U.S. Department of Justice
Office of Justice Programs
National Institute of Justice

National Institute of Justice

Research Report

Predicting Pretrial Misconduct with Drug Tests of Arrestees

Evidence from Eight Settings

A final report submitted to the National Institute of Justice

Predicting Pretrial Misconduct with Drug Tests of Arrestees

Evidence from Eight Settings

by
William Rhodes
Raymond Hyatt
Paul Scheiman

September 1994

Prepared for the National Institute of Justice, U.S. Department of Justice, by Abt Associates Inc., under contract #OJP-89-C-009. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Table of Contents

List of Tables

List of Figures

Acknowledgements

This report has benefited from comments by Virginia Baldau, the National Institute of Justice, and by three anonymous reviewers. Dana Hunt, Jan Chaiken, and Steven Kennedy of Abt Associates also gave valuable advice. John Spevacek was NIJ's capable Project Monitor. Remaining errors are the authors' responsibility.

Executive Summary

The tension between due process and crime control is never more evident than at the stage of pretrial release. Innocent until proved guilty, a defendant has a conditional right to freedom pending his day in court. But pretrial release can jeopardize the public when the legally innocent but predictably dangerous returns, unrestrained, to the streets. At some point, the public's demand for security must be balanced against the defendant's demand to be free.

Provided judges can identify defendants at high risk of pretrial misconduct, they can either detain them if the risk is excessive or release them and reduce the risk by supervision or other special conditions. Some see testing arrestees for recent drug use as one way to distinguish between those who will and those who will not commit pretrial misconduct. But drug testing is expensive. To be worthwhile for this purpose, drug tests must improve predictions based on other, more readily available data, such as a defendant's criminal history and community ties. Research reported here questions whether the incremental predictive power resulting from drug testing can always pass this test.

The research is an analysis of eight data sets. Each recorded arrestees' post-release misconduct (arrests and failure to appear), urine test results, and other factors (especially criminal records and community ties). Data are from Washington, D.C. (adults, 1984 to 1985), Manhattan, New York (1984), Washington, D.C. (juveniles, 1986 to 1988), Dade County, Florida (1987), Prince George's County, Maryland (1988 to 1989), Maricopa County, Arizona (1988), Milwaukee County, Wisconsin (1989), and Washington, D.C. (adults, 1989 to 1990).

The investigation used survival analysis to study time until rearrest. It used a probit model to analyze the occurrence of a failure to appear. In both case, the analysis showed whether a positive test for cocaine, heroin, or other illicit substance improved the prediction of misconduct after accounting for defendants' criminal records, community ties, and other factors commonly known by the court. Findings were:

- A positive test for opiates helped predict rearrest. The combined result across eight data sets was statistically significant and substantively large. Although a positive test for cocaine helped predict misconduct in some settings, the effect was not statistically significant in a combined test across all settings. Other drugs showed no consistent predictive power.

- A positive test for cocaine helped predict failure to appear. The combined result across eight data sets was statistically significant and substantively large. Other positive test results showed no consistent predictive power.

Overall, then, some evidence shows that drug test results predict pretrial misconduct. The evidence is inconsistent, however. In some sites, drug test results appear to contribute nothing toward predicting misconduct. In other sites, some combination of drug test results seems to predict either rearrest or failure to appear, but seldom both. There is scant evidence that arrestees who test positive, whatever the drug for which they tested positive, are more likely than those who tested negative to commit pretrial misconduct.

One problem is that urine test results cannot distinguish between heavy users and those whose use is more moderate. This distinction is important because criminal behavior generally, and pretrial misconduct specifically, increases with heavy drug use. Without some measure of heavy use, the roughly 60 percent of arrestees who test positive for an illicit substance look identical for purposes of prediction. Urine tests cannot readily identify the minority who engages in pretrial misconduct.

Better testing procedures may eventually improve the ability to predict misconduct by distinguishing heavy users from more moderate ones. Meanwhile, courts might make better use of available information including repeated measures of drug use, histories of substance abuse and treatment, and self-reports of the need for treatment. Of course, the study says nothing about another reason for drug testing, namely, to identify those who are in need of treatment and to see that they receive that treatment under supervision of judicial authority. This latter purpose may by itself justify pretrial drug testing.

Introduction

The link between drugs and crime is complex. Although many offenders use illegal drugs, others do not, and although many drug users commit crimes, some avoid criminal activity other than possession of illicit substances. No compelling evidence indicates either that drug use causes crime or that crime leads to drug use (Gandossy et al. 1980; Wish and Johnson 1986; Chaiken and Johnson 1988; Chaiken and Chaiken 1990).

Nevertheless, criminal activity is related to drug use. Arrestees are more likely to use illegal drugs than people who are not arrested.[1] People who use drugs heavily are more likely to be arrested than people who do not use drugs.[2] Furthermore, offenders typically commit more crimes when they use drugs heavily than when they abstain (Ball et al. 1981; Speckart, Anglin, and Deschenes 1989). Sustained drug use is one marker, although an imperfect one, for criminality.

The link between drug use and criminality has implications for the pretrial processing of arrestees. If arrestees who test positive for recent drug use are more prone to commit crimes than those who test negative, then judges can take special steps to protect the community. Of course, they can detain arrestees who test positive for recent drug use, but less restrictive measures are available. These include

---

[1] The Drug Use Forecasting system data from 23 sites (collected by the National Institute of Justice) indicate that during 1990 roughly 43 percent of arrestees tested positive for cocaine, 19 percent for marijuana, and 10 percent for opiates. Authorities consider the test used by DUF to be a conservative measure of recent drug use (see Mieczkowski et al. 1993; Visher and McFadden 1991). A positive test means the arrestee has probably used cocaine within two or three days of the arrest. In contrast, our analysis of the National Household Survey on Drug Abuse (the Household Survey) for 1991 (sponsored by the National Institute on Drug Abuse) indicates that fewer than one-half of one percent of Americans admitted using cocaine on a weekly basis during the year before the survey. Even after accounting for underreporting on the Household Survey, results indicate that arrestees are much more likely than other citizens to use drugs.

[2] The 1991 Household Survey questions a representative national sample about drug use, criminal activity, and arrests. An estimated 625,000 Americans admitted using cocaine on a weekly basis. About 76 percent of the weekly users admitted either to having been arrested or to having committed a criminal offense (not including drug use) during the year prior to the survey. About 20 percent of all others admitted to an arrest or a crime. Statistics are based on tabulations performed by Abt Associates, Inc. Also see Harrison and Gfroerer (1992).

supervised release, continued drug testing (with sanctions for violating court orders to abstain from drug use), and drug treatment.

Thus, drug testing has potential for improving criminal justice handling of pretrial releasees. However, drug testing is expensive for the court, and mandatory drug testing may encroach on Fourth Amendment protection against illegal searches and seizures.[3] Because of these costs, courts must examine and weigh potential benefits before adopting pretrial drug testing.

For the potential to be actualized, at least one question must be answered affirmatively. Can drug testing help the court distinguish between arrestees who would not commit pretrial misconduct and those who would? Pretrial misconduct comprises crime while on release (reflected in pretrial arrests) and failure to appear for court dates. Unless the answer is yes, pretrial urine testing may be wasteful.[4]

Actually, the question needs to be narrowed. Judges already know defendants' backgrounds. Criminal records are typically available. In many jurisdictions court employees obtain information on defendants' employment histories and community ties. The narrower question is this: Given what is already known about the defendant, do results from drug testing contribute additional information? That is, on top of other information at the judge's disposal, can knowledge of recent drug use improve predictions of pretrial misconduct?

---

[3] There appears to be little question about the constitutionality of pretrial drug testing, but legal opinion seems to be based on a balancing test. Pretrial drug testing is tolerated because it promotes the public's need for protection at a lesser cost of reduced liberty. If pretrial urine testing were useless when predicting danger to the community, constitutional question might require reconsideration.

[4] Judges can, as a condition of pretrial release, induce defendants to enter substance abuse treatment. The best evidence indicates that many compulsive drug users enter treatment because of criminal justice processing and that such treatment is at least as effective as treatment entered voluntarily (Leukefield and Tims, 1988). This study does not address this objective of pretrial drug testing, which may provide compelling justification for urine testing regardless of whether those who test positive are most likely to engage in pretrial misconduct. However, this justification assumes that urine tests at the time of arrests identify mostly compulsive users, that there are no less expensive way to identify them, and that effective substance abuse treatment is available. This report has some implications for the first two criteria although it provides no independent evidence regarding the third criterion.

This paper analyzes eight data sets that record pretrial misconduct of arrestees who were tested for recent drug use when booked into jail. The analyses show that urine test results have no consistent power to predict misconduct across sites and over time. The evidence suggests that the reader who has decisionmaking authority over pretrial detention should not be detaining arrestees based solely on their urinalysis results at arrest. This caution is especially pertinent in jurisdictions that have not, using their own data, validated a relationship between urinalysis test results and pretrial misconduct.

Others have analyzed these data previously, and some have reached conclusions that differ from ours. We discuss the data and the previous analyses in the next section.

Data and Previous Analysis

Six diverse sites tested arrestees for recent use of several illicit substances. Table 1 summarizes the sites, the programs they operated, and the data they collected. Sample sizes are the number of observations that entered our analysis, not the total collected. An appendix provides additional detail. Washington, D.C. provided three different data sets, corresponding to three different settings. To avoid confusion, we call the first Washington setting "D.C. adults, 1984," denoting that the data pertain to adult arrestees in 1984. We call the second Washington setting "D.C. juveniles," because the data pertain to juvenile arrestees who were processed through lockups between October 1986 and January 1988. The third Washington setting is called "D.C. adults, 1989-1990," to indicate that the data pertain to adults who were arrested in 1989 and 1990.

TABLE 1
SUMMARY OF EIGHT SITES, THEIR PROGRAMS, AND THEIR DATA

SITE	SUBJECTS	DATES	SPECIAL CONDITIONS	DRUGS TESTED	NUMBER OF CASES
District of Columbia (Adults 1984)	Adults, except those arrested for federal and minor crimes	June 1984 to January 1985	Experiment: • periodic testing • treatment	cocaine opiates PCP amphetamines methadone	5,689
District of Columbia (Juveniles)	Juveniles processed through lockups	October 1986 to January 1988	Experiment: • weekly testing • bimonthly testing • monthly testing	cocaine opiates marijuana PCP	2,137
District of Columbia (Adults 1989 to 1990)	Adults interviewed by DUF	1989 to 1990	• drug testing	cocaine opiates PCP other drugs	1,538
Prince George's County, Maryland	Adults booked	July 1988 to February 1989	Experiment: • drug testing	cocaine opiates marijuana PCP	1,072
Milwaukee County, Wisconsin	Adults booked for felonies, serious misdemeanors and outstanding warrants	February 1989 to December 1989	Experiment: • drug testing	cocaine opiates amphetamines bensodiazepines	830
Maricopa County, Arizona	Adults booked for felonies	Beginning summer 1988	Experiment: • drug testing	cocaine amphetamines other drugs	186
Manhattan, New York	Adults booked for felonies for non-drug law offenses	April 1984 to October 1984	None	cocaine opiates PCP methadone	1,893
Dade County, Florida	Adults booked for felonies excluding some serious crimes	June 1987 to July 1987	None	cocaine marijuana	1,294

The pretrial release program in Washington, D.C. was the prototype for replication programs in three other sites: Prince George's County (Maryland), Milwaukee County (Wisconsin), and Maricopa County (Arizona).[5] The Washington program also spawned a drug testing program for juvenile arrestees in the District of Columbia. In all five settings, the court received arrestees' urine test results, which the court used to randomly assign some releasees to experimental post-release supervision programs. Researchers conducted urine testing in Manhattan (New York) and Dade County (Florida), but they did not give test results to the court. Additionally, researchers collected data in Washington, D.C. during a period subsequent to the experimental phase of the Washington, D.C. project.

Each site recorded data about those arrestees, including criminal records, community ties, and the results of urine testing. Each monitored subsequent arrests and failure to appear for court dates for defendants released before trial. These data afforded tests of whether pretrial misconduct is higher among those who tested positive for recent drug use than for those who tested negative.

Of course, judges in all sites routinely imposed conditions on the behavior of defendants who were released pending trial. Some sites had special programs for substance abusers who, typically, had to submit to urine testing during the pretrial period. (As mentioned, some of these programs were experimental.) Judges sometimes imposed sanctions for noncompliance. We account for this special supervision in this study.[6]

We used data from these six sites to determine whether drug test results predict pretrial misconduct after accounting for a defendant's criminal record, community ties, participation in special

---

[5] Pima County (Arizona) and Multnomah County (Oregon) were additional replication sites for the Washington, D.C. program. We were unable to acquire the Multnomah County data, and the Pima County data proved unsuitable for analysis. Consequently, data from those two sites are not included in the analysis reported in this paper.

[6] Defendants were often subjected to pretrial release conditions in addition to the experimental conditions described here. The data did not always describe those other conditions and, consequently, our analysis could not take conditions of supervision fully into account. These additional conditions may account for some of the patterns that we attribute to drug use and other factors.

supervision programs, and time at risk. Pretrial misconduct means a rearrest during pretrial release or a failure to appear for a court date.[7] (We analyzed them separately.)

