Offender Programs Report

Social and Behavioral Rehabilitation in Prisons, Jails and the Community

Vol 3. No. 5 ISSN 1093-7439 Pages 65–80 January/February 2000

Probation and Parole Officer Style and Substance Abusing Offenders

by Sam Torres, Ph.D., and Robert M. Latta

To a significant degree, substance abuse and crime are intricately related. Developing an effective strategy to address the substance abuse problems of probationers and parolees is therefore a critical challenge for community-based corrections. In recent years, the supervision of offenders in the community, either on probation, parole, or supervised release, has become tantamount to the care and control of the drug and alcohol-abusing offender. While the prison population has soared in recent years, the number of offenders on probation and parole supervision has grown even more rapidly.

Drugs and Crime

In a major study, the National Center on Addiction and Substance Abuse (CASA) at Columbia University (Belenko, 1998) found that of the approximately 1.7 million men and women in jail or prison in 1998, fully 1.4 million had a history of substance abuse. Research reflects that drug and alcohol abuse is highly correlated with criminal behavior. (See, e.g., Deschenes, Turner, and Clear, 1992.) In a New York study, Wish (1987) found that, over a two- month period in 1986, 92% of all suspects, arrested, booked, and charged with robbery, and 81% charged with burglary, tested positive for cocaine use. Atmore and Bauchiero (1987) also found that 87% of inmates participating in a pre-release program in Springfield, Massachusetts had significant substance abuse problems. In a Canadian study, Zamble and Quinsey (1997) found that for the majority of parole violators, substance abuse is so entangled with other maladaptive behavior that they may be inseparable, and the use of intoxicants is certainly an important part of the antecedents of re-offending.

An Effective Supervision Strategy by a Federal Probation Office in Los Angeles

Punishments for drug possession and sales have become considerably more severe, but there has also been a renewed interest in treatment for drug addiction, and some of the prison- based programs, especially those based on the therapeutic community model,

See STYLE, next page

Science of Risk Prediction of Criminal Behavior

by Rodney V. McCormick, Ph.D., and Henry R. Cellini, Ph.D.

The issue of risk prediction of criminal behavior is of great interest to correctional administrators, policy makers, and clinicians. The determination of risk of potential recidivism can allow administrators to develop treatment programs based upon risk factors while simultaneously providing policy makers with the information necessary to support programs and sentencing alternatives in order to reduce harm in their communities.

One difficulty with risk prediction is that it involves extremely difficult mathematical and statistical information. In this article, we try to simplify the concepts of risk prediction for the non-statistician. We hope that our efforts are successful, but individuals will probably wish to pursue information found in the reference list.

In order to assess the risk of an individual for criminal recidivism, and more specifically violence and sexual violence, it is common to refer to previous studies to categorize the person and use the previous results to estimate the present risk. In order to apply previous research to an individual situation one must be convinced that the previous research is in fact "generalizable" to the present situation. One must examine the research based on certain specific factors of the research design. Understanding these factors and how they relate to the individuals or group of individuals whose future

See RISK PREDICTION, page 76

ALSO IN THIS ISSUE

RISK PREDICTION, from page 65

behavior you wish to predict is essential to the appropriate use of risk prediction research. These factors are:

- The sampling design;
- Reliability and validity of the observational instruments; and
- The appropriateness of the statistical inference.

Understanding these factors of the research design and how they relate to the individuals or group of individuals whose future behavior you wish to predict is essential to the appropriate use of risk prediction research.

The Sampling Design

The sampling design refers to the details about how the observations were made and to which groups, if any, the results can be generalized. For a more complete description of sample and other design issues, it is recommended that one review the Maryland Scale of Scientific Methods, as it provides a brief summary of such classic research texts as Campbell and Stanley and discusses their use in criminal justice research. Several broad approaches to research design and sampling include:

Ecological Studies: Crime rates and exposures to risks are measured in each of a series of populations and their relationship is examined. Often the information about crime and exposure to crime is abstracted from published statistics such as the FBI's Uniform Crime Reports or already assembled research databases, and therefore does not require expensive or time consuming data collection.

