words, is that aggression is a trait-like behavior that represents a stable characteristic of individuals.

Olweus' work (1978, 1979, 1981) is generally taken as evidence for the stability of aggressive and, by extension, violent behavior (e.g., Loeber & Stouthamer-Loeber, 1998). Aggressive behavior, however, was not a necessary requirement for inclusion in Olweus' study. Olweus (1979) defined aggression as

> any act or behavior that involves, might involve, and/or to some extent can be considered as aiming at, the infliction of injury or discomfort; also manifestations of inner reactions such as feelings or thought that can be considered to have such an aim are regarded as aggressive responses. (Olweus, 1973, p. 240; cited in Olweus, 1979, p. 856).

Given this definition, Olweus excluded such variables as competitiveness, dominance, repression of aggressive thoughts, results based on projective instruments, and other indirect manifestations of aggressive or other tendencies. He did, nonetheless, include measures containing such variables as irritability (Block, 1971, cited in Olweus, 1979, p. 856), verbal protest (Olweus, 1977, cited in Olweus, 1979, p. 859), and a peer inventory including items assessing who does not obey the teacher, who says "give me that," who gives dirty looks or sticks out their tongue, who makes up stories and lies to get other children into trouble, who bothers others, who starts fights over nothing, who pushes or shoves children, who is always getting into trouble, who says mean things, and who takes other children's things without asking (Eron, Huesmann, Lefkowitz, & Walder, 1972, p. 254). While these are certainly unpleasant activities, these items do not necessarily measure violent or threatening behavior.

Several other efforts have been either promoted or cited as evidence for the stability of antisocial behavior, and violence in particular (e.g., Farrington, 1991; Loeber, 1982; Mossman, 1994), but none of them has focused exclusively on systematically assessing the evidence for the stability of violent behavior nor attempted to estimate its concurrent relationship with other antisocial behaviors. Despite the paucity of evidence on the stability of these relationships, the assumption of the continuity and stability of antisocial behavior forms a fundamental justification for policies of graduated sanctions and selected or indicated intervention (see e.g., Krisberg & Howell, 1998).

Using Prediction to Improve Intervention Practice

Whether implicit in the operationalization or stated explicitly in the goals, preventive interventions attempt to either enhance protective factors or reduce or ameliorate risk factors and their relationships with outcomes. Because causal experimentation with many of the factors proposed as significant in the development of violence is unethical or impossible, scientists rely on estimates of natural covariation. When paired with theory, these estimates have explanatory value in predicting later behavior. This evidence provides a science-based justification for targeting interventions but does not provide all the evidence that is useful and appropriately considered when identifying individuals for selective or indicated intervention.

Gordon (1983) argues that prevention efforts can be "operationally classified on the basis of the population groups among which they are optimally used" (p. 107). He further suggests using the term "universal intervention" to refer to interventions that are generally beneficial and "can be advocated confidently for the general public" (p. 108). Interventions that are "advisable for population subgroups distinguished by age, sex, occupation, or other evident characteristics, but who are perfectly well" are candidates for selective intervention (p. 108). These interventions are appropriately applied to subgroups who, while not displaying violence, are at risk for later violent behavior. When an individual has a "demonstrable condition that identifies the individual as at higher-than-average risk for the future development of a problem," the individual is eligible for indicated intervention (p. 108). This framework links one's status on different kinds of predictors with the interventions that may vary in cost, intrusiveness, and other issues of concern to citizens or prevention planners.

When identifying youth for selective or indicated intervention, the primary concern lies with such issues as whether those identified are likely to benefit from preventive intervention, whether those not identified might benefit from intervention, whether participation in an effective intervention is likely to substantially impact the outcome of interest, and other issues associated with identification and exposure, or not, to intervention. These and other matters surrounding the justification of identifying individuals and groups for intervention to prevent violence can be effectively addressed by assessing the conditional relationships between selection variables and the later display, or not, of violence. Because selection requires that individuals be placed above or below some criterion value of eligibility, selection requires dichotomizing the predictor on some level of eligibility.

The 2 x 2 table. When the predictor and criterion can be dichotomized into present and absent conditions, a 2 x 2 contingency table can be used to map all the possible interrelations of conditions (See Figure 1). While all four cells of the contingency table are necessary to characterize the relationship between two variables, the 2 x 2 contingency table is useful for measuring and demonstrating the conditional probability that an antecedent condition above or below a stated threshold will result, or not, in an undesired outcome.

--------- Figure 1 about here ---------

Because these conditional relationships vary independently of the strength of correlation, knowing the overall strength of correspondence between a selection variable and an outcome tells little about the potential error terms associated with selection for preventive intervention. For example, Figure 2 shows the inferred cell values of the 2 x 2 table at different levels of selection when the correlation between a predictor and an outcome is set at $r = .34$ and the outcome rate is fixed at 15 percent. This graph is read vertically, as if slicing a loaf of bread. Each "slice" shows the different percentages of true negatives, false negatives, false positives, and true positives at each percentage of selection when the outcome and correlation are fixed. Thus, when the cutting score (the value at which the predictor is dichotomized) identifies 10 percent of the population at risk (positively identified on the predictor), half of that group (5 percent of the population) can be expected to engage in the outcome while the other half is unlikely to (true and false positives respectively). With 90 percent of the population not predicted to be at risk, 10 percent of the population are likely to fall below the cutting score and still engage in the outcome (false negatives) while 80 percent of the population are not expected to show the outcome (true negatives).

--------- Figure 2 about here ---------

Although establishing the cutting score that separates eligibility from ineligibility can engender considerable debate, it is essentially a trade-off between the positive predicted value (PPV) of a predictor and its sensitivity when outcome rates are fixed. Sensitivity indexes the proportion of those who display the outcome who were positively identified by the predictor (see Figure 1, A/(A+C)). Sensitivity thus estimates the potential influence an effective intervention could have in reducing later violent behavior by estimating the percentage of those who display the outcome who would be identified as eligible to participate in preventive intervention. Sensitivity is improved by relaxing eligibility criteria so more individuals are eligible for intervention. All other things being equal, by estimating the potential impact an effective intervention could have in reducing later violent behavior, predictors that show high sensitivity to the outcome may be an intervention planners' first choice in targeting interventions.