Researchers have previously analyzed these same data. In Visher's review of those analyses (1992), she concludes that ". . . drug test results appear to improve the classification of defendants according to the risk of pretrial misconduct in three sites: Washington, D.C., Manhattan, and Dade County" (p. 117). She also concludes that "At two of these four sites [that replicated the D.C. experiment], defendants who tested positive for illegal drugs at arrest were at significantly greater risk of pretrial arrest or FTA, after taking into account factors usually considered by the arraignment judge" (p. 123).

Visher's comments are thoughtful and useful. However, our own conclusions based on a review of those analyses are equivocal. Tables 2 and 3 summarize regression results from seven of those studies, which are discussed below.

Researchers have adopted various specifications for using drug test results in their regression models, and the tables show the specification used in each study: four used a probit model, two used a logistic model, and one used survival analysis. The tables report a t-score test statistic for each parameter estimate associated with the drug-test measure used in the regression. The absence of a t-score means that a drug test measure (reported in column 1) was not used in the analysis (identified in row 1). A heavy box appears around a t-score greater than 1.96; a lighter box (in interrupted line) appears around a t-score greater than 1.64 but less than 1.96.[8]

Some researchers used probit as the estimating technique. For these studies, a positive t-score implies a high risk of rearrest or of failing to appear for a court date. Other researchers used logit as

[7] Data documentation does not always define "failure to appear." Although failure to appear seems to mean a "willful failure," generally resulting in a bench warrant, courts probably differ in their tolerance for missed court dates. Differential tolerance may account for some of the patterns observed here.

[8] A t-score in excess of 1.96 implies that a parameter is statistically significant at $p < 0.05$. A t-score in excess of 1.64 implies that a parameter is significant at $p < 0.10$.

Table 2 – Summary of Regression Results from Previous Analyses of Pretrial Rearrest Rates													
	Goldkamp, Gottfredson, and Welland Dade County probit		Smith, Wish, and Jarjoura Manhattan probit				Toborg, Bellassai, Yezer, and Trost D.C. probit	Goldkamp, Jones, and Gottfredson Prince George's County logit	Goldkamp, Jones, and Gottfredson Milwaukee County logit	Rhodes D.C. (juveniles) survival model	Smith and Polsenber Washington, D.C. probit		
Cases	1374	1374	1967		1967		3852	352	299	2137	1284	1284	1284
Specification	model 1	model 2	model 1	model 2	model 1	model 2	model 1	model 1	model 1	model 1	model 1	model 2	model 3
heroin			1.75	1.78			1.51			0.24			
cocaine	-1.06	-1.00	0.69	0.89			-0.48	-1.26		-1.46	4.85		
PCP			3.23	3.39			1.67			0.54			
methadone			-0.96	-0.85									
marijuana	0.00									0.04			
cocaine or marijuana	2.03	2.32											
cocaine and marijuana		0.09											
opiates and cocaine							-1.08						
opiates and PCP							0.27						
PCP and cocaine							0.53						
three or more drugs							2.63						
any positive								1.70	-0.86			4.29	
none positive								-1.65					
1 positive								1.33					
2 positive													
3 positive													
Number positive					2.58	2.88							4.00
2 or more positive								1.40		0.44			

Note: Model 2 is a reduced form specification of model 1, except in Date County, where model 2 adjusts for selection bias while model 1 does not.

Table 3 -- Regression Results from Previous Analyses of Pretrial Failure to Appear Rates

	Goldkamp, Gottfredson, and Welland Dade County probit		Smith, Wish, and Jarjoura Manhattan probit				Toborg, Bellassai, Yezer, and Trost D.C. probit	Goldkamp, Jones, and Gottfredson Prince George's County logit	Goldkamp, Jones, and Gottfredson Milwaukee County logit	
Cases	1374	1374	1965	1965	1965	1965	3852	352	260	260
Specification	model 1	model 2	model 1	model 2	model 1	model 2	model 1	model 1	model 1	model 2
heroin			1.66	1.61			1.84			
cocaine	0.84	1.10	1.88	1.86			3.45	-0.26	0.25	2.09
PCP			0.31	0.75			-1.76			
methadone			-0.02	-0.08						
marijuana	-0.99									
cocaine and marijuana		-0.20								
cocaine or marijuana	-0.56	-0.70								
opiates and cocaine							2.62			
opiates and PCP							-1.64			
PCP and cocaine							0.52			
three or more drugs							-0.57			
any positive								-0.36	0.01	
none positive								0.33		
1 positive								-0.66		
2 positive										
3 positive										
0 or 1 positive									0.13	
2 or more positive								0.25	-0.18	
Number positive					2.75	2.92				

Note: Model 2 is a reduced form specification of model 1, except in Dade County, where model 2 adjusts for selection bias while model 1 does not.

the estimating technique. For these studies, a negative t-score reflects a higher risk. A positive t-score reflects a higher risk in the single study that used an exponential survival model.

Washington, D.C.

Toborg et al. (1989) analyzed data for adult arrestees (during 1984) in the District of Columbia. They concluded that "urine test results make a consistent, significant, incremental contribution to pretrial risk classification for arrestees in the District of Columbia." This summary is too strong, however. Toborg and her colleagues found that a positive test for PCP predicted rearrest, but the effect was weak,[9] and curiously, when used in combination with other drugs PCP use did not predict a rearrest. A positive test for three or more drugs predicted rearrest, but fewer than 2 percent of defendants were positive for three or more drugs. Furthermore, no other drugs predicted a rearrest.

Their conclusion is more justified for failure to appear. They reported that the likelihood of failure to appear increased with the use of cocaine and opiates.[10] Inexplicably it decreased with the use of PCP.[11]

Visher and Linster (1990), who analyzed a subset of the same adult rearrest data using a different technique, report the following: "Of particular interest . . . is the relative importance of the drug test variables as risk factors. . . . during the early months after release the model associates a substantially increased risk for drug use" (p. 169). This finding seems to support the Toborg conclusions. However, Visher and Linster also report that "the relative risk is somewhat less for subjects testing positive only

[9] The t-statistic for PCP was 1.67. This is significant either at 0.05 in a one-tailed test of significance or at 0.10 in a two-tailed test.

[10] The t-score for cocaine was 3.4. It was 1.8 for heroin and 2.6 for cocaine and heroin used in combination. The authors claimed that a positive test for cocaine increased the probability of a failure to appear by about 0.15. According to our calculations the increment is somewhat less than 0.11--still an appreciable size by either account.

[11] The t-score for PCP was 1.8.

for PCP than it is for cocaine, amphetamines, or opiate users . . . " (p. 169). This conclusion differs from that of Toborg and her colleagues.

Smith and Polsenberg (1992) conducted still another analysis of Washington, D.C. data. Their data comprised subjects who were arrested during 1990.[12] Their analysis was limited to rearrests.

Unlike Toborg and colleagues, who examined the narrow issue of misconduct during pretrial release, Smith and Polsenberg answered the broader question: Does a positive urine test for cocaine (or for illicit substances generally) predict rearrest during a follow-up period that might extend beyond the date of the arrestee's case termination?

They concluded that ". . . the presence of drugs in a urine sample taken shortly after arrest (i.e., drug-positive arrestee test results) is significantly associated with an increased probability that the person subsequently will be arrested for a new crime" (p. 376). However, their statistical analysis supports only the narrower conclusion that a positive test for cocaine predicts a rearrest, because they did not include drugs other than cocaine in their regression specifications.

Replications of the Washington, D.C. Program

Goldkamp et al. (1990b) analyzed data from Prince George's County. They conclude that drug test results are not useful for predicting rearrest or failure to appear for a court date. However, their regression analysis seems to show some predictive power from knowledge of drug test results, as shown by t-scores that are greater than 1.64.[13]

---

[12] By this date, the District had abandoned the experimental phase of post-release supervision. Judges were free to assign conditions of post-release supervision as they saw fit.

[13] The Goldkamp et al. analysis is difficult to interpret, because they appear to have used a regression model specification in which their drug test variables are highly correlated. This high correlation tends to make the t-scores small even when drug testing is a useful predictor of misconduct. Nevertheless, for the regression based on rearrest as a dependent variable, the researchers' measure of association (pseudo-R^2) increased from 0.49 before drugs were included in the regression to only 0.51 after drugs were included in the regression. This pseudo-R^2 is not affected by collinearity. The increase implies that drug test results cannot make a large improvement in prediction of rearrests. The pseudo-R^2 did not improve for the prediction of failure to appear.

They also analyzed data from Milwaukee. They conclude that drug test results are not useful for predicting rearrest. A positive test for cocaine helped predict failure to appear, but the effect was deemed small. Furthermore, the parameter estimate had the wrong sign, implying that arrestees who tested positive for cocaine were more likely to appear for court dates.

The D.C. Pretrial Services Administration also set up an experimental pretrial drug testing program for juveniles. Data were collected over a 15-month period. In his analysis of the D.C. juvenile data, Rhodes (1991) reported that drug test results did not help predict rearrests.

Other Settings

Wish, Cuadado, and Magura (1988) collected and first analyzed pretrial misconduct data from Manhattan. They reported that failure to appear is especially high among those arrestees who test positive for drug and admit they need drug treatment. FTA appears less likely for arrestees who test positive but deny the need for treatment. Those arrestees who test negative had the lowest FTA rates.

We comment later on this important study. Here, however, we note two points. First, the researchers seem to have included in their analyses the records of arrestees who were not at liberty prior to case disposition, that is, those whose cases were dismissed prior to release. We are uncertain how this affected the analysis. Second, the researchers deemed the statistical analysis to be preliminary.

Subsequently, Smith, Wish and Jarjoura (1989) reanalyzed the Manhattan data using more rigorous statistical methodology. According to their reanalysis, the number of drugs for which a defendant tested positive helped predict failure to appear. Tests for cocaine and for heroin both had modest correlations with failure to appear. The number of positive drug tests also helped predict rearrest, and arrestees who tested positive for PCP had significantly higher rearrest rates, while those who tested positive for opiates had marginally higher rearrest rates.

Belenko, Mara-Drita, and McElroy (1992) also reanalyzed these data and concluded that ". . . at least in New York City, universal drug screening of arrestees is not likely to be a cost-effective mechanism for identifying defendants at high risk for pretrial failure to appear." (p. 3)

This conclusion is overdrawn. Derived from four regression equations, it seems to rest on two observations. The first is that the parameter associated with a positive drug test is smaller than both the parameter associated with the release recommendation (which is based on several factors) and the parameter associated with an indication of prior failures to appear. However, this observation is subject to interpretation.[14] Moreover, observing that a positive drug test is only the third best predictor is not tantamount to rejecting it as being a useful predictor. Belenko's second observation is that prediction accuracy does not increase demonstrably when drug test results are added to a regression that already includes both the release recommendation and an indication of failure to appear. Unfortunately, this comparison is invalid, because it is based on two overlapping, but different, data files.[15] One cannot determine the incremental predictive power of drug tests from the results presented in his paper.

[14] The importance of a parameter to prediction cannot be judged by the size of that parameter relative to the size of other parameters. One must also consider the variance of the independent variable with which the parameter is associated. Researchers often use standardized parameters (sometimes called beta scores) for this purpose. Furthermore, Belenko does not report the statistical significance of any parameters.

[15] The first model specification comprised the following variables: release recommendation, prior failure to appear, affidavit charge, and age. The second model included these four variables and also an indication of a positive EMIT test. The first regression was based on 3,140 cases, and the second was based on only 2,418 cases.

Finally, Goldkamp, Gottfredson, and Weiland (1990a) analyzed Dade County data. Drug test results were not helpful when predicting failure to appear, but they contributed somewhat toward predicting rearrests.[16]

Summary of Earlier Findings

Overall, then, some evidence indicates that drug test results can help predict pretrial misconduct. The evidence is inconsistent, however. In some sites drug test results appeared to contribute nothing toward predicting misconduct. In other sites, some combination of drug test results (such as the number of drugs for which the tests were positive) or some individual drug test results (such as cocaine or heroin) seemed to predict either rearrests or failure to appear, but seldom both. There is scant evidence that arrestees who test positive, regardless of the drug for which they tested positive, are more likely than those who tested negative to commit pretrial misconduct.

Comparing and contrasting findings across these studies is complicated because researchers have chosen different ways to specify their models There is no standard for specifying the drug-test variables, no common set of control variables, and no agreement about the statistical issues that must receive attention during parameter estimation. Conducting meta-analysis and making unequivocal summary statements is hindered by this lack of standardization.

Recognizing this problem, the National Institute of Justice asked us to reanalyze these data by using a common model specification and statistical procedure. This reanalysis tests two null hypotheses. The first states that drug tests cannot help predict pretrial arrests after taking into account other factors

[16] The exact effect is difficult to discern. The regression model has a dummy variable for a positive cocaine test, a dummy variable for a positive marijuana test, a dummy variable coded one when either the cocaine or marijuana test was positive, and apparently a dummy variable that indicated that the drug tests were positive for both. This specification is collinear. Because this model could not possibly be estimated using ordinary least squares regression, we conclude that the authors must have estimated a model that differed somewhat from what was represented in the report. Our interpretation is that this fourth variable did not actually enter the model.

(criminal records, community ties, post-release supervision, and time at risk). The second hypothesis is similar: drug test results cannot help predict failure to appear after taking other factors into account.