Longitudinal Studies: Subjects are followed over time with continuous or repeated monitoring of risk factors or outcomes, or both. Such investigations vary enormously in their size and complexity. At one extreme a large population may be studied over decades. Outcomes such as re-incarceration and incidence of crime can be related to employment status, housing, and other variables measured at successive censuses.

Case-Control and Cross Sectional Studies: One of the drawbacks of using a longitudinal approach to investigate the causes of crime with low base rates is that large and lengthy studies may be required to give adequate statistical power in order to see very small differences. An alternative that avoids this difficulty is the case-control or case-referent design. In a case-control study persons who have committed a crime are identified and their past exposure to suspected risk factors is compared with that of controls or referents who did not commit a crime.

Experimental Studies: The survey designs described above are all observational, meaning that investigators study people as they find them. Thus, subjects exposed to a risk factor often differ from those who are unexposed in other ways, which independently influence their risk of criminal behavior. If such confounding influences are identified in advance then allowing for them in the design and analysis of the study may be possible. There is still, however, a chance of unrecognized confounding factors. Experimental studies are less susceptible to confounding because the investigator determines who is treated and who is not. In particular, if treatment is allocated randomly and the number of groups or individuals randomized is large then even unrecognized confounding effects become statistically unlikely. There are, of course, ethical constraints on experimental research in humans, and it is not acceptable to expose subjects deliberately to potentially serious hazards. This limits the application of experimental methods in the investigation of risk factors, although it may be possible to evaluate preventive strategies experimentally. For example, to test the hypothesis experimentally that child abuse causes criminality, rather than assign two groups of kids to be raised with and without abuse, it would suffice to randomly select high risk families and assign them to abuse prevention treatment or to a control group.

Observational Instrument Reliability and Validity

Whether it is an electronic meter attached to a person or a questionnaire, there must be a method to determine whether the instrument is accurate and that it indeed measures what it says it measures. Various measures of reliability that are important include:

Internal Consistency: Accuracy is typically gained by measuring something several times and taking an average. Measuring someone's height over time is an obvious example of taking a measure several times and then averaging the values to obtain an accurate measure. In a questionnaire, it is common to ask the same question several different ways to create a "sub- scale." Reliability is measured using either split half correlation (a procedure that partitions the variables in a scale into two subsets, computes reliability coefficients for each subset of variables, and determines correlation according to the extent to which the two halves of the scale measure the same thing) or Cronbach's Alpha (the reliability coefficient based on the internal consistency of items within a test, ranging in value from 0 to 1; a negative value for alpha indicates that items on the scale are negatively correlated and the reliability model is inappropriate).

Inter-Rater Reliability: When two or more people are recording their observations, one may indicate the reliability by saying the percentage of the times the two raters agree on their observations. Adjusting for chance agreement, one would use Fleiss' Kappa for inter-rater agreement. Kappa measures the agreement between two raters when both are rating the same object. The difference between the observed proportion of cases in which the raters agree and the proportion expected by chance is divided by the maximum difference possible between the observed and expected proportions, given the marginal totals. A value of 1 indicates perfect agreement. A value of 0 indicates that agreement is no better than chance.

Stability Over Time: If one is measuring a static dimension of people, making a repeated observation after, say, three weeks one can then correlate the two observations to see how stable that dimension is over time. Presumably, a dynamic risk factor would have poor stability over time. Pearson's or Spearman's correlation coefficients are used in this situation depending on the type of variable being examined. [Pearson's R is a measure of linear association between two variables. The value of R ranges between -1 (a perfect negative relationship in which all points fall on a line with negative slope) and +1 (a perfect positive relationship in which all points fall on a line with positive slope). A value of 0 indicates

See RISK PREDICTION, next page

RISK PREDICTION, from page 76

no linear relationship. Spearman's correlation coefficient is a product-moment correlation coefficient suitable for ordinal data. It is equivalent to ranking observations and computing a Pearson R.]

Criterion-Related Validity: Does the instrument measure what it says it measures? It is common to compare an instrument to a "gold standard" to indicate that they indeed measure the same thing.