Of course, all things are never equal, and resources, stigma, and the opportunity costs of being labeled "at risk" are often prominent concerns when identifying individuals for intervention. As these concerns come to the fore, maximizing the proportion of those identified for intervention who are likely to display violence gains prominence. This relationship is known as the positive predicted value (see Figure 1, A/(A+B)). As eligibility criteria are tightened so that selection is more discriminating and fewer individuals are identified for intervention, a greater percentage of those so identified are likely to display the outcome. Unfortunately, the cost of improving PPV accuracy is a decrease in sensitivity so that increasingly fewer individuals who are likely to display violence are identified by the process.

When selecting individuals for intervention, such as when schools identify youth at risk for violence, the primary concern is whether the selection criteria improve prediction accuracy and reduce prediction error. Are selected youth likely to benefit from preventive intervention (estimated by PPV)? Will successful intervention have an appreciable impact on the prevalence of violence (estimated by sensitivity)? Could youth not selected for intervention have benefited from prevention services (estimated by specificity; see Figure 1, D/(C+D))? Correlations, odds ratios, Receiver Operating Characteristics, and other measures estimating the entire summed relationship between variables, provide little guidance in answering these ethical and programmatic concerns (Derzon & Lipsey, 1999a).

These concerns are particularly salient in the context of estimating the predictive relationships of antisocial behaviors to violence. Individuals engaging in these activities are sometimes subject to school sanctions such as suspension, and many of these behaviors are subject to legal prosecution and punishable under the law (see Bailey, this issue). Given the potential seriousness of being identified on either the predictor or the outcome in this set of relationships, it is perhaps appropriate to frame the concern with identification around the idea of reducing error. While the evidence will show that antisocial behavior is significantly predictive of violence, it is in agreement with Loeber and Stouthamer-Loeber that the stability of violence by no means approaches "the stability of intelligence over time ... [it] is not necessarily stable for all individuals and under all measurement conditions" (Loeber & Stouthamer-Loeber, 1998, pp. 100–101).

Method

The data for the current investigation were drawn from an ongoing meta-analysis of prospective longitudinal studies on the development of antisocial behavior and substance use. A keyword bibliographic search of computer databases identified more than 18,000 reports potentially containing evidence on the prospective influence of risk factors on antisocial behavior, violence, youth alcohol or tobacco use, adult alcohol misuse, or marijuana use at any age. Researchers reviewed each abstract and made a determination of likely eligibility. To be eligible, a document had to report cross-sectional or longitudinal quantitative findings from a study following individuals over time. This process identified nearly 5,000 reports for retrieval.

These reports were retrieved, sorted according to their study of origin, coded for eligibility and type of antisocial or substance-using behavior measured, and bibliographies were scanned for additional reports. All reports containing eligible data were coded by trained personnel on nearly 100 items related to study methods and procedures, subject and cohort characteristics, predictor and outcome measures and constructs, and other variables hypothesized to influence effect size magnitude. Ultimately, 1,055 reports from 402 independent studies contributed nearly 30,000 estimates of the relationship of some risk or protective factor with current or later antisocial or substance-using behavior.

This database was reviewed and effect sizes indexing the relationship of an earlier display of antisocial or substance-using behavior with a current or later report of crimes against persons were identified for analysis. This procedure identified 1,040 effect sizes from 82 documents based on 58 studies and 38,254 subjects.

To create the categories in this set of syntheses, this study used "bottom-up" methods to distill 12 predictor categories based on type of behavior from the diverse constructs measured in individual studies. This involved a series of iterative judgments that sorted the construct data into increasingly refined clusters. "Crimes against persons" was operationalized as those activities that involve any kind of interpersonal violence, battery, extortion, or other form of threatening behavior. The different predictor categories can be distinguished by refinement. "Recidivism" is a particular form of criminal activity, thus measures estimating repeat offending were coded as recidivism rather than criminal activity. It should be emphasized that this is not an empirical clustering. Different judges would perhaps group the measures differently and different groupings might reasonably be defined for different purposes.

To adjust for bias introduced from study methods or procedures, each effect size was statistically adjusted using mixed effects weighted regression analysis to approximate the finding that would have been obtained from a study of known characteristics. Details of the procedure used to adjust the observed data can be found in Derzon and Lipsey (1999b). When modeling the current data, three variables were found to systematically influence effect size magnitude, and the data were statistically adjusted to control for these characteristics. Relationships tended to be stronger when studies were performed in the United States ($\hat{a} = .42$), used general populations ($\hat{a} = .33$), and used samples that were younger at the time the outcome was measured ($\hat{a} = .15$). Together, these three variables explained 24 percent of the variability among effect sizes. Since effect sizes would be more comparable after controlling for these study features, effect sizes were adjusted to be representative of data that would have been obtained from a general population in the United States who were 18 years old at the time the outcome was measured.

When the same study sample provided multiple estimates of similar relationships, these partially redundant estimates were averaged within each study sample prior to synthesizing across study samples to eliminate statistical dependencies. These procedures reduced the number of effect sizes for synthesis to 161 independent aggregated effect sizes. Before being analyzed across study samples, all effect sizes were $\underline{Z}$-transformed to normalize the distribution of effect sizes and to make the sampling error variance independent of the population correlation (Becker & Hedges, 1989; Rosenthal, 1994). In addition, standard meta-analytic weighting procedures were used in all analyses to give greater weight to effect size estimates based on larger samples (Hedges & Olkin, 1985). However, to keep the largest samples from overwhelming the contribution of smaller samples, those larger than N = 700 were recoded (Winsorized) to 700, which, therefore, became the maximum sample size value used in the weighting procedure.

Results

Correlational Findings

The first set of analyses breaks out the data by whether the predictor and outcome were prospectively or concurrently measured. While the eligibility criteria specified that studies had to be prospective to be eligible, both longitudinal and cross-sectional estimates were coded when they were provided. Thus, while all the data in this study come from prospective studies, nearly a third of the effect sizes coded were cross-sectional. Having both types of data allows estimating whether these antisocial behaviors represent a developmental trajectory culminating in crimes against persons or if the individuals responding to these items were as likely to have been committing crimes against persons at the time the antisocial predictors were measured.