Analytical Issues Regarding Pretrial Misconduct

We discuss analytical issues that arise when analyzing pretrial misconduct. Some issues are related to the way that the dependent variable is specified. We chose a survival model to analyze rearrest data. This choice forced us to deal with other problems, including missing data for the time of release until rearrest, unmeasured heterogeneity, and selection bias. These problems and our approach to overcoming them are described in the next section. We chose a probit model to analyze failure to appear. Additional issues that arise when analyzing failure to appear are discussed subsequently.

Rearrest

There are numerous ways to specify the dependent variable. One way is to treat a rearrest as being measured on a nominal scale: it either occurred or it did not occur. All but one of the original analyses followed this approach, and either probit or logit was the estimation method.

Probit and logit are acceptable for analyzing site-specific data. However, findings from one site are not readily comparable to findings from other sites unless the analysis incorporates some measure of time at risk (in this case, the time that an arrestee is free and awaiting trial). Rearrest rates cannot be compared across sites without this measure because they are sensitive to defendants' opportunities to commit crimes, and opportunity is a function of time at liberty. Few studies included a measure of time at risk, a limitation that hampers making cross-site comparisons from published results.[17]

---

[17] When the analysis is specific to a single site, the analyst need not take into account time at risk unless time until the end of the follow-up period is correlated with the regression's independent variables.

We adopted a survivor model, using time until rearrest as the dependent variable. This approach is more direct than introducing time at risk as a variable in a probit or logit model.[18] A survivor model has a second advantage. Treating the timing of rearrest as a dependent variable retains more information than do dichotomous specifications (such as the probit and logit models), which disregard the timing of events.[19] Using a survivor model should produce more precise parameter estimates.

With time until rearrest as the dependent variable, several options exist for conducting the analysis (Kalbfleisch and Prentice, 1980; Lancaster, 1990). A parametric survival model based on an exponential decay function is used here. Justification follows from inspection of Kaplan-Meier survivor plots.[20] For one site (Maricopa County, Arizona), the assumption of an exponential decay function was questionable. In this one site, we also used a Cox proportional hazard model, which requires no explicit parametric assumptions about the decay function.[21]

---

[18] Time at risk is not equivalent to the length of a study's follow-up period. Defendants are only at risk until their case reaches a disposition. Average disposition times may vary with defendant characteristics, and average disposition times differ across sites.

[19] Another way to specify the dependent variable is to count the number of rearrests during the period of pretrial release. (Regression based on the poisson process might be appropriate.) However, the number of rearrests (some of which may result in confinement) affects the amount of time at liberty to commit crimes, so the number of rearrests is a questionable specification for the dependent variable.

[20] The exponential decay function has a constant hazard function. (The hazard is often defined as the instantaneous probability of a failure (rearrest) by individuals who have not yet failed.) Although inspection of these data seems to justify use of the exponential decay function, Visher and Linster's (1990) analysis suggests that the hazard (for releasees in Washington, D.C.) is time variant. Their finding might be attributed partly to their assumption that all heterogeneity is measured. (We discuss measured and unmeasured heterogeneity later in this paper.) Thus, there may be no conflict between their evidence that hazards are time variant and our assumption that they are time invariant when unmeasured heterogeneity is taken into account. At any rate, the approach used by Visher and Linster is impractical for our application: It requires larger data sets than those available to us, and its properties are unknown when unmeasured heterogeneity and selection bias are introduced.

[21] Using a Cox proportional hazard model for the entire analysis was considered. However, although it is suitable for estimating the additional predictive power of drug testing, it does not appear to be readily adaptable to dealing with unmeasured heterogeneity and selection bias. (See Lancaster, 1990, p. 263.) Stephen Kennedy convinced the analysts that our attempts to adapt the proportional hazard model were in error.

The rest of the discussion serves to develop extensions of this basic model. Let t represent the time until rearrest for a defendant with characteristics described by a column vector X. (X includes information about drug tests, control variables, and a constant.) The variable t (the timing of a rearrest) has the density function:

$$\Phi(t) = \lambda e^{-\lambda t} \qquad \textbf{(equation 1)}$$

where $\lambda = e^{\beta X}$ and β is a row vector of unknown parameters. An i subscript, denoting the i^{th} observation, is implicit. The probability that the arrestee is arrest free as of time T is written:

$$P(T<t) = e^{-\lambda T} \qquad \textbf{(equation 2)}$$

To be recorded in these data, a rearrest had to have occurred either before the end of the follow-up period or before the defendant's case disposition. Observations are said to be "censored" when the defendant was arrest-free. Thus, when no arrest occurred, T denotes the time of censoring.

The parameters can be estimated provided either t (when an arrest occurred) or T (when the observation was censored) are known. Maximum likelihood procedures, programmed in GAUSS, are used to estimate the β parameters and their asymptotic standard errors. Equation (1) is the likelihood when an arrest occurred. Equation (2) is the likelihood when censoring occurred.[22]

---

[22] The timing of censoring events is assumed to be independent of the timing of rearrest. For example, case dispositions are not accelerated for defendants who are deemed likely to be rearrested during the pretrial period, and the length of the follow-up period does not depend on the probability that a defendant will be rearrested.

Although equations (1) and (2) comprise the ingredients of the basic model used in this analysis, several special problems arose. All the solutions developed are dependent on the exponential model's appropriateness to this study.[23]

Problem 1: Missing data for t. In some sites, an appreciable amount of data on the timing of a rearrest was missing. Discarding those cases as "missing" would have biased the results, because a case could be missing only when an arrest had occurred. Instead of discarding the case, we used the more limited information that a rearrest had occurred sometime during the follow-up period.

Using this more limited information requires a third term in the likelihood function. The probability that an arrest occurred sometime during the follow-up period is written:

$$P(T \geq t) = 1 - e^{-\lambda T} \qquad \textbf{(equation 3)}$$

Thus, the likelihood function used equations (1) when an arrest occurred and the date was known, (2) when censoring occurred, and (3) when an arrest occurred and the date was unknown.

This adaptation of the exponential survivor model generalizes to data files where we know T but not t. Such data are often analyzed using a probit or logit model. As we mentioned above, the probit and logit models (which specify the dependent variable as a dichotomy) discard information about the timing of events. In this instance, they discard the timing of the end of the follow-up period. The survivor model used here employs all the data, making it theoretically superior to probit and logit approaches for this application.

---

[23] Maltz (1984) has argued for using a split-population model for recidivism data. Other researchers (Schmidt and Witte 1988; Rhodes 1989) have applied this model to similar problems. However, Rhodes has argued that interpretation of parameters from the split-population model is ambiguous. At any rate, with the data at hand, only the left end of the distribution can be estimated (about 15 percent of the defendants are rearrested), so the split-population model cannot be readily distinguished from an exponential model with a steeper decay.

This modification has another advantage over the probit approach. Parameters estimated by using the probit procedure are proportional to an unknown scale factor. Because this scale varies across sites, parameter estimates are proportional from one site to another even when the underlying population parameters are equivalent. In contrast, parameter estimates based on equations (1), (2), and (3) are measured on the same scale across sites. Thus, even when we measure a rearrest as a dichotomy (that is, when the timing of the rearrest is unknown), the parameter estimates are measured in the same units across sites.

Problem 2: Unmeasured heterogeneity. As specified to this point, the model assumes that the X vector captures all differences across defendants, or "measured heterogeneity." In reality, X probably fails to capture all differences across defendants. "Unmeasured heterogeneity," that is, differences not captured in X, have implications for the analysis: parameter estimates will be biased when estimation occurs in the presence of unmeasured heterogeneity and censoring.

The bias resulting from unmeasured heterogeneity increases as time until censoring (time until disposition or the end of the follow-up period) decreases.[24] Unmeasured heterogeneity probably makes little difference for site-specific analyses, provided we consider estimates to be conditional on the length of the follow-up period.[25] However, it complicates cross-site comparisons. Disposition times vary markedly across the sites, and consequently, so does the bias. Vaiid comparisons require that unmeasured heterogeneity be taken into account.

[24] Unmeasured heterogeneity has no effect when the data are not censored.

[25] Nevertheless, analyses that do not account for unmeasured heterogeneity will result in biased parameter estimates even within a single site. The bias is probably unimportant if the analyst wants to predict pretrial outcomes for a similar follow-up period. If the analyst develops a prediction instrument based on a 90-day follow-up and uses this to predict for a 120-day follow-up, the resulting predictions will overstate rearrest rates. If the analyst uses this prediction for a 30-day follow-up, the resulting predictions will understate the rearrest rate. Thus, the extent to which unmeasured heterogeneity is a problem depends on how the analyst plans to use the results.

One solution is to rewrite λ as $\lambda = e^{\beta X + \epsilon}$ where ϵ is an error term that represents unmeasured heterogeneity. There are several approaches to introducing unmeasured heterogeneity into a survivor model. (These approaches are often called mixture models. See Lancaster 1990; Yamaguchi 1986) Here we assume that ϵ is distributed identically across defendants, independent of X, and as normal with a mean of $-0.5\sigma^2$ and a standard deviation of σ.[26] Then $\lambda = e^{\beta X}$ is replaced with $\lambda = e^{\beta X + \epsilon}$ in equations (1), (2), and (3). Because ϵ cannot be observed, the resulting likelihood function is estimated after integrating over ϵ, so we rewrite (1), (2), and (3) as:

$$\Phi(t) = \int_{\epsilon=-\infty}^{\epsilon=\infty} \lambda e^{-\lambda t} \phi(\epsilon) d\epsilon \qquad \text{(equation 4)}$$

$$P(T<t) = \int_{-\infty}^{\infty} e^{-\lambda T} \phi(\epsilon) d\epsilon \qquad \text{(equation 5)}$$

$$P(T \geq t) = 1 - \int_{\epsilon=-\infty}^{\epsilon=\infty} e^{-\lambda T} \phi(\epsilon) d\epsilon \qquad \text{(equation 6)}$$

Here ϕ is the normal density function. Unfortunately, equations (4), (5), and (6) have no closed form expression. Their solution requires numerical methods.

The correction for heterogeneity is sensitive to the assumptions made about the distribution of ϵ. Alternative assumptions would probably lead to parameter estimates that differ somewhat from those estimated here. However, the normal distribution is consistent with prior research on criminal careers

[26] These assumptions assure that $E[e^{\epsilon}] = 1$. In turn, this expectation assures that, when data are uncensored, β will be estimated consistently whether or not unmeasured heterogeneity occurs.

(Chaiken and Chaiken 1982; Blumstein et al. 1986; Spelman 1994).[27] If ϵ is normal with a mean of $-0.5\sigma^2$ and a standard deviation of σ, then λ is distributed as lognormal with a mean of $\beta X-0.5\sigma^2$ and a standard deviation of σ. This is to say that λ is skewed. Assuming that ϵ is normal is also convenient for dealing with the next problem, selection bias.

Problem 3: Selection bias. Recognition of unmeasured heterogeneity leads to a third problem. What if judges can observe ϵ (at least partly) while researchers cannot? What, furthermore, would happen if judges act on their information about ϵ? Specifically, what if judges tend to detain defendants who they predict are more likely to engage in pretrial misconduct, and their predictions depend on both X and ϵ? Selectivity by judges might lead to what analysts know as selection bias (Maddala, 1983).

Selection bias may be relatively unimportant in analyzing data from a single site,[28] but it complicates cross-site comparison. Suppose judges in one site know more about ϵ than judges in another site, or suppose judges are more willing or able to act on ϵ in one site than they are in another site. Then

---

[27] Researchers generally have little or no theoretical or empirical basis for specifying the distribution of ϵ or how it enters the model. Some (Flinn and Heckman 1982) have recommended nonparametric approach, but these are not always applicable. Unfortunately, parameter estimation can be very sensitive to the parametric model adopted (Trussell and Richards 1985). The best approach is to adopt distributional assumptions consistent with theory and empirical evidence, and we have taken that approach here.

[28] The importance of selection bias depends on the uses that the researcher intends to make of his or her findings. See Rhodes (1985).

When selection bias occurs but it is ignored in the analysis, the parameter estimates from the survival analysis are reduced form parameters. That is, the estimates are a combination of parameters from both the exponential survival equation and the selection equation. When the selection equation differs across sites, these reduced form parameter estimates cannot be readily compared.

Thus, the reason for taking unmeasured heterogeneity and selection bias into account is to estimate structural equations instead of reduced form equations. The structural equations reflect releasee behavior. With structural equation, we can predict releasee misconduct controlling differential treatment of defendants by criminal justice officials. The reduced form equations reflect a combination of releasee behavior and actions and reactions by criminal justice authorities--for example, a tendency to detain high-risk arrestees. The reduced form equations only allow us to predict misconduct conditioned on those actions and reactions. Because one reason for developing predictions is to modify CJS authorities actions and reactions, the structural equations have more value than the reduced form equations.

parameter estimates are not directly comparable across sites: judicial selectivity will lead to cross-site differences in parameter estimates even when population parameters are identical across sites.

Dealing with selection bias requires two steps. The first is to model the process by which judges make release decisions. The second is to use this first model of how judges make release decisions to adjust the statistics derived from the survival model.