The Appropriateness of the Statistical Inference

Given the variables being observed in conjunction with the sampling design, decisions must be made about decision rules and which statistical procedures will be used to draw conclusions. It would require several thick statistics books to survey all the procedures that could be used in the many different research situations. (For more information, the administrator or program person is referred to an introductory statistics book to obtain the necessary foundation.) There are several basic levels of statistical procedures commonly used: descriptive, inferential, and predictive.

At the descriptive level, rates, proportions, percentages, and averages are used along with bar graphs to tell a story numerically about the sample. There is no concern at this point to suggest that the sample is representative of a larger population.

By using simple random sampling or possibly a more complex sampling design it is possible to use a sample to infer something about a population from which the sample was drawn. Therefore the use of inferential statistics and a proper sampling design is essential if one desires to use the results of the analysis in the future with people from that same population. Because it is possible to draw any one of a multitude of possible random samples from a large population, there is going to be sampling variability to contend with. Typically, when the value of a population percentage or average is stated, it is stated to a particular level of confidence within the range of values called a confidence interval.

The predictive level of statistics is a variation of inferential statistics. Based on one or more factors and prior statistical results, a prediction is made for the outcome for an individual. The prediction is often stated in terms of a probability or an odds-ratio (see Figure 1 above). For example, based on prior research one may have an equation that uses gender, age, number of prior convictions for violent offenses, and years of education to estimate the probability that a particular person will commit a violent offense within two years of release.

Figure 1

Person	Gender	Age	Prior	Years Education	Odds Ratio
A	M	50	0	15	1.00
B	F	19	1	10	1.30
C	M	23	3	9	2.10

Figure 2

	Violent	**Not Violent**	
Dangerous	A=155 Hits True Positve	B=24 False Alarm False Positive	**A+B=179**
Safe	C=32 Miss False Negative	D=205 True Negative	**C+D=237**
	A+C=187	**B+D=229**	

These numbers indicate that person B is 30 percent more likely to commit a violent crime in the next two years than is person A, while person C is 110 percent more likely to commit a violent crime within that time frame than person A.

To say that a person is 2.1 times more likely than another to commit a violent crime is a bit deceptive if one is not aware of the base rates. Pretend we know that, out of 100,000 individuals with person A's characteristics, 8 of them will commit violent crimes in two years. This yields a probability of 8 per 100,000. So, the probability that a person of person C's group (with like characteristics) will commit a violent crime within the same time frame is 16.8 per 100,000. So, while the person C is far more likely to be one of the 21 people in his or her group of 100,000 who might commit a violent crime, the probability of violence in this situation is actually small because the base rate is quite small. As such, if one were to predict that none of the person C types would commit a crime, the prediction would be very close to the truth because of the low rate.

The example in Figure 2 is a case-control study of 416 people, 187 of whom are considered the "cases." This means that their outcomes were in the area of interest to the study, i.e., violent behavior. An assessment was made and each person was determined to be either dangerous or safe, and then was tracked for the next years for violent crimes.

After the year, 155 out of 179 of those assessed as dangerous were found to be violent, while 32 out of 237 of those assessed to be safe were in fact violent. Figure 3 on page 78 provides many of the descriptive statistics used in research to assess the viability of an assessment/predictive instrument.

Note again how the base rate affects the probabilities of a correct prediction. To quote from Fleiss: "If the disease is not too prevalent—if it affects, say, less than 1% of the population—the false negative rate will be quite small, but the false positive rate will be rather large." Figure 4 is from Fleiss and shows the effect of base rates on the probability of either a false positive or a false negative:

If this were a case of predicting violence, and the base rate were 1/100, then one should expect for every 100,000 declared "safe," 51 would turn out to be violent; but yet if the base rate were 1/1000, then only five would turn out to be violent. However, if the base rate were 1/100, then for every 100,000 declared dangerous, 51,000 would turn out to be non-violent; while if the base rate were 1/1000, then 91,300 would turn out to be non-violent.