As can be seen in Figure 3, regardless of the timing of measures, all antisocial constructs were significantly and positively related to crimes against persons (none of the error bars crosses the zero criterion). Statistically adjusting the data for study features increased the grand mean for all values from an observed value of r = .205 to r = .339. No cross-sectional estimates of the relationship of severity of crime with crimes against persons were identified. Indicating no significant differences between the two estimates, the error bars for cross-sectional and longitudinal estimates overlap on 6 of the 11 constructs for which both could be estimated.

Among the five constructs whose estimates differed, three differences favored the longitudinal estimate (i.e., criminal activity, fighting, and other antisocial behavior).

--------- Figure 3 about here ---------

Because nearly half the estimates differed according to time of measurement and because the relative paucity of cross-sectional estimates would not constrain analysis, the rest of this investigation is limited to prospective estimates in which the antisocial predictor is measured at one point in time and the crimes against persons outcome is measured at some point later in time. Many have noted that the early display of antisocial behavior portends particularly serious consequences for youth (Farrington et al. 1990). To examine the effect of age, the data in each predictor category were broken into five age groups ranging from six years old and under to age 18 and older at the time the antisocial predictor data was collected (Figure 4). It is worth emphasizing that these data do not necessarily describe behaviors that occurred within these age groupings but reflect the age at which the predictor was measured.

--------- Figure 4 about here ---------

In addition to the symbols used to indicate age, Figure 4 uses height within category to aid interpretation. Within each category the samples are arrayed by age, with the youngest measured estimate closest to each category's bottom line and estimates from increasingly older samples nearer the upper line. Using these aids it is relatively easy to observe that the relationship of antisocial predictors with crimes against persons generally increases in strength as youth age. These changes are rarely so radical that differences are significant in adjoining age groups, but for the categories of criminal activity and other antisocial behavior in which there appear to be clear age trends, estimates for the oldest group are clearly larger than those for the smaller.

While no category shows clear negative trends in which relationships are stronger for each successively younger age category, the six-years-old and under estimates are particularly strong for the categories of recidivism and crimes against persons. For the categories of fighting, aggressive-disruptive behavior, other antisocial behavior, and criminal activity the six-year-old-and-under estimates are the weakest relationship obtained. Since these terms imply legal definitions it is worth emphasizing here that these behaviors were grouped by type. Thus, "recidivism" includes being in multiple fights (a more refined category than fighting), hitting teachers three times or more, as well as being arrested more than once, whereas the category "crimes against persons" includes all violent and threatening behaviors regardless of whether or not subjects faced adjudication.

Conditional Findings

Because fewer than a third of the studies reported conditional data on these prospective relationships, it was necessary to infer the 2 x 2 cell values for nearly 7 out of every 10 effect sizes. To calculate these missing values, the study synthesized the selection and outcome rates from the general population samples to estimate the mean selection and overall outcome rate of violent and threatening behavior in these samples (Derzon, 1996; Derzon & Lipsey, 1999b; Lipsey & Derzon, 1998). These values (32 percent for the selection rate and 15 percent for the outcome rate) and each case's adjusted correlation were then used to solve a simultaneous equation that estimated the missing cell values. These inferred and observed cell values were then weighted by the Winsorized sample size and synthesized using the meta-analytic techniques described above. By following these procedures, all longitudinal estimates were able to contribute to the conditional findings, but the procedure standardizes the findings more than would be expected from field-generated data. The synthesized values obtained for each cell of the 2 x 2 contingency table are displayed graphically in Figure 5.

--------- Figure 5 about here ---------

In Figure 5 the standardized cell values for each construct category are displayed as a horizontal bar that equals 100 percent of the synthesized sample. Thus, the first predictor in Figure 5, criminal activity, can be read as combining evidence from 13 estimates in which a mean of 17 percent of the sample engaged in prior criminal activity (selection rate). On average, 9 percent of the combined sample was predicted to engage in crimes against persons and did (true positives), while 8 percent of the sample so identified did not (false positives). Of those positively identified as having a history of criminal activity, 53 percent engaged in later crimes against persons (PPV; true positives/selection rate) leaving 47 percent of those so identified who did not commit a later crime against a person (positive prediction error = 100 – PPV).

Focusing still on the horizontal bar representing criminal activity, 83 percent of the sample had no history of prior criminal activity (100 – selection). Of these, 24 percent nonetheless committed a crime against a person (negative error rate [NER] = 100 – selectivity). An average of 29 percent of those contributing evidence to this estimate displayed the outcome (true positives and false negatives); of these, 31 percent engaged in prior criminal activity and had been positively identified by the predictor (sensitivity; true positives/outcome rate).

While each value, and its relationship to other values, carries information that is potentially useful to such decision makers as school principals and school psychologists, it is usually a subset of that information that is of core concern for any particular decision. When the task is selecting individuals for intervention, and if the desire is to maximize the accuracy of selection so that a majority of those selected are likely to display the behavior, then the PPV of a criterion variable is of central concern. From the data presented in Figure 5, the strongest predictors in this regard are shown to be severity of crime, recidivism, and prior criminal activity. More than half of those who had engaged in these behaviors later committed a violent or threatening act. This can be contrasted with the general category of other antisocial behavior in which fewer than 20 percent of those so identified later engaged in violent or threatening behavior. Overall, slightly more than 60 percent of those who engaged in a prior antisocial or substance-using behavior did not go on to engage in a reported crime against a person.

Given the seriousness of violent and threatening behavior, it may be desirable to select as many persons who are likely to engage in that behavior for intervention services as possible. In this instance, selecting individuals based on their status on predictors that are sensitive to the outcome may be most prudent. On average, 57 percent of those who committed a crime against a person were identified by the severity of an earlier crime. Similarly, nearly half of those who engaged in later violent or threatening behavior had earlier engaged in serious or index crime. By contrast, only 23 percent of those who committed violent or threatening behavior would have been identified by earlier marijuana use (see also Derzon & Lipsey, 1999c). Across outcomes, 66 percent of those who committed a crime against a person were not identified by an earlier antisocial behavior.

Finally, estimating the negative error rate (NER) can provide critical guidance to decision makers trying to balance the distribution of resources across selective, indicative, and universal interventions. While these indicators show generally strong PPV and sensitivity with the outcome, on average nearly 25 percent of those not identified by prior antisocial or substance-using behavior committed a later crime against a person. Because these individuals had no prior behavior identifying them as likely to be violent, they would not have been eligible for, nor received, selective or indicative intervention services based on these criteria. Among these predictors, severity of crime showed the highest negative error rate (37 percent), while tobacco use showed the lowest negative error rate (16 percent). The percentage of samples that was not identified as at risk, but that, nonetheless, committed a later crime against a person ranged from between 11 percent for the category of other antisocial behavior and 21 percent for the category of severity of crime.