The model of judicial decision making that is used here assumes that release depends on the value of a latent variable Z, where $Z = X\alpha + \epsilon_2$. Here α is a column vector of parameters, and ϵ_2 is a random error term. The judge releases the arrestee when $X\alpha + \epsilon_2 > 0$ or, equivalently, when $\epsilon_2 > -X\alpha$. Thus the probability of release is written:

$$\textit{probability of release} = \int_{-\alpha X}^{\infty} \phi(\epsilon_2)d\epsilon_2 \qquad \textbf{(equation 7)}$$

If ϵ_2 has a normal distribution with mean of 0 and a standard deviation of 1, then α is estimated usir maximum likelihood (a probit model).

Further, our model assumes that ϵ and ϵ_2 have a bivariate normal distribution. Based on these assumptions equations (4), (5) and (6) are rewritten so that the density of ϵ is conditional on $\epsilon_2 > -X\alpha$. That is, $\phi(\epsilon)$ is replaced with:

$$\phi(\epsilon|\epsilon_2 \geq -\alpha X) = \int_{\epsilon_2=-\alpha X}^{\epsilon_2=\infty} \phi_b(\epsilon,\epsilon_2)d\epsilon_2 \qquad \textbf{(equation 8)}$$

where ϕ_b is the bivariate normal density function. As before, no closed form expression is available, and solution of the likelihood function requires numerical methods. We must estimate one additional parameter (ρ), that is, the correlation between ϵ and ϵ_2.

Implementation problems. In one jurisdiction the timing of rearrests did not appear to follow an exponential distribution. Although the parameters were estimated as if rearrests followed an exponential distribution in any case (because it was assumed that the exponential hazard approximated the true hazard), a Cox proportional hazard model was also estimated as a check. For this one site (Maricopa County), the approach developed to compensate for unmeasured heterogeneity and selection bias was inoperative[29] and was not applied.

Adjustments for selection bias require that the system of equations be "identified." (Equations 4, 5, and 6 comprise one system; equation 7 is the second system.) In this context, identification requires that some parameters in the β vector be zero when their counterparts in the α vector are not zero. The data available for analysis do not always meet this condition.[30] Consequently, correcting for selection bias was sometimes impracticable.

---

[29] Our approach to dealing with unmeasured heterogeneity (and hence with selection bias) is sensitive to the exponential distribution being the correct distribution. In Maricopa County, σ was estimated as zero. This result is implausible because it is unlikely that X captures all variation across defendants.

[30] If judges take pretrial misconduct into account when deciding on release, then the same variables that predict misconduct could predict release. That is to say, the β parameter vector would never have a zero element when the α vector is nonzero. In a mathematical sense, such a restriction is too severe. Nonlinearities in the system of equations can be used to identify the parameters. Smith, Wish, and Jarjoura (1989) used this justification in their analysis of the Manhattan data; Toborg et al. (1989) apparently used this justification in their analysis. But as Maddala (1983) has pointed out, this is not a very satisfying identification restriction.

One way to identify the equations is to use the identity of the judge (or other bail official) in the probit model. This identification restriction was used by Rhodes (1985) in his study of pretrial misconduct in Federal courts. However, we discovered that in some courts a single judge is responsible for most bail hearings. Consequently, this solution would not work in all sites.

A second method for identifying the equation is to exploit special properties of the data in two sites. In Prince George's County and in Milwaukee County the researchers collected drug test results before and after implementing an experiment. Before the experiment, the drug tests were not made known to the judge, so the drug tests could not affect the release decision. After implementing the experiment, the drug test results were made known to the judge and did influence the release decision. This special property of these data allow us to identify the parameters. Essentially, drug test results are a different variable for purposes of the α and β vectors.

A third approach assumes that release practices vary over time, so that time itself could identify the equations. This approach was abandoned, however, when it was discovered that a single judge tends to make all release decisions. With no evidence of rotation among judges at the pretrial stage, little justification remained for using elapsed time to identify parameters.

Release dates were sometimes missing from the data. Since evidence indicates that releases generally occur within a few days of arraignment for all but few defendants, our solution was to assume that release occurred on the date of arraignment when the release date was missing. However, a dummy variable was introduced into the regression specification when this assumption was made. We expected that the parameter associated with this dummy variable would be positive, partly correcting for the shorter time at risk. That is, to the extent that estimated release dates are wrong, defendants will appear to be less recidivistic than they are in practice.

Finally, we had to make numerous assumptions about recoding variables. The important decisions are documented below. If the reader's interest requires further clarification, computing codes (in SPSS and GAUSS) are available for review.

Failure to Appear

Analysts have used probit and logit models to analyze failure to appear. (The dependent variable is a dichotomy.) We follow that practice (using a probit model), although the approach is not altogether satisfactory.

The basic problem is that defendants can fail to appear for a scheduled court date only when they have an opportunity to do so. Opportunity varies both across defendants and across sites. It varies across defendants, for example, because some have their cases concluded almost immediately. Dismissals are typical, but early guilty pleas occur often. An early case termination (especially a dismissal) means that the defendant has very little opportunity to fail to appear. It varies across sites, for example, because every court has its own pace for administering justice. When delays are long, opportunities for failing to appear may be negligible for short follow-up periods. When court delay is short, the probability of a failure to appear may be greater.

We find no good way to account for these differences. Ostensibly, we might introduce a variable that controls for the method of case disposition (or that the case is still open), or even a variable that reflects time until case disposition (or, again, that the case is still open). The problem with either approach is that the occurrence of a disposition, and its timing, are both themselves functions of failure to appear. That is, defendants who abscond are least likely to have dispositions; also, they are most likely to have lengthy case disposition times. Including either the method of case disposition or the time until case disposition in the statistical model confuses the causal order. In short, this approach does not work,[31] and we have no good control for the opportunity to fail to appear for scheduled court dates.

It would be valuable if the data provided a measure of the number of court dates that the defendant faced. Even better, it would be useful to know the nature of those court events. Many may be nonthreatening (evidence hearing as opposed to sentencing hearing, for example) and pose little risk of a failure to appear. Most of our data files exclude such information (Manhattan data provide a count of scheduled court dates), so analysis is necessarily ambiguous.

Another problem is that defendants may be reincarcerated during the pretrial period. Reincarceration is most likely to occur if the defendant is rearrested. Once the defendant is confined, the probability of a failure to appear falls to near zero. Unfortunately, our data are not very informative about rearrest, reconfinement, and re-release. Anyway, statistical analyses of data with two "competing events" are extremely complicated.[32] Given other problems with these data, we did not attempt to deal with those complications.

Another reason that defendants are reincarcerated is that a judge may revoke their bail following conviction for the instant offense. Judges may reason that a pending prison sentence gives the defendants too much incentive to abscond. Unfortunately, this judicial behavior will introduce a bias into models

[31] This problem could be overcome by using an instrumental variable (such as predicted time until disposition) in place of actual time. However, this solution raises the issue of parameter identification. That is, some variables that affect disposition time must not affect failure to appear decisions. These data did not provide suitable identification restrictions.

[32] We have discussed the problem with competing events elsewhere (Rhodes, 1986).

of failure to appear, as defendants who face long sentences may have reduced opportunities to fail to appear for sentencing hearings.

Beyond these concerns, selection bias is a potential problem when analyzing failure to appear. The analytic issues are the same as those raised in developing the survival model. So, too, is the solution. Using equation (7) we estimate the probit model conditional on a release having occurred.

The problems that arise when estimating the probability of failure to appear are serious. We cannot overcome these problems by statistical adjustments. Rather, the solution requires richer data that record opportunities for failing to appear for scheduled court dates. Additional data collection is beyond this study, and we deal the best we can with the data at our disposal.

Statistical and Substantive Significance

In this report we use three tests of statistical significance, two for use within each site, and one for use across sites. First, within each site we use a likelihood ratio test to determine whether arrestees who tested positive for recent drug use have misconduct rates that differ from those of arrestees who tested negative.[33] This test does not establish whether arrestees who test positive have higher or lower rates, only that they have different ones. Second, within each site we use t-scores to test whether those who test positive for cocaine (for heroin, for marijuana, and so on) have higher misconduct rates than those who do test negative.[34] Unlike the first test, this second one is focused on specific drugs, and it

---

[33] The unconstrained model allows all parameters to vary freely in a regression. The constrained model fixes to zero the parameters associated with drug testing, and allows all other parameters to vary freely. The likelihood ratio is the value of the likelihood for the unconstrained model divided by the likelihood for the constrained model. Minus two times the logarithm of this ratio is distributed as Chi-square with degrees of freedom equal to the number of restrictions (i.e., the number of different drug tests included in the model). The null hypothesis is rejected when $p < 0.05$. The test is discussed in G. Judge et al. (1985).

[34] The regression parameter estimates are normally distributed in large samples. (That is, the asymptotic distribution is normal.) Thus, the t-score (the parameter estimate divided by its standard error) is distributed as standard normal under the null hypothesis that drug use does not predict misconduct. The null hypothesis is rejected using a one-tailed test at $p < 0.05$. The test is discussed in Judge et al. (1985).

establishes the direction (higher, lower) of the difference in misconduct rates between those who test positive and those who test negative. Third, we use meta-analysis to combine results from across sites to make a global statement about whether those who test positive for cocaine (for heroin, for marijuana, and so on) engage in pretrial misconduct more frequently than arrestees who test negative.[35] This third test is especially useful because the snippets of information from each individual site can be combined to support a much stronger general conclusion.

Although tests of statistical significance are important in judging whether drug tests help predict pretrial misconduct, they do not tell the entire story. Social scientists tend to use conservative tests. In this context they would stack the deck against concluding that tests for recent drug use are correlated with misconduct.[36] This bias is appropriate for the social scientist, because drawing a wrong conclusion about the relationship between two variables (testing positive for drug use and rearrest) is seen as more costly for the advancement of science than failing to establish that the two are related.

To emphasize this point, consider a simulation. Assume that arrestees who test negative for recent drug use have a $\lambda=e^{6.0}$ and that arrestees who test positive for recent drug use have a $\lambda=e^{5.75}$. These assumptions imply that about 20 percent of those who test negative and about 25 percent of those who test positive will be rearrested within 90 days of release. Assume also that the follow-up period has an exponential distribution with $\lambda=e^{4.87}$ (about 50 percent have their cases terminated within 90 days) and

[35] The individual t-scores are distributed as standard normal. Thus, when the t-scores associated with the parameter estimates for an individual drug (such as cocaine) are summed across sites, and divided by the square root of the number of sites, the result is also distributed as standard normal. The statistic can be weighted to reflect different sample sizes across sites. We call the test a weighted or unweighted Stouffer combined test. This test is described by Wolf (1986).

[36] Hypothesis testing involved two types of errors, called type-1 and type-2. A type-1 error occurs when a null hypothesis (e.g. that drug tests do not predict misconduct) is rejected when in fact it is true (e.g. in fact drug tests do not predict misconduct). A type-2 error occurs when a null hypothesis is not rejected when in fact it is false (eg. in fact drug tests do predict misconduct). Social scientists frequently set the chance of a type-1 error at a small value, usually 0.10, or 0.05, or 0.01. They frequently ignore the possibility of a type-2 error, which can be appreciable.

is truncated at 90 days. This simulation approximates conditions that might be observed if drug test results have substantively meaningful predictive power.

We first simulated data for 200 arrestees who tested negative for recent drug use and for 200 arrestees who tested positive. We then estimated the t-score for the parameter associated with a positive drug test.[37] We repeated this simulation 25 times, and computed the mean and standard deviation for the sample of 25 t-scores. Figure 1 shows the mean and ± one standard deviation for the results.

Next, we simulated data for 400 arrestees who tested negative and 400 who tested positive. We repeated the steps in the previous paragraph (25 repetitions of this sample of 800) and plotted the results in figure 1. Finally, we performed the same simulation for 600 arrestees who tested positive and negative, 800 arrestees who tested positive and negative, up to 2800 arrestees who tested positive and negative. These choices for sample sizes correspond to those samples available for our analysis.

Figure 1 shows that the risk of failing to reject the null hypothesis (drug test results do not predict misconduct) is considerable, especially when sample size is as small as is found in Maricopa County (only 186 cases). The test has extremely low power. Even when the sample sizes are moderate, as in Dade County (1,294 cases), a type-2 error is more likely than not. The probability of a type-2 error remains appreciable even in large samples, such as those from the District of Columbia (5,689 cases) and Manhattan (1,893 cases). Thus, using standard tests of statistical significance to judge drug testing as effective or ineffective poses substantial risk of misguided policies.

Meta-analysis, when applied to statistical analyses from all seven settings, provides a more powerful test of the null hypothesis. Meta-analysis plays an important role in the following discussion.

The Analysis

The following section discusses the analysis of rearrests and failure to appear.

---

[37] The regression model has a constant and one variable, a dummy variable coded 1 for those who test positive and coded 0 for those who test negative.

Figure 1

Simulation of T-Scores for the Drug Positive Parameter

Rearrests. Regression results appear in a series of eight tables (4–11), which follow. The first column of each table identifies the variables that entered the analysis. We intended variables to be comprehensive of factors that researchers often report as being useful when predicting rearrests of defendants on bail. Of course, we included results from drug testing; however, we sometimes dropped test results for substances that were reported infrequently in the jurisdiction. Sometimes we combined several infrequently observed substances into a category "other drug." *To be included in this analysis a defendant had to have had a urine test result.*[38]

We always included criminal history variables. These were the logarithm[39] of the number of previous arrests, the logarithm of the number of recent previous arrests (defined variously), the logarithm of the number of previous incarcerations, the logarithm of the number of previous failures to appear, whether the defendant was on probation or parole, and whether the defendant had a warrant outstanding. (Because the logarithm of 0 is undefined, we added 1 to the independent variables before taking logarithms.) Not all variables were available in every data file. Thus, we sometimes substituted a different variable for one in the list. For example, the number of previous felony convictions was a proxy for the number of previous incarcerations.