The ROC

The ROC: Receiver Operator Characteristic (Kramer, M.S., *Clinical Epidemiology and Biostatistics: A Primer for Clin-*

See RISK PREDICTION, next page

Figure 3

Statistic	Formula		Comments
Correlation	(B*C-A*D)/((A+B)*(C+D)*(A+C)*(B+D))^0.5	-0.73	Strong relationship between violence and dangerousness
Relative Improvement Over Chance	(A*(A+B+C+D)-(A+B)*(A+C))/((A+B)*(B+D))	76%	This is 76% better than tossing a coin
Odds Ratio	A*D/(B*C)	41.4	A person rated dangerous is over 40 times more likely to be violent than a person rated safe to be not violent.
Chi-square	((A+B+C+D)*((A*D-B*C)^2))/((A+B)*(A+C)*(C+D)*(B+D))	220	With one degree of freedom this is a significant chi-square value
Base rate	(A+C)/(A+B+C+D)	45%	45% of the group is violent
Selection ratio	(A+B)/(A+B+C+D)	43%	43% of the group is rated dangerous
Hit rate/ sensitivity	A/(A+C)	83%	83% true positives
Specificity	D/(B+D)	90%	90% of the non-violent were rated safe
False alarm rate (1- specificity)	B/(B+D)	10%	10% false positives
Proportion correct	(A+D)/(A+B+C+D)	87%	87% of the ratings were correct
Positive predictive power	A/(A+B)	87%	87% of those rated dangerous were violent
Negative predictive power	B/(B+D)	10%	10% of those were not violent were rated dangerous

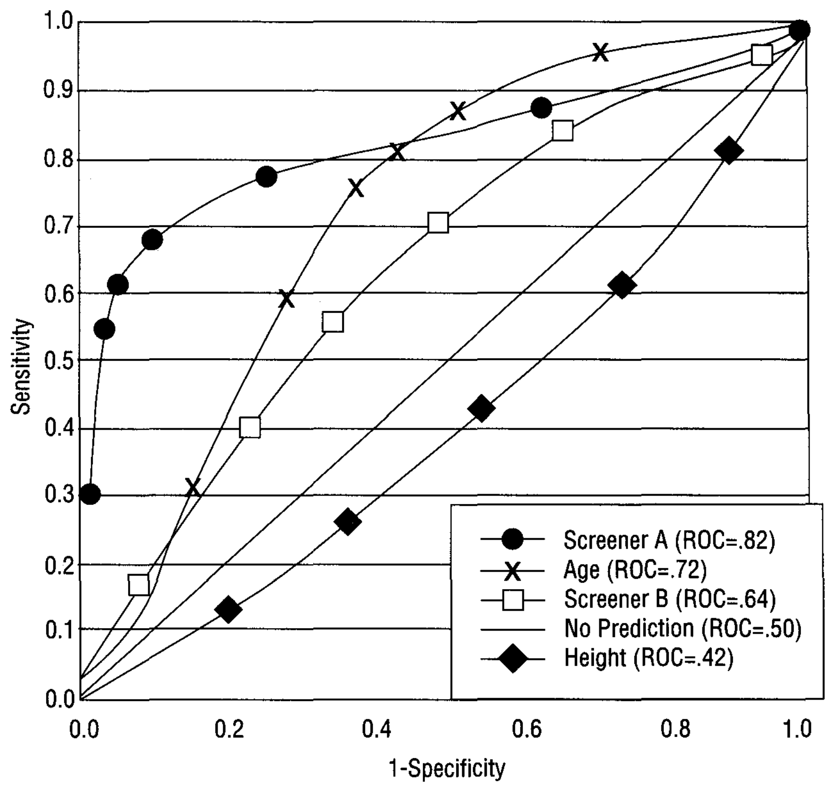
RISK PREDICTION, from page 77

ical Investigators and Decision-Makers (1991)) is a measure, now in common use, of the predictive value of an assessment instrument. In assessment instruments where higher values indicate higher values of abnormality, choosing a lower cutoff, for example, will result in greater sensitivity and will miss fewer cases. On the other hand, higher values of the cutoff will result in greater specificity, meaning less misclassification of people who will not commit further crime. This reciprocal relationship between sensitivity and specificity is always found when a cutoff point is chosen from a continuous scale. This relationship can be shown on the assessment instrument's "receiver operating characteristic" curve.