Discussion

In contrast to several earlier reviews of the stability of antisocial behavior, the current analysis found considerably more modest relationships between the earlier antisocial behavior and the likelihood of later violent or threatening behavior. After summarizing the findings of nearly 60 independent prospective studies documenting the natural development of violence, a grand mean correlation of .205 was observed, a far cry from the .63 observed by Olweus for aggressive tendency disorder and cited by others as the stability of violence. After statistically adjusting effect sizes to estimate the relationships that would have occurred if all studies had been conducted in the United States using general populations with the violence outcome measured when subjects were 18 years of age, the overall grand mean correlation between all forms of antisocial and prior substance-using behavior with violent and threatening behavior increased to .339. Interestingly, in these data, cross-sectional correlations were not reliably stronger, or weaker, than prospective estimates.

Despite the apparent strength of these relationships, certain errors associated with selecting persons into preventive intervention on the basis of their status on these predictors can be anticipated because of their probablilistic nature. Depending on the construct category identified for selection, positive prediction errors (100 – PPV) ranged from a high of 81 percent for the category of other antisocial behavior to a low of 35 percent for severity of crime. Similarly, these predictors failed to identify between 77 and 43 percent of those who later engaged in crimes against persons (100 – sensitivity; marijuana use and severity of crime, respectively). Overall, the false positive selection error rate for antisocial predictors of crimes against persons was 60 percent. On average, these predictors failed to identify 66 percent of those engaging in one or more violent or threatening behaviors.

While the relationships between antisocial and substance-using predictors and later crimes against persons are among the strongest predictive relationships observed using meta-analytic techniques (e.g., Derzon, 1996; Lipsey & Derzon, 1998), they by no means support the conclusion that violent behavior is a fundamentally immutable, trait-like activity of individuals. Although the sample-based data available to this meta-analysis cannot support the sophisticated model-fitting that might describe the particular intersection of person, setting, and circumstance that is the likely cradle from which violence springs, the data summarized here provide support for both selective and indicative intervention, albeit with modest ambitions for each. Selecting on these factors will not identify all those who are likely to later engage in violence for intervention, nor will all those selected for intervention based on these predictors engage in violence. However, given the sensitivity of these predictors, successful interventions directed at individuals displaying these antecedent behaviors have the potential to reduce later violence by as much as a third, given that the predictors correctly identify 33 percent of those engaging in violent or threatening behavior.

Because these data suggest that most individuals who engage in antisocial or substance-using behaviors are unlikely to engage in later violence, interventions directed at these individuals should be tempered. While severe or punishing interventions may be desired for extracting retribution, the findings summarized here suggest that many of those who would be recipients of these interventions are not likely to commit later violent offenses, regardless of their exposure to intervention. With few exceptions, notably severity of crime, recidivism, and criminal activity, most of those who engage in the activities summarized here do not display later violence. These findings should give pause as schools contemplate zero tolerance interventions that are highly stigmatizing, limit students' current or future opportunity, or are otherwise harmful to recipients.

To the extent that these results are generalizable, this evidence of the improbability of antisocial individuals engaging in later violence has implications for how researchers estimate the effectiveness of programs designed to reduce or prevent violence in the absence of a comparison group. When effectiveness studies use comparison group designs, the comparison group acts as a referent for the group exposed to the intervention. Differences between the two observed values at posttest (preferably after controlling for pretest differences) is the change attributable to intervention. When studies do not use a comparison group, but estimate the program-attributable change using only observations of a single group before and after the intervention, then that observed change is tested against assumptions of what the outcome rate would have been for the sample in the absence of intervention. If the observed amount of violence is less than the assumed value, the intervention is considered a success.

If violence is assumed to be a relatively immutable trait of individuals, then selective or indicative prevention programs which document 30, 40, or even 50 percent success rates sound pretty good. The findings summarized here, however, suggest that probabilistically speaking, fully 60 percent of those selected for intervention would not be expected to engage in later violence. By this measure, programs documenting 30 to 50 percent success rates may be having iatrogenic effects, producing, in effect, more violence than they prevent.

In addition to providing support for the potential utility of selective and indicative preventive intervention, these findings point to the need to provide support for universal interventions directed at general populations. On average, nearly a quarter of those not showing prior antisocial or substance misusing behavior engaged in later violent or threatening behavior. These individuals could perhaps have benefited from intervention but would not have been identified to receive indicated or selective intervention services. Because none of these construct categories successfully identified all who would benefit from prevention intervention, it is always desirable to supplement selective or indicative interventions with some form of universal intervention.

While this assessment of the evidence for the stability and continuity of violence benefits from the multiple strengths of meta-analysis in its comprehensiveness, its statistical reliability, and its sophisticated management of sampling and methods bias, it also suffers from the limitations of all archival research: it must work from and within the evidence available from an established record. This synthesis found relatively few cross-sectional estimates of the interrelations between various antisocial and substance misusing behaviors and later crimes against persons. Fewer than half the mean cross-sectional estimates were based on more than three independent samples.

Similarly, fewer than a third of the estimates provided information on the conditional relationship between the predictors and later crimes against persons. Including in the conditional analysis all evidence summarized in the correlational analysis meant making some assumptions of the data. First, the cutting scores used by the subset of researchers reporting conditional data were accepted as meaningful. Second, the synthesized outcome rate (15 percent) and synthesized cutting score (32 percent) were applied uniformly across construct categories. While it is unlikely that the outcome rate varies across construct categories, the assumption that cutting scores do not vary in meaningful ways across construct categories seems less defensible. More published data on these conditional relationships would have made it possible to estimate the mean cutting score for each predictor category. Having this information, however, would not necessarily ameliorate the concern that the cutting score synthesized for each predictor is the appropriate score for each purpose. As shown in Figure 2, different cutting scores can radically change contingent relationships when other features of the relationship are held constant.