---

[38] There seemed to be no meaningful alternative to discarding cases where urine tests were not available. This is troubling: Urine testing is almost always voluntary, and arrestees who have recently used an illicit substance may have an incentive to refuse to take the test. Such refusals appear to be rare. In sites where the missing data rate is high, missing urine testing results appear to arise most frequently for administrative reasons. For example, the site may have gone for periods of time without testing anyone.

[39] A logarithmic transformation was used because it dampens the effect of extreme values of the independent variables. A more detailed analysis of these data would probably find other structural forms superior for predicting. However, searching for alternative structural forms is impractical when multiple data sets are analyzed--search times are too high. Because our interest was focused on the parameter estimates for drug test results, we were less concerned with the structural form for the rest of the model. We recognize that alternative structural forms might lead to somewhat different inferences about the effectiveness of drug tests when predicting misconduct.

Table 4
Regression Results From Dade County: Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)	Parameters (T-scores)
Time at risk	50.763			
	33.979			
Arrested	0.332			
	0.745			
Constant	1.000	**-5.458**	**-6.387**	**-6.387**
		-5.856	**-4.821**	**-4.816**
Cocaine	0.733	**0.431**	**0.613**	**0.613**
	0.443	**1.699**	**1.804**	**1.805**
Marijuana	0.438	**0.760**	**1.052**	**1.052**
	0.496	**2.213**	**2.176**	**2.176**
Both drugs	0.373	**-0.668**	**-0.924**	**-0.924**
	0.484	**-1.784**	**-1.765**	**-1.765**
Condition: pretrial supervision	0.090	0.174	0.236	0.235
	0.286	0.673	0.661	0.659
Condition: TASC	0.153	0.184	0.264	0.264
	0.360	0.931	0.945	0.947
Prior felony conviction	0.266	**0.312**	**0.435**	0.435
	0.442	**1.800**	**1.721**	1.533
Ln(Number prior arrests)	1.224	**0.444**	**0.623**	**0.623**
	1.012	**5.143**	**4.273**	**4.264**
On probation or parole	0.050	-0.138	-0.238	-0.239
	0.217	-0.457	-0.548	-0.496
Outstanding warrant	0.191	**0.428**	**0.590**	**0.589**
	0.393	**2.763**	**2.443**	**2.414**
Offense Seriousness	3.499	0.002	0.012	0.012
	1.991	0.027	0.090	0.086
Offense severity unknown	0.218	0.449	0.361	0.362
	0.413	1.005	0.569	0.519
Ln(Length of Residence)	5.236	**-0.108**	**-0.142**	**-0.141**
	1.625	**-1.832**	**-1.708**	**-1.702**
Has phone	2.886	**0.082**	**0.107**	**0.107**
	3.799	**3.080**	**2.793**	**2.745**
Is married	0.137	0.322	0.437	0.436
	0.344	1.517	1.431	1.428
Is employed	0.726	0.070	0.117	0.117
	0.446	0.408	0.488	0.473
Age/10	2.934	-0.648	-0.749	-0.749
	0.863	-1.349	-1.157	-1.157
((Age/10)^2)/10	0.935	0.649	0.698	0.698
	0.626	0.930	0.755	0.756
Sigma			1.598	1.597
			3.888	3.860
Adjusted-rho				0.000
				0.000
Number of cases	1,294	1,294	1,294	1,294
Likelihood ratio test		0.10	0.07	0.07

Table 5
Regression Results From District of Columbia (Adults, 1984): Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
Time at risk	96.197		
	51.325		
Arrested	0.180		
	0.384		
Constant	1.000	-4.384	-4.401
		-13.746	-13.313
Cocaine	0.182	-0.046	-0.045
	0.386	-0.406	-0.382
Heroin	0.199	0.170	0.173
	0.399	1.590	1.577
PCP	0.324	0.039	0.040
	0.286	0.437	0.438
Metadone	0.029	0.149	0.151
	0.168	0.802	0.791
Poly-drug	0.196	0.152	0.156
	0.397	1.117	1.112
Condition #1	0.094	0.106	0.105
	0.291	0.930	0.901
Condition #2	0.163	0.143	0.146
	0.370	1.592	1.583
Condition #3	0.056	0.256	0.259
	0.230	1.953	1.923
Ln(Prior convictions)	0.629	0.512	0.523
	0.699	9.760	9.079
Open case pending	0.247	0.161	0.162
	0.550	2.940	2.871
Number probation terms	0.138	-0.100	-0.104
	0.403	-1.312	-1.330
Number parole terms	0.109	-0.313	-0.317
	0.358	-3.196	-3.162
Offense seriousness	5.148	-0.030	-0.030
	4.261	-3.558	-3.526
Felony	0.352	-0.231	-0.236
	0.478	-3.420	-3.384
Ln(Time at address)	0.015	0.136	0.136
	0.122	0.636	0.619
Employment	0.152	-0.051	-0.050
	0.359	-0.527	-0.505
In school	0.265	-0.271	-0.276
	0.441	-3.347	-3.322
Married	0.081	-0.129	-0.135
	0.273	-1.070	-1.090
High school graduate	0.506	0.039	0.040
	0.500	0.602	0.607
Age/10	2.799	-1.108	-1.134
	0.831	-5.798	-5.712
(Age/10)^2/10	8.527	0.114	0.117
	5.869	4.533	4.501
Sigma			0.407
			1.061
Number of cases	5,689	5,689	5,689
Chi-square		0.002	0.002

Table 6
Regression Results From District of Columbia (Juveniles): Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
Constant	1.000	-1.282	-0.788
		3.994	-1.622
Cocaine	0.137	-0.224	-0.371
		1.136	-1.460
Opiates	0.006	0.187	0.139
		0.408	0.237
Marijuana	0.130	-0.006	0.007
		0.044	0.037
PCP	0.243	0.061	0.075
		0.567	0.533
Poly-drug		0.143	0.193
		0.598	0.623
Experimental program		0.001	0.005
		0.125	0.436
Number of prior arrests	1.007	0.098	0.160
		5.620	5.705
Age	14.744	0.047	0.011
		2.288	0.337
Sigma			1.345
			13.447
Number of cases	2.137	2,137	2,137

Table 7

Regression Results From Maricopa County: Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
Time at risk	77.177		
	41.184		
Arrested	0.118		
	0.324		
Constant	1.000	-6.471	
		-1.744	
Cocaine	0.360	0.603	0.538
	0.481	1.287	1.134
Amphetamine	0.151	-0.198	-0.346
	0.359	-0.355	-0.617
Other drug	0.070	-0.042	-0.025
	0.256	-0.059	-0.034
Age/10	2.898	0.560	0.478
	0.885	0.218	0.184
((Age/10)^2)/10	0.918	-2.006	-2.096
	0.609	-0.459	-0.474
Ln(Number of charges)	0.309	0.596	0.628
	0.425	1.335	1.413
Prior FTAs	0.118	0.782	0.916
	0.324	1.429	1.629
Time at present address	3.430	-0.162	-0.123
	1.085	-0.812	-0.600
Number of cases	186	186	186
Likelihood ratio test		0.58	

Table 8
Regression Results From Milwaukee County: Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)	Parameters (T-scores)
Time at risk	76.433			
	26.602			
Arrested	0.157			
	0.364			
Constant	1.000	**-7.081**	**-7.203**	**-7.124**
		-5.958	**-5.535**	**-5.462**
Cocaine	0.568	-0.121	-0.149	-0.121
	0.496	-0.180	-0.208	-0.168
Opiate	0.036	0.336	0.372	0.400
	0.187	0.648	0.669	0.717
Amphetamine	0.013	-0.608	-0.616	-0.584
	0.114	-0.564	-0.550	-0.522
Benzodiazepines	0.063	-0.653	-0.704	-0.722
	0.243	-1.468	-1.450	-1.486
Poly-drug	0.600	0.497	0.540	0.497
	0.490	0.707	0.716	0.656
Experimental groups: special supervison	0.221	-0.073	-0.102	-0.156
	0.415	-0.342	-0.438	-0.657
Ln(Times incarcerated last 5 years)	0.177	**0.368**	**0.385**	**0.393**
	0.398	**1.817**	**1.691**	**1.719**
Ln(Number of prior arrests)	0.937	**0.446**	**0.487**	**0.471**
	0.711	**3.076**	**2.705**	**2.710**
On probation or parole	1.923	0.033	0.008	-0.021
	0.271	0.108	0.024	-0.061
Ln(Number of felony FTAs)	0.037	0.614	0.613	0.569
	0.173	1.509	1.387	1.266
Ln(Number of misdemeanor FTAs)	0.441	0.152	0.147	0.163
	0.634	1.105	0.985	1.082
Outstanding warrant	0.212	**-0.450**	**-0.489**	**-0.514**
	0.409	**-1.740**	**-1.707**	**-1.799**
Offense seriousness	0.471	-0.167	-0.177	-0.146
	0.309	-0.537	-0.534	-0.435
Age/10	2.866	0.201	0.176	0.241
	0.806	0.302	0.247	0.335
((Age/10)^2)/10	0.886	-0.531	-0.520	-0.606
	0.568	-0.527	-0.483	-0.556
Sigma			0.727	0.820
			0.926	1.388
Adjusted-rho				-1.357
				-0.935
Number of cases	830	830	830	829
Likelihood ratio test		0.27	0.28	0.29

Table 9
Regression Results From Manhattan: Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
Time at risk	132.056		
	133.730		
Arrested	0.239		
	0.426		
Constant	1.000	-6.092	-6.342
		-8.767	-7.368
Cocaine	0.448	0.028	0.037
	0.497	0.227	0.232
Opiates	0.269	0.277	0.339
	0.444	2.011	1.873
PCP	0.133	0.106	0.125
	0.339	0.649	0.580
Methadone	0.160	-0.316	-0.372
	0.367	-1.385	-1.255
Poly-drug	0.229	0.237	0.299
	0.420	1.170	1.115
Ln(Previous convictions)	0.379	-0.123	-0.157
	0.768	-1.494	-1.474
Ln(Previous arrests)	1.263	0.304	0.346
	1.192	5.218	4.517
Previous warrants	0.500	0.202	0.271
	0.500	1.484	1.531
Outstanding criminal case	1.038	0.046	0.045
	1.498	1.326	1.041
Number of previous probations	0.200	0.020	-0.039
	0.528	0.212	-0.322
Number of previous parole	0.065	0.229	0.193
	0.284	1.157	0.747
Number of previous revocations	0.077	0.430	0.549
	0.274	2.569	2.431
Offense seriousness	5.774	-0.002	-0.006
	4.957	-0.189	-0.478
Felony charge	0.883	-0.134	-0.210
	0.322	-0.969	-1.132
Ln(Time at address)	0.975	0.573	0.731
	0.157	1.485	1.554
Employed	0.349	-0.315	-0.375
	0.477	-2.879	-2.664
In school	0.058	-0.140	-0.274
	0.234	-0.670	-1.004
Married	0.273	0.136	0.116
	0.446	1.172	0.770
Highschool graduate	0.425	0.055	0.060
	0.494	0.537	0.450
Age/10	2.793	-0.631	-0.716
	0.918	-1.767	-1.605
((Age/10)^2)/10	8.642	0.054	0.056
	6.207	0.989	0.836
Standard error			1.244
			10.042
Number of cases	1,893	1,893	1,893
Chi-square		0.02	0.04

Table 10
Regression Results From Prince George's County: Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)	Parameters (T-scores)
Time at risk	96.68			
	38.88			
Arrested	0.12			
	0.36			
Constant	1.00	**-6.47**	**-6.61**	**-6.77**
		-5.19	**-4.96**	**-4.95**
Cocaine	0.65	0.39	0.43	0.41
	0.48	1.23	1.25	1.21
Opiates	0.13	**0.56**	**0.59**	**0.66**
	0.33	**2.07**	**2.01**	**2.19**
Marijuana	0.30	-0.01	0.00	0.02
	0.46	-0.04	0.00	0.09
PCP	0.23	-0.09	-0.09	-0.11
	0.42	-0.36	-0.33	-0.40
Poly-drug	0.80	-0.07	-0.10	-0.17
	0.40	-0.16	-0.22	-0.37
Ln(Number of incarcerations last 5 years)	0.81	-0.16	-0.15	-0.12
	0.68	-0.65	-0.62	-0.47
Ln(Number of prior arrests)	0.72	**0.77**	**0.80**	**0.74**
	0.80	**2.98**	**2.87**	**2.68**
Ln(Number of recent prior arrests)	0.51	-0.09	-0.10	-0.13
	0.61	-0.35	-0.37	-0.48
Ln(Number of outstanding warrants)	0.14	0.18	0.21	0.27
	0.35	0.70	0.75	0.93
Ln(Prior FTA)	0.28	0.14	0.13	0.15
	0.45	0.53	0.47	0.53
Condition: experimental group	0.28	-0.05	-0.04	0.02
	0.45	-0.20	-0.15	0.06
Ln(Length of residence in county)	3.38	0.11	0.12	0.12
	1.35	1.44	1.47	1.48
Married	0.11	**0.45**	**0.49**	**0.53**
	0.32	**1.67**	**1.65**	**1.73**
Has phone	0.82	0.22	0.22	0.23
	0.39	0.84	0.81	0.82
Age/10	2.87	-1.13	-1.21	-1.19
	0.74	-1.62	-1.58	-1.54
((Age/10)^2)/10	0.88	1.61	1.74	1.74
	0.49	1.63	1.58	1.56
Adjustment for release date	0.38	0.15	0.14	0.06
	0.48	0.64	0.56	0.22
Sigma			0.67	0.82
			0.94	1.39
Adjusted-rho				2.32
				1.05
Number of cases	1072	1072	1072	1055
Likelihood ratio test	0.54	0.15	0.15	0.15