Figure 4
Error Rates Associated With a Screening Test

P(B) (Base Rate)	False Positive	False Negative
1/million	.9999	0
1/1000	.913	.00005
1/100	.510	.00051

Graph 1: Receiver Operating Characteristics Curve



See RISK PREDICTION, next page

Figure 5: Calculated Values Used in Graph 1

Height (cm)	Sensitivity	1-Specificity	Partial	Screener B	Sensitivity	1-Specificity	Partial
cutoff	0.00	0.00	ROC	cutoff	0.00	0.00	ROC
190	0.00	0.01	0.00	100	0.03	0.00	0.00
175	0.13	0.21	0.01	45	0.03	0.01	0.01
170	0.26	0.36	0.03	35	0.17	0.08	0.04
165	0.43	0.55	0.06	30	0.40	0.23	0.05
160	0.61	0.72	0.09	28	0.55	0.34	0.09
155	0.81	0.88	0.11	26	0.70	0.48	0.13
145	0.98	0.99	0.10	24	0.84	0.65	0.24
100	1.00	1.00	0.01	20	0.97	0.92	0.08
		Total ROC	**0.42**			**Total ROC**	**0.64**
Screener A	**Sensitivity**	**1-Specificity**	**Partial**	**Age**	**Sensitivity**	**1-Specificity**	**Partial**
cutoff	0.00	0.00	ROC	cutoff	0.00	0.00	
200	0.31	0.01	0.00	80	0.09	0.06	0.00
130	0.55	0.03	0.01	70	0.31	0.16	0.02
120	0.61	0.04	0.01	60	0.59	0.27	0.05
110	0.68	0.09	0.03	50	0.76	0.37	0.07
100	0.77	0.25	0.11	45	0.82	0.43	0.04
90	0.87	0.62	0.30	40	0.87	0.51	0.07
75	0.97	0.98	0.33	30	0.95	0.70	0.18
50	1.00	1.00	0.02	16	1.00	1.00	0.29
		Total ROC	**0.82**			**Total ROC**	**0.72**

RISK PREDICTION, from page 78

This is done on a graph of sensitivity versus one minus specificity. Examples of which are in Graph 1 and Figure 5. Two screening instruments, A and B, and two other variables, age and height, are included as examples. Figure 5 lists the cutoff values that were used in the calculations.

ROC is a measure of how well the instrument predicts by considering all possible cutoff values. The ROC value is simply the area under the ROC curve. Note that the statistical software SPSS version 9 includes routines to calculate these values. ROC can also be estimated by adding up a series of rectangular and triangular areas under the graph lines. For example, the area under the "no prediction" line is simply the area of a triangle that equals one half the product of the triangle's base and the triangle's height. This value is 0.50. Values of ROC close to one (the area of a square in this graph) mean that the instrument has an excellent predictive value. Values of ROC close to 0.50 have no predictive value. However, ROC values less than 0.50 and close to zero will result in too many false alarms (false positives).

In the example, Screening Instrument A has a ROC of .82, making it a very good prediction instrument overall. Screening Instrument B, on the other hand, with a ROC of .65 is not as good as simply using the person's age (ROC of .72). Using a person's height (ROC of .42) as a predictor is of little value since its ROC is less than and close to .50.

It is our hope that this attempt to provide a simplified overview of the science of risk prediction is useful to the reader. Sometimes simplifying a complex set of mathematical concepts leads to greater confusion. If more knowledge is sought, the reader is advised to review the following references.

References

Fleiss, J.L., *Statistical Methods for Rates and Proportions,* 2d ed., New York, NY: John Wiley & Sons (1981).

Ott, R.L., *An Introduction to Statistical Methods and Data Analysis,* 4th ed., Duxbury, MA: Duxbury Press (1993).

Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.A., *Violent Offenders: Appraising and Managing Risk.*, Washington, DC: American Psychological Association (1998).

Dr. Rodney V. McCormick is president of McCormick Research Services, and Dr. Henry R. Cellini is president of the Training and Research Services, Incorporated. Both can be reached through Training and Research Services at PMB 227, 8200 Montgomery NE, Albuquerque, NM, 87109; via email at tricellini@aol.com; via fax at (505) 299-9454; and phone at (505) 293-3058. ■