Finally, this synthesis only assesses the bivariate relationships between individual predictors and later violent and threatening behavior. The question can reasonably be raised whether combining items would yield even greater predictive ability. The predictive strength of combinations of risk variables depends on the intercorrelations among the variables. If they are highly intercorrelated, they are functionally redundant for predictive purposes, and their predictive ability combined is no greater than that of the best of the variables alone. However, if two or more risk variables are individually predictive but not highly intercorrelated, their predictive strength will be greater in combination than for any one alone. It should be noted, however, that the multivariate techniques that improve predicative strength may not decrease all forms of error when applied to selecting persons for intervention.

Combining variables to improve selection involves specifying the multiple variables on which an individual is above some cutting score. In other words, people are at risk when they are above some set of criterion values for multiple variables. Selection on multiple variables is generally accomplished through one of various techniques known as multiple gating (Dishion & Patterson, 1993; Feil, Severson, & Walker, 1998; Sprague et al. this issue) or CART modeling (e.g., Lemsky, Smith, Malec, & Ivnik, 1995) or done less formally using profiles developed through multivariate modeling. Combining multiple variables that are not intercorrelated increases positive predictive value but has the effect of decreasing sensitivity. In other words, more of those identified through these multiple predictors will be likely to engage in the outcome, however those positively identified by the combined predictors will represent an increasingly smaller percentage of those who eventually engage in the outcome. When the outcome is violent and threatening behavior, concern with the shrinking sensitivity of multiple indicators is warranted.

In sum, the evidence summarized here fails to support the contention that early antisocial behavior is deterministic of violence; most of those who engage in early antisocial or substance misusing behaviors do not engage in later violence. The display of early antisocial behavior, nonetheless, is modestly to strongly predictive of violence (Cohen, 1988) and has utility for selecting individuals for interventions that may be more intrusive than universal interventions. Conditional analysis of these predictors, however, suggests that while increased sanctions are sometimes warranted, the severity of those sanctions should be tempered by the knowledge that a majority of those included in the intervention are unlikely to display later violence. Finally, it is noted that nearly one-fourth of those not identified as likely to be violent because of prior antisocial behavior did, in fact, engage in later crimes against persons and that typically a majority of those who committed a violent or threatening act had no prior history of antisocial or substance-using behavior. Thus, prevention strategies should include all three forms of intervention for maximum effectiveness. Combined, this evidence argues that interventions should be positive, not punitive; the inherent errors of selection preclude doing harm.

References

Antonucci, R. V. (1994). Advisory opinion on student discipline. Boston, MA: Massachusetts Department of Education. Available online: http://www.doe.mass.edu/lawsregs/advisory/discipline/AOSD1.html (July 28, 2000).

Bailey, K. (this issue). Legal implications of profiling students for violence. Psychology in the Schools.

Becker, B. J., & Hedges, L. V. (1989). Synthesizing research on organizational participation. Research on the Sociology of Organizations, 7, 203–231.

Bennett, W. J. (Ed.). (1993). The book of virtues: A theory of great moral stories. New York: Simon and Schuster.

Block, J. (1971). Lives through time. Berkeley, CA: Bancroft Books.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum

Derzon, J. H. (1996). A meta-analysis of the efficacy of various antecedent behaviors, characteristics, and experiences for predicting later violent behavior (Doctoral dissertation, Claremont Graduate School, 1996). Dissertation Abstracts International, 57, 748.

Derzon, J. H., & Lipsey, M. W. (1999a). What good predictors of marijuana use are good for: A synthesis of research. School Psychology International, 20, 69–85.

Derzon, J. H., & Lipsey, M.. W. (1999b). Predicting tobacco use to age 18: A synthesis of longitudinal research. Addiction, 94, 995–1006.

Derzon, J. H., & Lipsey, M. W. (1999c). A synthesis of the relationship of marijuana use with delinquent and problem behaviors. School Psychology International, 20, 57–68.

Dishion, T. J, & Patterson, G. R. (1993). Antisocial behavior: Using a multiple gating strategy. In M. I. Singer (Ed.) Handbook for screening adolescents at psychosocial risk (pp. 375-399). New York: Lexington Books.

Dwyer, K, Osher, D., & Warger, C. (1998). Early warning, timely response: A guide to safe schools. Washington, DC: U.S. Department of Education.

Eron, L. D., Huesmann, L. R., Lefkowitz, M. M., & Walder, L. O. (1972). Does television violence cause aggression? American Psychologist, 27, 253–263.

Farrington, D. P. (1991). Childhood aggression and adult violence: Early precursors and later outcomes. In D. J. Pepler & K. H. Rubin (Eds.), The development and treatment of childhood aggression (pp. 5-29). Hillsdale, NJ: Lawrence Erlbaum.

Farrington, D. P., Loeber, R., Elliott, D. S., Hawkins, J. D., Kandel, D. B., Klein, M. W., McCord, J., Rowe, D. C., & Tremblay, R. E. (1990). Advancing knowledge about the onset of delinquency and crime. In B. B. Lahey & Kazdin, A. E. (Eds.), Advances in clinical child psychology, Vol. 13, (pp. 283-342). New York: Plenum Press.

Feil, E. G., Severson, H. H., Walker, H. M. (1998). Screening for emotional and behavioral delays: The Early Screening Project. Journal of Early Intervention, 21, 252–266.

Gordon, R. (1983). An operational classification of disease prevention. Public Health Reports, 98, 107–109.

Gottfredson, M., & Hirschi, T. (1987). The methodological adequacy of longitudinal research on crime. Criminology, 25, 581–614.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press.

Krisberg, B., & Howell, J. C. (1998). The impact of the juvenile justice system and prospects for graduated sanctions in a comprehensive strategy. In R. Loeber, & D. P. Farrington (Eds.), Serious & violent juvenile offenders: Risk factors and successful interventions (pp. 346–366). Thousand Oaks, CA: Sage.

Lemsky, C. M., Smith, G., Malec, J. F., & Ivnik, R. J. (1995). Identifying risk for functional impairment using cognitive measures: An application of CART Modeling. Neuropsychology, 10, 368–375.

Lipsey, M. W., & Derzon, J. H. (1998). Predictors of serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), Serious and violent offenders: Risk factors and successful interventions (pp. 86–105). Thousand Oaks: Sage.

Loeber, R. (1982). The stability of antisocial and delinquent child behavior: A review. Child Development, 53, 1431–1446.