Table 11
Regression Results From District of Columbia (Adults, 1989 - 1990): Rearrests

District of Columbia Adults Regression Results: Rearrests

Variables	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
Time at risk	129.319		
	140.963		
Arrested	0.219		
	0.414		
Constant	1.000	-4.552	-4.749
		-6.857	-4.870
Cocaine	0.590	0.333	0.452
	0.492	2.357	2.191
Heroin	0.170	0.057	0.054
	0.376	0.249	0.156
Marijuana	0.098	0.015	-0.103
	0.298	0.072	-0.331
PCP	0.103	0.201	0.144
	0.304	0.834	0.405
Other	0.069	-0.411	-0.505
	0.253	-1.421	-1.258
Poly-drug	0.278	-0.029	0.060
	0.448	-0.113	0.158
Condition: drug testing	0.346	0.440	0.685
	0.476	3.410	3.559
Condition: treatment referral	0.141	-0.055	-0.148
	0.348	-0.321	-0.565
Ln(Number of prior arrests)	1.014	0.730	0.907
	0.930	5.740	4.768
Ln(Number of prior convictions)	0.642	-0.553	-0.761
	0.747	-3.320	-3.118
On probation	0.130	0.369	0.301
	0.337	2.181	1.186
On parole	0.057	0.685	0.894
	0.231	3.117	2.700
Misdemeanor/Felony	0.631	-0.665	-0.762
	0.483	-5.788	-4.448
Married	0.101	0.155	0.180
	0.302	0.783	0.634
In school	0.047	-0.545	-0.554
	0.211	-1.854	-1.328
Employed	0.322	-0.254	-0.358
	0.467	-2.026	-1.947
Age/10	2.876	-1.092	-1.288
	0.838	-2.478	-2.013
((Age.10)^2/10)	0.897	1.045	1.105
	0.571	1.577	1.159
Sigma			1.689
			11.745
Number of cases	1,538	1,538	1,538
Likelihood ratio test		0.06	0.14

We always included community ties variables. These were the logarithm of the time that the defendant lived in the community, the availability of a telephone, whether he or she was married, and whether he or she was employed. Not every variable was reported in each data set. Furthermore, when the rate of missing data was high for any one of these variables, we used our judgment to drop that variable from the analysis.

Other variables included a measure of offense seriousness and the defendant's age. We measured offense seriousness in various ways, depending on information provided in the data. We measured age as age and age-squared. (To facilitate the computing algorithm we divided age by 10 and we divided age-squared by 1000. Neither adjustment affects any parameter estimates, except by alternating the scale.) We excluded race and gender from the analysis, reasoning that judges could not legitimately use these variables to predict pretrial misconduct.[40]

Upon release, defendants are subjected to various special conditions imposed upon them by the court. Sometimes these conditions include drug testing; often they include reporting to a third party periodically. We included the imposition of special supervision conditions as control variables in the regression specifications. Beyond the special experimental programs introduced in all sites[41] except Dade County and Manhattan, we are uncertain that these control variables are comprehensive of all special supervision imposed on defendants.[42]

---

[40] Inclusion of age is arguable. We reasoned that judges might legitimately consider a criminal record in light of a defendant's age, discounting lengthy records somewhat for defendants who are old. Excluding race and gender from the regressions has a disadvantage. If these omitted variables are correlated with urine test results, when all other variables are held constant, then the parameter associated with a urine test will be biased.

[41] Supervision programs were experimental in Washington, D.C. when the 1984 data were collected. Those programs had been institutionalized by the time that the 1989–1990 data were collected.

[42] It was difficult to determine when a defendant had been selected for the various experimental programs in the District of Columbia (adults). Random group assignment by the on-site researchers was clear. Judicial placements seemed clear. However, judges often made placement in violation of project's random assignment.

The tables are organized so that parameter estimates and t-scores associated with the drug test variable appear near the top of each table. These statistics are surrounded by a solid border. If the data contained information about pretrial urine testing supervision, parameters associated with that supervision are reported next. These are followed by parameters associated with criminal history, offense seriousness, community ties, age, and other variables.

We used no empirical search procedures to identify a parsimonious model. (Many of the previous analyses of these data used step-wise or other search procedures.) Although search procedures may be justifiable for site-specific analysis, cross-site comparisons are most accurate when conducted on models that have the full model specification.[43]

We used no test-retest procedures. Although researchers commonly use test-retest procedures when developing prediction instruments, our focus is more narrow. We do not develop prediction instruments; rather, we simply ask whether drug test results can improve predictions. We judge whether drug test results could improve predictions based on the statistical significance of the parameters

---

[43] As a simplification, we specify λ as $\lambda=\exp(\beta_0+X_1\beta_1+X_2\beta_2+X_3\beta_3)$ where X_1 is the first control variable, X_2 is the second control variable, and X_3 is the drug test result. We allow for some correlation between X_1 and both X_2 and X_3. If X_1 were dropped from the model's specification, then the estimates of β_2 and β_3 will increase or decrease depending on whether the correlation between X_1 and either X_2 and X_3 is positive or negative. The test statistic (t-score) associated with β_3 will also be affected. We are especially concerned that β_3 and its t-score not be affected this way, so we have declined to drop variables from our chosen model specification.

If X_1 and X_2 are highly correlated, the parameter estimates for β_1 and β_2 may appear large, and β_1 and β_2 may not be statistically significant, even when they are substantively large. This correlation between X_1 and X_2 in no way affects the estimate of β_3, however. Thus, dropping X_1 from the model's specification can bias the estimate of β_3, while the inclusion of X_1 has no cost. This conclusion argues in favor of using a full model specification.

One might argue that inclusion of X_1 inflates the standard error of β_3 when X_1 and X_3 are highly correlated. This objection misses the point of the analysis, however. If X_1 and X_3 are highly correlated, then X_1 is already a suitable predictor of being arrested. Results from the drug test result, reflected in X_3, simply adds no additional useful information.

associated with a positive drug test. Test-retest procedures have no advantages, and significant disadvantages, when used for this purpose.[44]

Nevertheless, readers may look for a summary measure of how well we have been able to distinguish between releasees who are rearrested or fail to appear and those who are not arrested or do not fail to appear. We use predictions, based on the regression equations in each setting, to assign all arrestees to one of five equal-sized categories based on the predicted probability that they will be rearrested or fail to appear. Then, in figures 4 through 10, we report the rearrest rates and the failure to appear rates for individuals in each of these five categories.[45]

The summary measure does not show by how much drug test results improve our ability to predict pretrial misconduct. Therefore, we also provide estimates of the differences between the probability of rearrest (failure to appear) for two hypothetical arrestees: one tested negative for recent drug use and the other tested positive. We evaluated these probabilities at the means for all variables, other than drug test results, that appeared in the X vector, and report them in figures 2 and 3.

As mentioned, the first column of tables four through eleven identifies the variables that entered the analysis. The second column provides descriptive statistics for those variables. The standard errors appear below their respective means.

The third column reports regression results for the statistical model that has neither unmeasured heterogeneity nor selection bias. Asymptotic t-scores appear below the parameter estimates. A reader might consider a t-score of greater than 1.64 to be statistically significant for the drug test results. (This

---

[44] Test-retest procedures are often used when developing prediction equations because predictions based on a calibration sample will "shrink" when applied to a validation sample. Shrinkage means that predictions based on the calibration sample will tend to be less accurate in the replication sample.

However, the concept of shrinkage does not apply to the parameter estimates. In large samples these are distributed normally about the true population parameter. (This presumes that the model is specified correctly.) They would differ between a calibration and validation sample, of course, but they would not be systematically larger in one than in another. On the other hand, tests of statistical significance would be less precise in a split-sample due to decreased sampie size. Thus, test-retest procedures have no advantages and significant disadvantages when used to test the hypotheses posed here.

[45] Our ability to predict outcomes is overstated in these figures. See the previous note.

t-score has a type-1 error of 0.05 for a one-tailed test of significance.) We have written parameters that have t-scores of 1.64 or more in bold. In some instances the sample is small, and the t-scores are correspondingly suspect.[46] More weight should probably be placed on analyses conducted with large data sets rather than on analyses with small ones.

The second column also presents a likelihood ratio test statistic. The full model used all the variables listed in column one, including the drug test results. The restricted model included all the variables listed in column one, excluding the drug test results. The tables report the level of significance for the resulting Chi-square statistic. For example, the number 0.17 indicates that we would have considered the likelihood ratio statistic as statistically significant at $p<0.17$.

We estimated the basic exponential survival model for all the sites. As noted earlier, an exponential survival model did not seem to fit the data from Maricopa County. Thus, we also estimated a Cox proportional hazard model for that site. Results appear in column three for Maricopa County only. We describe column three for other sites next.

Except in Table 7, column three reports results from estimating the exponential survivor model with unmeasured heterogeneity. We estimate one additional parameter, σ, the standard deviation of the distribution of ϵ.

Column four reports results for the exponential model with unmeasured heterogeneity as adjusted for selection bias. As was indicated earlier, we could not always estimate this model. When estimates were impossible, the column is missing. Otherwise, the column has one additional parameter, called the adjusted rho. This can be used to derive ρ, the correlation between ϵ and ϵ_2, by using the formula:

$$\rho = \frac{2}{1+e^{adjusted\ rho}} - 1$$

[46] Small sample properties for the standard errors of the parameter estimates are unknown. This is to say that the t-scores are accurate only for large samples. The definition of large is not precise, but samples of a few hundred would not seem to qualify.

Although this specification of ρ is cumbersome, it facilitates the computing algorithm by constraining ρ to be between -1 and +1.

Failure to appear. Variable selection and model specification were the same for failure to appear. Tables twelve through eighteen report results.

Column 1 identifies variables. Column 3 reports results from the basic probit model. Column 4 reports results after adjusting for selection bias. As before, we could not always estimate the selection bias model, and when we could not, the column is blank.

Discussion: Predicting Misconduct Using Urine Test Results

Table 19 summarizes statistical tests of whether drug test results for individual drugs help predict rearrest once other factors are taken into account. We have boxed parameter estimates that are statistically significant.[47] A positive parameter estimate indicates that a defendant who tested positive is more likely than a defendant who tested negative to be rearrested. Beyond this, interpretation of the parameter estimates requires the use of equation (1) through equation (6).

Tests for recent use of cocaine were conducted in all eight settings. This is understandable. Cocaine[48] has been the most prevalent drug of abuse, at least among arrestees, since the middle of the 1980s. It remains the most prevalent drug of abuse today. Authorities are understandably concerned about the relationship between cocaine use and criminal activity.

Unfortunately, no clear picture of the predictive power of a positive test for cocaine emerges from this analysis. The strongest evidence that a positive test predicts rearrest comes from adults in Washington, D.C., during 1989 and 1990. Table 19 shows this propensity with a box around the results

---

[47] In all tests reported in this paper the critical value for the test statistic has been set to 0.05=p using a one-tail test of significance.

[48] Drug tests cannot determine the mode of administration. Some cocaine is used as powder, some is injected (often with heroin), and most is smoked as crack. We use the term "cocaine" generally.