Loeber, R., & Stouthamer-Loeber, M. (1998). Juvenile aggression at home and at school. In D. S. Elliott, B. A. Hamburg, & K. R. Williams (Eds.), Violence in American schools (pp. 95–126). New York: Cambridge University Press.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. Journal of Consulting and Clinical Psychology, 62, 783–792.

Olweus, D. (1973). Personality and aggression. In J. K. Cole & D. D. Jensen (Eds.), Nebraska Symposium on Motivation, Vol. 20. Lincoln, NE: University of Nebraska Press.

Olweus, D. (1977). Aggression and peer acceptance in adolescent boys: Two short-term longitudinal studies of ratings. Child Development, 48, 1301–1313.

Olweus, D. (1978). Aggression in the schools: Bullies and whipping boys. Washington, DC: Hemisphere.

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. Psychological Bulletin, 86, 852–875.

Olweus, D. (1980). The consistency issue in personality psychology revisited -- With special reference to aggression. *British Journal of Social and Clinical Psychology, 19,* 377-390.

Olweus, D. (1981). Continuity in aggressive and withdrawn, inhibited behavior patterns. Psychiatry and Social Science, 1, 141–159.

Rosenthal, R. (1994). Statistically describing and combining studies. In H. Cooper & L. V. Hedges (Eds.), The handbook of research synthesis (pp. 231–244). New York: Russell Sage.

Sprague, J., Walker, H., Wagner, L., Stieber, S., Simonsen, B. & Nishioka, V. (this issue). Predicting violence and delinquency using school and community data sources. Psychology in the schools.

Author Notes

James H. Derzon is Associate Director for Research at the Hamilton Fish Institute at The George Washington University.

The author wishes to thank and acknowledge Dr. Ali Habibi. This work would not have been possible without his unflagging attention to detail and tireless assistance in coding the primary studies. This work was supported by the Center for Evaluation Research and Methodology at the Vanderbilt Institute for Public Policy Studies, the Office of Juvenile Justice and Delinquency Prevention (97-MU-FX-K012), and a grant from the National Institutes of Mental Health (MH51685). The points of view or opinions in this document are solely those of the author.

Correspondence concerning this article should be sent to James H. Derzon, Hamilton Fish Institute, 2121 K Street NW, #200, Washington, DC 20037-1830. Electronic mail may be sent via the Internet to Derzon@GWU.Edu.

Figure Captions:

Figure 1: The 2 X 2 contingency table

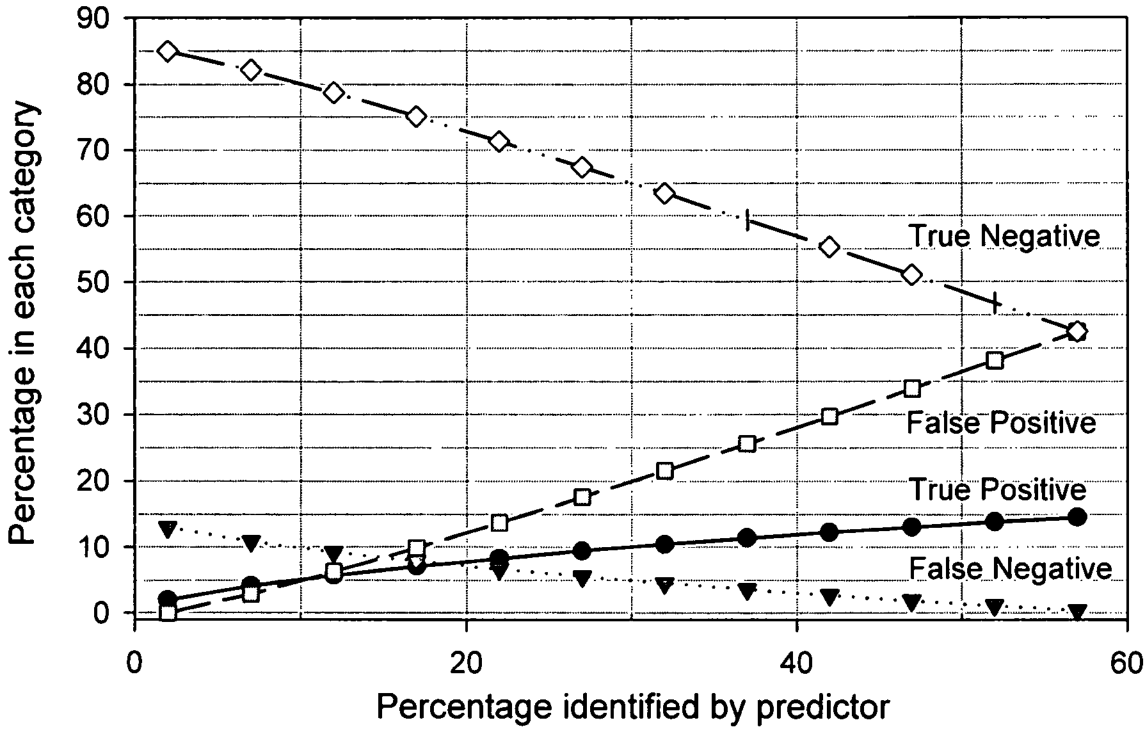
Figure 2: Inferred cell values when outcome rate is set at 15 percent and the underlying correlation is .34

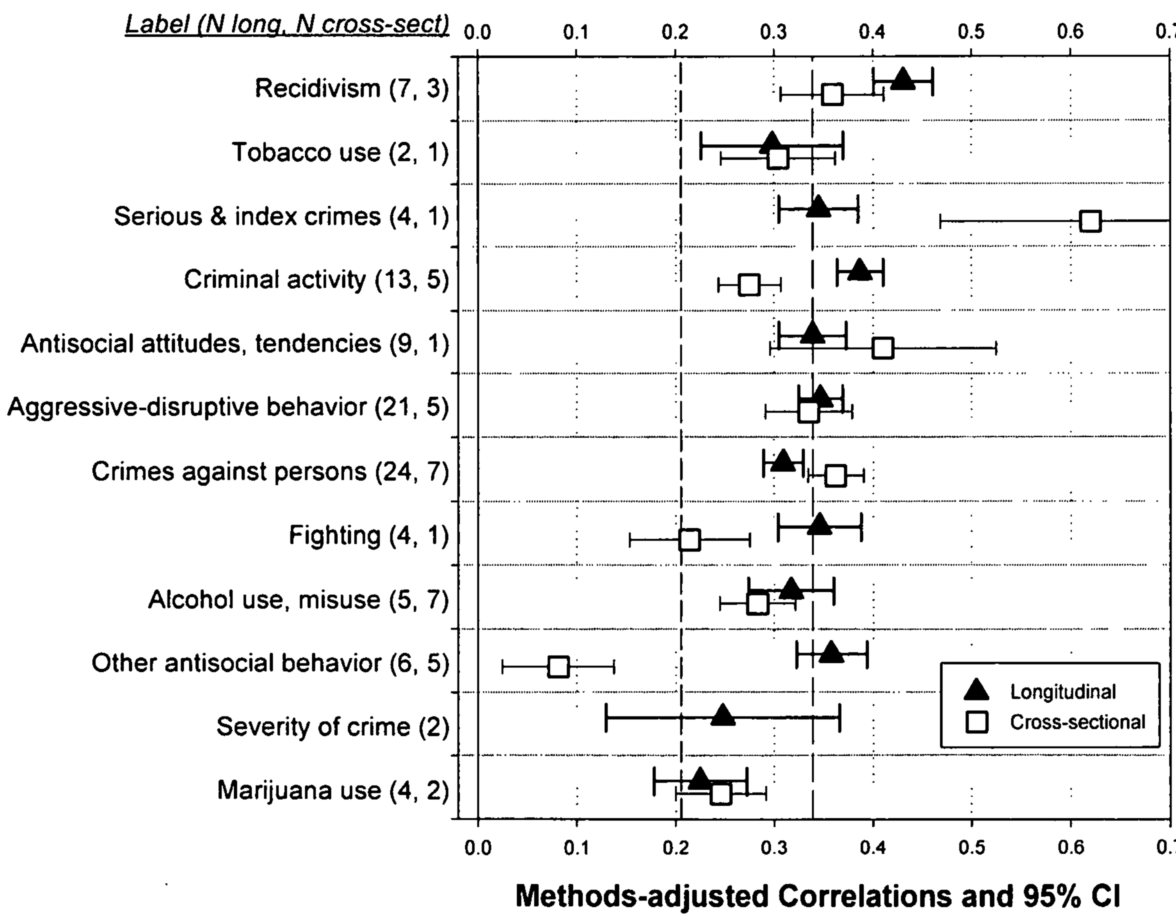
Figure 3: Antisocial predictors of crimes against persons

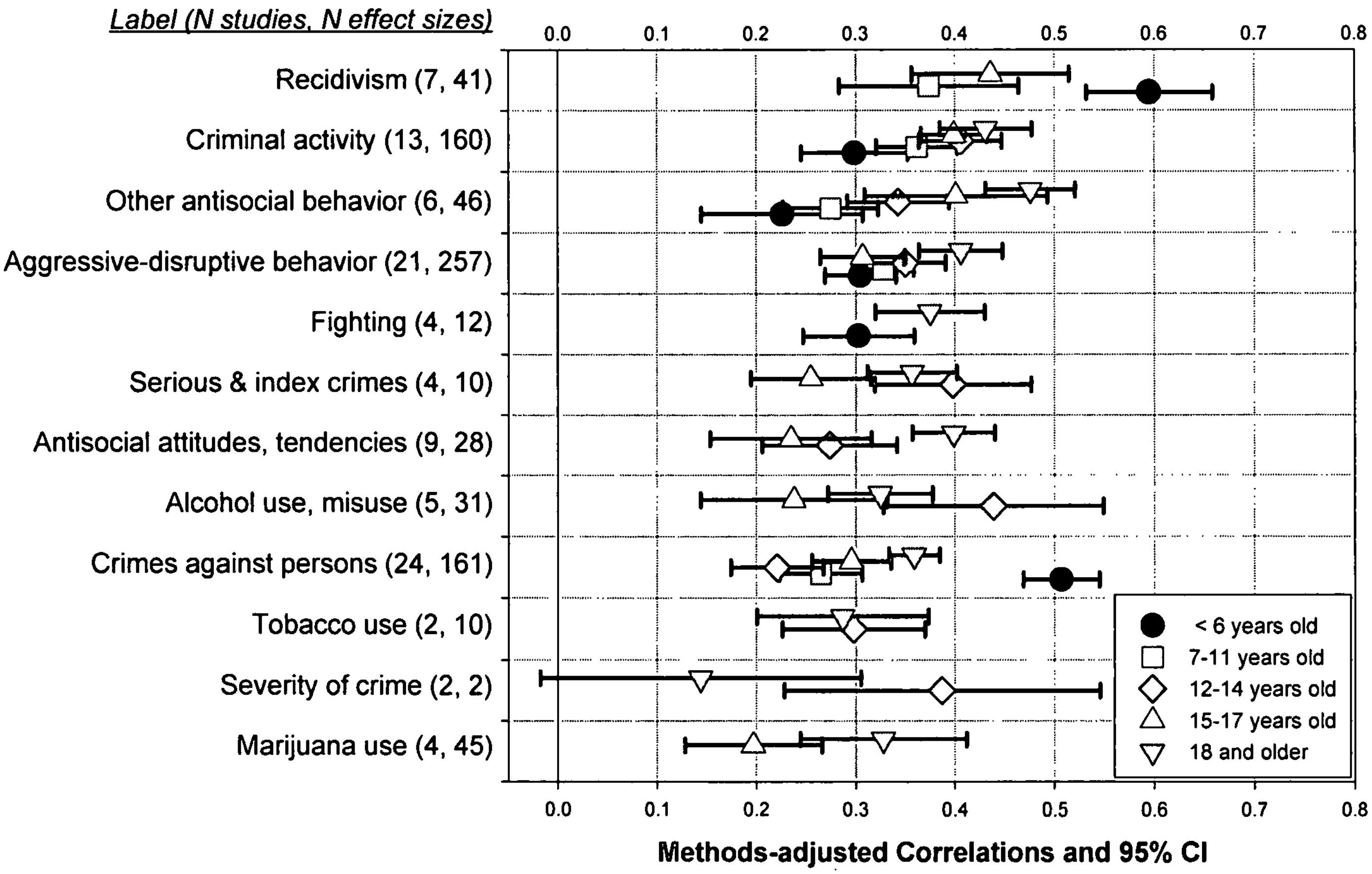
Figure 4: Antisocial predictors of crimes against persons by age of measurement

Figure 5: Observed and inferred conditional relations of antisocial predictors of crimes against persons where NER equals the percentage of those negatively identified by the predictor who were violent, PPV equals the percentage of those positively identified by the predictor who were violent, and Sens equals the percentage of those who were violent who were positively identified by the predictor.