Table 12
Regression Results From Dade County: Failure to Appear

Variable	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
FTA	0.101		
	0.302		
Constant	1.000	**-1.802**	**-1.798**
		-2.880	**-3.084**
Cocaine	0.733	0.140	0.120
	0.443	0.930	0.854
Marijuana	0.438	-0.316	-0.319
	0.496	-1.150	-1.206
Both drugs	0.373	0.297	0.318
	0.484	1.000	1.116
Condition: pretrial supervision	0.090	0.227	0.230
	0.286	1.290	1.390
Condition: TASC	0.153	**0.554**	**0.633**
	0.360	**4.350**	**5.115**
Prior felony conviction	0.266	-0.138	-0.217
	0.442	-1.000	-1.605
Ln(Number prior arrests)	1.224	-0.018	-0.031
	1.012	-0.290	-0.492
On probation or parole	0.050	-0.293	**-0.379**
	0.217	-1.070	**-1.658**
Outstanding warrant	0.191	**0.333**	**0.289**
	0.393	**2.710**	**2.545**
Offense Seriousness	3.499	-0.051	-0.069
	1.991	-0.810	-1.136
Offense severity unknown	0.218	**-1.054**	**-1.146**
	0.413	**-3.200**	**-3.621**
Ln(Length of Residence)	5.236	0.050	0.051
	1.625	1.440	1.578
Has phone	0.403	-0.142	-0.141
	0.491	-1.290	-1.397
Is married	0.137	-0.081	-0.074
	0.344	-0.510	-0.491
Is employed	0.726	-0.064	-0.015
	0.446	-0.550	-0.145
Age/10	2.934	0.308	0.283
	0.863	0.970	0.963
((Age/10)^2)/10	0.935	-0.435	-0.410
	0.626	-0.970	-0.993
Adjusted-rho	0.874		3.357
	0.474		5.982
Number of cases	1,294	1,294	1,294
Chi-square		0.27	0.26

Table 13
Regression Results From District of Columbia (Adults, 1984): Failure to Appear

Variable	Means (Standard errors)	Parameters (T-scores)
Failure to appear	0.176	
	0.381	
Constant	1.000	-0.203
		-1.000
Cocaine	0.182	0.107
	0.386	1.500
Heroin	0.199	0.081
	0.399	1.170
PCP	0.324	-0.142
	0.286	-2.490
Methadone	0.029	-0.207
	0.168	-1.480
Poly-drug	0.196	-0.052
	0.397	-0.580
Condition #1	0.094	0.045
	0.291	0.610
Condition #2	0.163	-0.027
	0.370	-0.440
Condition #3	0.056	-0.080
	0.230	-0.860
Ln(Prior convictions)	0.629	0.059
	0.699	1.760
Open case pending	0.247	-0.158
	0.550	-3.790
Number probation terms	0.138	-0.052
	0.403	-1.000
Number parole terms	0.100	-0.159
	0.358	-2.490
Offense seriousness	5.148	-0.018
	4.261	-3.730
Felony	0.352	-0.134
	0.478	-3.100
Ln(Time at address)	0.015	-0.027
	0.122	-0.170
Employment	0.152	-0.163
	0.359	-2.710
In school	0.265	-0.156
	0.441	-3.290
Married	0.081	-0.282
	0.273	-3.510
High school graduate	0.506	0.065
	0.500	1.600
Age/10	2.799	-0.272
	0.831	-2.240
(Age/10)^2/10	8.527	0.031
	5.869	1.920
Number of cases	5,689	5,689
Chi-square		0.000

Table 14
Regression Results From Maricopa County: Failure to Appear

Variables	Means (Standard errors)	Parameters (T-scores)
Failure to appear	0.257	
	0.438	
Constant	1.000	-0.705
		-0.560
Cocaine	0.356	**0.424**
	0.480	**2.010**
Amphetamine	0.158	**0.455**
	0.366	**1.780**
Other drug	0.064	-0.014
	0.246	-0.030
Ln(Number of charges)	0.303	-0.040
	0.427	-0.170
Prior FTAs	0.124	-0.044
	0.330	-0.150
Time at present address	3.376	-0.008
	1.091	-0.080
Age/10	2.899	0.121
	0.931	0.150
((Age/10)^2)/10	0.927	-0.574
	0.677	-0.450
Number of cases	202	202
Chi-square		0.02

Table 15
Regression Results From Milwaukee County: Failure to Appear

	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
Failure to appear	0.180		
	0.384		
Constant	1.000	**-2.213**	-2.188
		-3.060	
Cocaine	0.567	0.523	0.526
	0.496	1.270	
Opiate	0.036	-0.007	-0.003
	0.187	-0.020	
Amphetamine	0.013	-0.136	-0.136
	0.115	-0.270	
Benzodiazepines	0.063	0.320	0.319
	0.243	1.380	
Poly-drug	0.600	-0.310	-0.318
	0.490	-0.720	
Experimental groups, special supervison	0.220	**0.375**	0.367
	0.414	**2.980**	
Ln(Times incarcerated last 5 years)	0.178	0.140	0.140
	0.398	0.890	
Ln(Number of prior arrests)	0.938	**-0.204**	-0.206
	0.712	**-2.120**	
On probation or parole	1.923	**0.569**	0.566
	0.272	**2.260**	
Ln(Numbef of felony FTAs)	0.037	**0.852**	0.845
	0.173	**3.120**	
Ln(Number of misdemeanor FTAs)	0.440	**0.520**	0.521
	0.634	**5.900**	
Outstanding warrant	0.211	-0.115	-0.118
	0.408	-0.840	
Offense seriousness	0.472	-0.177	-0.170
	0.309	-0.940	
Age/10	2.866	-0.032	-0.027
	0.806	-0.100	
((Age/10)^2)/10	0.886	0.008	0.002
	0.568	0.020	
Adjusted-rho	1.193		-0.157
	0.749		
Number of cases	829	829	829
Chi-square		0.26	0.26

Note: Standard errors could not be estimated for the selection bias model.

Table 16
Regression Results From Manhattan: Failure to Appear

Variable	Means (Standard errors)	Parameters (T-scores)
Failure to appear	0.356	
	0.479	
Constant	1.000	-0.299
		-0.800
Cocaine	0.448	0.116
	0.497	1.550
Opiates	0.269	-0.004
	0.444	-0.050
PCP	0.133	0.090
	0.339	0.800
Methadone	0.160	0.087
	0.367	0.580
Poly-drug	0.229	-0.115
	0.420	-0.830
Ln(Previous convictions)	0.379	-0.184
	0.768	-3.570
Ln(Previous arrests)	1.263	-0.029
	1.192	-0.760
Previous warrants	0.500	0.001
	0.500	0.010
Outstanding criminal case	1.038	0.008
	1.498	0.340
Number of previous probations	0.200	-0.078
	0.528	-1.180
Number of previous parole	0.065	-0.184
	0.284	-1.310
Number of previous revocations	0.077	-0.027
	0.274	-0.220
Offense seriousness	5.774	-0.018
	4.957	-2.860
Felony charge	0.883	-0.190
	0.322	-2.050
Ln(Time at address)	0.975	0.163
	0.157	0.830
Employed	0.349	-0.165
	0.477	-2.460
In school	0.058	-0.255
	0.234	-1.830
Married	0.273	-0.098
	0.446	-1.360
Highschool graduate	0.425	0.193
	0.494	3.030
Age/10	2.793	0.105
	0.918	0.540
((Age/10)^2)/10	8.642	-0.020
	6.207	-0.700
Number of cases	1,893	1,893
Chi-square		0.72

Table 17
Regression Results From Prince George's County: Failure to Appear

Variable	Means (Standard errors)	Parameters (T-scores)	Parameters (T-scores)
Failure to appear	0.213		
	0.410		
Constant	1.000	**-1.318**	**-1.093**
		-2.100	**-5.596**
Cocaine	0.658	**0.353**	**0.335**
	0.475	**2.360**	**2.347**
Opiates	0.124	0.037	0.026
	0.330	0.250	0.184
Marijuana	0.301	-0.010	-0.019
	0.459	-0.090	-0.176
PCP	0.235	-0.117	-0.099
	0.424	-0.980	-0.853
Poly-drug	0.798	0.000	0.035
	0.402	0.000	0.189
Condition: experimental group	0.276	**0.353**	**0.287**
	0.447	**3.540**	**2.690**
Ln(Number of incarcerations last 5 years)	0.811	0.154	0.130
	0.683	1.370	1.319
Ln(Number of prior arrests)	0.720	0.035	0.071
	0.798	0.260	0.511
Ln(Number of recent prior arrests)	0.513	0.030	0.061
	0.614	0.210	0.413
Ln(Number of outstanding warrants)	0.144	**0.543**	**0.476**
	0.351	**4.340**	**3.555**
Ln(Prior FTA)	0.278	0.001	0.009
	0.448	0.010	0.067
Ln(Length of residence in county)	3.380	-0.047	-0.046
	1.344	-1.390	-1.423
Married	0.114	0.148	0.127
	0.318	1.040	0.933
Has phone	0.821	0.018	-0.037
	0.384	0.150	-0.288
Age/10	2.870	0.066	0.024
	0.740	0.170	0.417
((Age/10)^2)/10	0.878	-0.120	-0.081
	0.493	-0.210	-0.636
Adjustment for release date	0.082	**-0.433**	**-0.389**
	0.274	**-2.270**	**-2.158**
Adjusted-rho			-0.906
			-1.598
Number of cases	1,055	1,055	1,055
Chi-square		0.02	0.02

Table 18

Regression Results From District of Columbia (Adults, 1989 - 1990): Failure to Appear

Variables	means (Standard errors)	parameters (T-scores)
Failure to appear	0.193	
	0.394	
Constant	1.000	-0.446
		-1.050
Cocaine	0.590	**0.285**
	0.492	**3.030**
Heroin	0.170	0.151
	0.376	0.970
Marijuana	0.098	0.024
	0.298	0.170
PCP	0.103	-0.062
	0.304	-0.370
Other	0.069	0.123
	0.253	0.760
Poly-drug	0.278	-0.098
	0.448	-0.570
Condition: drug testing	0.346	**0.390**
	0.476	**4.380**
Condition: treatment referral	0.141	-0.024
	0.348	-0.200
Ln(number of prior arrests)	1.014	0.066
	0.930	0.730
Ln(number of prior convictions)	0.642	-0.186
	0.747	-1.620
On probation	0.130	0.039
	0.337	0.320
On parole	0.057	0.161
	0.231	0.920
Misdemeanor/Felony	0.631	-0.113
	0.483	-1.440
Married	0.101	0.113
	0.302	0.900
In school	0.047	**-0.462**
	0.211	**-2.280**
Employed	0.322	**-0.188**
	0.467	**-2.230**
Age/10	2.876	-0.222
	0.838	-0.810
((Age.10)^2/10)	0.897	0.105
	0.571	0.260
Number of cases	1,538	1,538
Likelihood ratio test		0.000

Table 19

Summary of Regression Results for Rearrests

Drug Type	Dade County	D.C. Adult, 1984	D.C Juveniles	Manhattan	Maricopa County	Milwaukee County	Prince George's County	D.C. Adults, 1989-1990
Marijuana	1.052		0.007				0.023	-0.103
	2.176		0.037				0.088	-0.331
Cocaine	0.613	-0.045	-0.371	0.037	0.538	-0.121	0.415	0.452
	1.805	-0.382	-1.460	0.232	1.134	-0.168	1.207	2.191
Opiates*		0.173	0.139	0.339		0.400	0.664	0.054
		1.577	0.237	1.873		0.717	2.192	0.156
Amphetamines					-0.346	-0.584		
					-0.617	-0.522		
Benzodiazapines						-0.722		
						-1.486		
PCP		0.040	0.075	0.125			-0.107	0.144
		0.438	0.533	0.580			-0.401	0.405
Other drugs					-0.025			-0.505
					-0.034			-1.258
Methadone		0.151		-0.372				
		0.791		-1.255				
Poly-drug	-0.924	0.156	0.193	0.299		0.497	-0.167	0.060
	-1.765	1.112	0.623	1.115		0.656	-0.371	0.158
Number of cases	1,294	5,689	2,137	1,893	186	829	1,072	1,538

* Denotes significant at $p < 0.05$ in a weighted Stouffer test.

pertaining to cocaine in D.C. However, a positive test for cocaine did not help to predict a rearrest among adults in D.C. during 1984. Nor did a positive test help to predict rearrest among D.C. juveniles.

Consistent with results for adults in Washington during 1989–1990, those who tested positive for cocaine in Dade County are more likely than who tested negative to be rearrested. Although the effect of a positive drug test is statistically significant, it is not conclusive. First, the effect is not strong,[49] and second, other ways of examining the Dade County data have lead to different conclusions (Goldkamp, Gottfredson, and Weiland 1990a).

Moreover, a positive test for recent cocaine use was not statistically significant in any of the other four sites. In two sites (Prince George's County and Maricopa County) the effect was positive. (We heavily discount the Maricopa County results.[50]) In Manhattan the parameters show virtually no difference between those who tested positive for cocaine and those who tested negative. In Milwaukee, those who tested positive are arrested less frequently than those who tested negative, although the results did not approach statistical significance.

On balance we find no consistent evidence supporting the assertion that detecting recent cocaine use helps predict rearrests once other factors have been taken into account. Nevertheless, we cannot altogether discount the evidence from Dade County, Washington, D.C. (1989-1990), and Prince George's County.[51] A cautious conclusion is that the use of a positive test for cocaine use has not been established.

---

[49] The t-score is 1.805, which is just statistically significant at $p<0.05$.

[50] Maximum likelihood procedures were used to estimate parameters. The standard errors, used in the t-scores, have only an asymptotic justification. They may be inaccurate for small samples. We deem this a problem in Maricopa County where the sample size was only 186 cases.

[51] The effect was not statistically significant using either a weighted or an unweighted Stouffer combined test. See note 33. The unweighted Stouffer test had a score of 1.61, which approaches statistical significance. Given our concern with type two errors, we might accept this as evidence that a positive test for recent use of cocaine helps predict rearrest. However, a weighted Stouffer test had a score of only 0.31, a value that does not approach statistical significance. The differences between these two tests is attributable to the large sample sizes for Washington, D.C. adults (1984) and juveniles. Findings from these two settings dominate the weighted Stouffer test but play a lesser role in the unweighted Stouffer test.