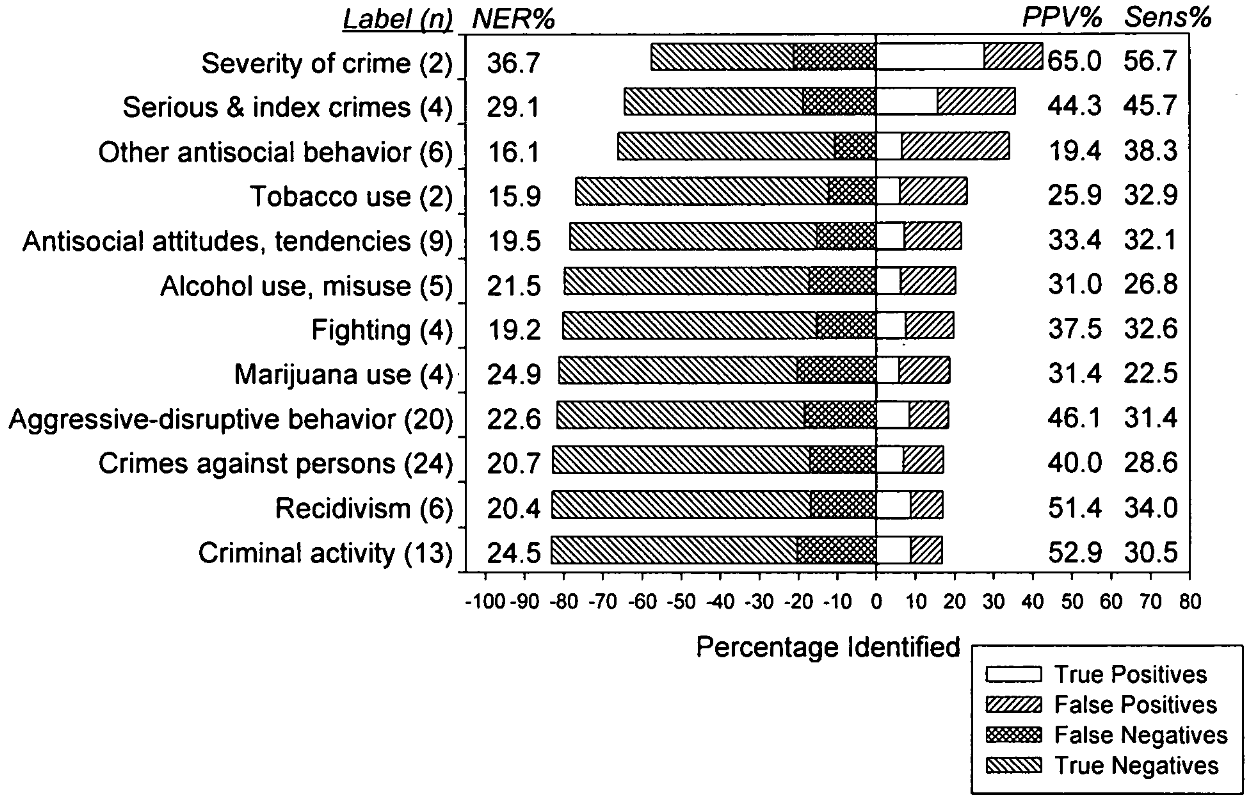
		Time 2		
		Criterion Present	Criterion Absent	
Time 1	Predictor Present	A (True Positives)	B (False Positives)	A + B (At Risk)
	Predictor Absent	C (False Negatives)	D (True Negatives)	C + D (Not At Risk)
		A + C (Outcome)	B + D (No Outcome)	N (Total)


90
80
70
60
50
40
30
20
10
0
Percentage in each category
True Negative
False Positive
True Positive
False Negative
0
20
40
60
Percentage identified by predictor


Label (N long, N cross-sect)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Recidivism (7, 3)
Tobacco use (2, 1)
Serious & index crimes (4, 1)
Criminal activity (13, 5)
Antisocial attitudes, tendencies (9, 1)
Aggressive-disruptive behavior (21, 5)
Crimes against persons (24, 7)
Fighting (4, 1)
Alcohol use, misuse (5, 7)
Other antisocial behavior (6, 5)
Severity of crime (2)
Marijuana use (4, 2)
Longitudinal
Cross-sectional
Methods-adjusted Correlations and 95% CI



Label (N studies, N effect sizes)
Recidivism (7, 41)
Criminal activity (13, 160)
Other antisocial behavior (6, 46)
Aggressive-disruptive behavior (21, 257)
Fighting (4, 12)
Serious & index crimes (4, 10)
Antisocial attitudes, tendencies (9, 28)
Alcohol use, misuse (5, 31)
Crimes against persons (24, 161)
Tobacco use (2, 10)
Severity of crime (2, 2)
Marijuana use (4, 45)
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
< 6 years old
7-11 years old
12-14 years old
15-17 years old
18 and older
Methods-adjusted Correlations and 95% CI


Label (n)
NER%
PPV%
Sens%
Severity of crime (2) 36.7 65.0 56.7
Serious & index crimes (4) 29.1 44.3 45.7
Other antisocial behavior (6) 16.1 19.4 38.3
Tobacco use (2) 15.9 25.9 32.9
Antisocial attitudes, tendencies (9) 19.5 33.4 32.1
Alcohol use, misuse (5) 21.5 31.0 26.8
Fighting (4) 19.2 37.5 32.6
Marijuana use (4) 24.9 31.4 22.5
Aggressive-disruptive behavior (20) 22.6 46.1 31.4
Crimes against persons (24) 20.7 40.0 28.6
Recidivism (6) 20.4 51.4 34.0
Criminal activity (13) 24.5 52.9 30.5
-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80
Percentage Identified
True Positives
False Positives
False Negatives
True Negatives