Heroin use may be different. Heroin use was prevalent among arrestees in four settings: Manhattan, the District of Columbia, (adults, 1984, and adults, 1989-1990), and Prince George's County. The test of the utility of drug testing was statistically significant for heroin in two of the four settings, it approached statistical significance in the third, and was positive but insignificant in the fourth. Although the effect was not statistically significant in Milwaukee and the District of Columbia (juveniles), heroin use was relatively rare in these two settings. Indeed, in the District of Columbia, the rate for positive tests for opiates (0.5 percent of juvenile arrestees) approached the false positive rate for the test.[52] Altogether, then, we cannot ignore what appear to be statistically and substantively strong relationships between drug test results for heroin and pretrial rearrest rates.[53]

Furthermore, the difference between rearrest rates for those who tested positive and negative for heroin might be considered substantively large. Figure 2 summarizes predictions of the probability of a rearrest within 90 days for those who tested negative on all drugs and those who tested positive for opiates only. The regressions were estimated at the mean for all variables other than drug test results.

The bulk of the evidence seems to indicate that positive tests for other drugs have scant utility when predicting pretrial arrests. Poly-drug use does not seem to provide any special improvement to predictions of being rearrested.[54] A positive test for marijuana use seems to improve predictions in

---

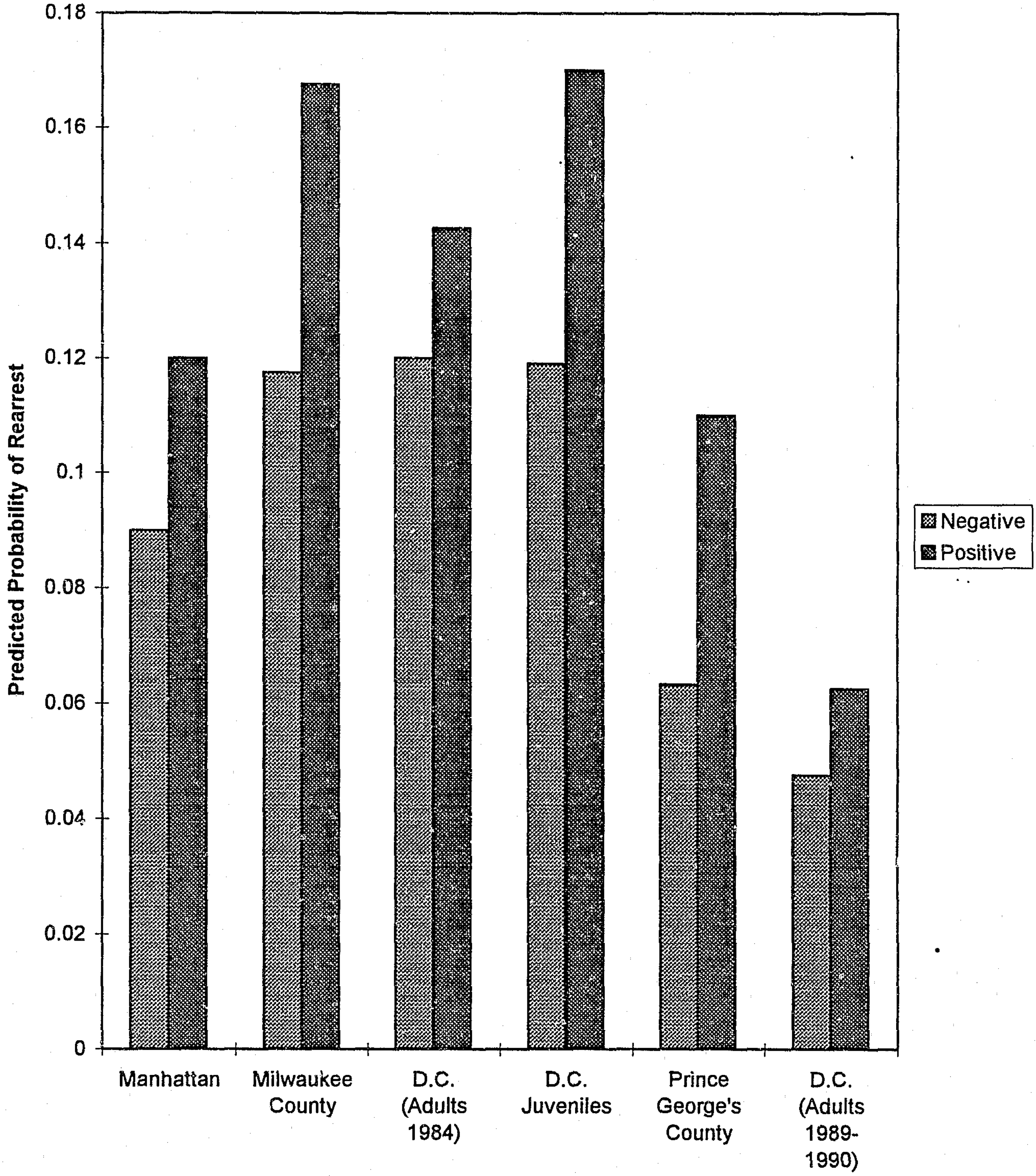
[52] Visher and McFadden (1991) define a false positive test as "a test result indicating positive for a given drug when the drug is actually absent in a urine sample or present in concentrations below the designated cutoff level" (p. 3). According to Visher and McFadden, the EMIT test (used in all settings) has a false positive rate of 2.2 percent. Furthermore, the urine test for opiates may identify drugs that are used for medical purposes: Percodan, Dilaudid, Demerol, and Codeine.

[53] Both a weighted and unweighted Stouffer combined test causes us to reject the null hypothesis at $p<0.05$.

[54] When a defendant tested positive for two drugs (X and Y) we set a dummy variable for drug X equal to 1, we set a dummy variable for drug Y equal to 1, and we set a dummy variable for poly-drug use equal to 1. This model specification demands some care when interpreting the parameter for the poly-drug variable. Consider the results from Dade County. Suppose that a defendant tested positive for marijuana only. He would get a drug test score of 1.052 in the prediction. This is the parameter estimate for a positive test for marijuana. Suppose that a defendant tested positive for just cocaine. He would get a drug test score of 0.613. This is the parameter estimate for cocaine. But if he tested positive for both cocaine and marijuana, then his overall score would be $1.052+0.613-0.924 = 0.741$.

Figure 2

Predicted Probability of Rearrest Within 90 Days: Those Who Tested Positive for Recent Opiate Use



Dade County. Overall, however, a positive test for heroin appears to be the only drug test that helps predict pretrial rearrests.

Table 20 summarizes statistical tests of whether drug test results for individual drugs help predict a failure to appear once other factors are taken into account. A positive test for recent cocaine use appears to provide some improvement in the ability to predict failure to appear. The improvement is statistically significant in three settings: Prince George's County, Maricopa County, and Washington, D.C. (adults, 1989-1990).[55] Moreover, the effect is positive in all sites (no data is available about failure to appear among District of Columbia juveniles). When results are combined across sites, the ability of cocaine tests to predict failure to appear is statistically significant.[56] As figure 3 indicates,[57] the effect might be considered substantively significant in several sites.

With regard to other drugs, however, drug test results do not seem to be useful in predicting failure to appear except for PCP in the District of Columbia (adults, 1984) and amphetamines in Maricopa County. Inexplicably, however, recent PCP users in D.C. appear to be least likely to fail to appear for court dates.

Using variables other than drug tests to predict misconduct. Our analysis shows that variables other than (or in addition to) drug test results are correlated with pretrial misconduct. A criminal history seems to be the best predictor of being rearrested during the pretrial period. The number of previous

---

[55] We discount the results heavily from Maricopa County because of the small sample size. See note 48.

[56] That the parameter estimates are positive in all sites, that they are statistically significant in three sites, and that they approach statistical significance in three other sites, provides evidence that cocaine use helps predict failure to appear. Consider, first, the probability that all parameter estimates would be positive when the null hypothesis was in fact true (cocaine does not predict failure to appear): $0.5^7=0.007$. We would reject the null hypothesis on the basis of this test statistics. Consider, second, both the weighted and unweighted Stouffer combined test. Based on these test statistics, we again reject the null hypothesis ($p<0.05$).

[57] Figure 3 is based on the regressions with failure to appear as the dependent variable. The regressions were evaluated at the mean values for all variables other than drug test results.

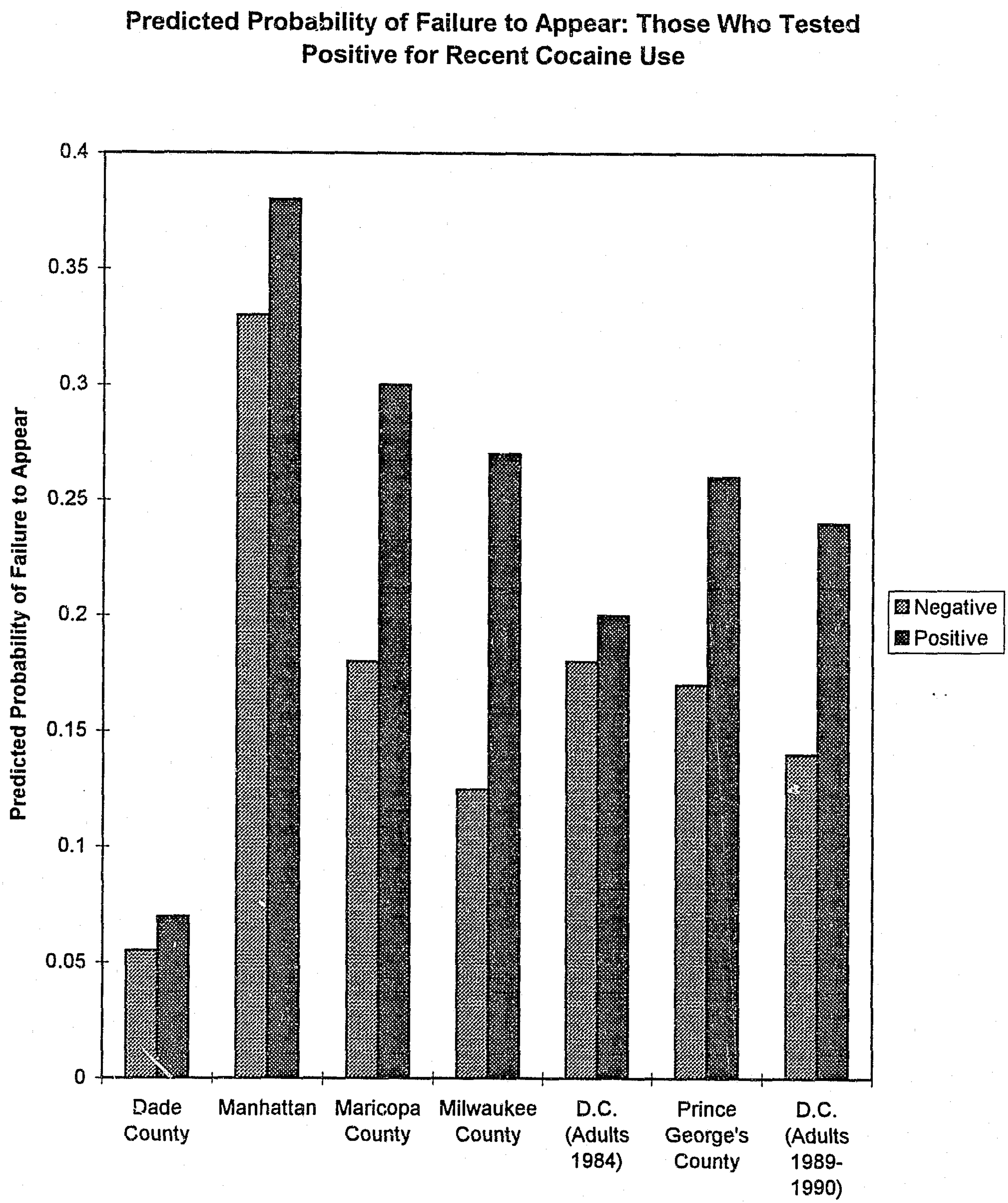
Table 20

Summary of Regression Results for Failure to Appear

Drug Type	Dade County	D.C. Adults 1984	Manhattan	Maricopa County	Milwaukee County	Prince George's County	D.C. Adults 1989-1990
Marijuana	-0.319					-0.019	0.024
	-1.206					-0.176	0.170
Cocaine*	0.120	0.107	0.116	0.424	0.523	0.335	0.285
	0.854	1.500	1.550	2.010	1.270	2.347	3.030
Opiates		0.081	-0.004		-0.007	0.026	0.151
		1.170	-0.050		-0.020	0.184	0.970
Amphetamines				0.455	-0.136		
				1.780	-0.270		
Benzodiazapines					0.320		
					1.380		
PCP*	0.090	-0.142				-0.099	-0.062
	0.800	-2.490				-0.853	-0.370
Other drugs				-0.014			0.123
				-0.030			0.760
Methadone		-0.207	0.087				
		-1.480	0.580				
Poly-drug	0.318	-0.052	-0.115		-0.310	0.035	-0.098
	1.116	-0.580	-0.830		-0.720	0.189	-0.570
Number of cases	1,294	5,689	1,893	186	829	1,072	1,538

* Denotes significant at $p < 0.05$ in a weighted Stouffer test

Figure 3


Predicted Probability of Failure to Appear: Those Who Tested Positive for Recent Cocaine Use
Predicted Probability of Failure to Appear
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Dade County
Manhattan
Maricopa County
Milwaukee County
D.C. (Adults 1984)
Prince George's County
D.C. (Adults 1989-1990)
Negative
Positive

arrests (or convictions when arrests was unknown) is highly significant in all the regressions. The importance of other indexes of a criminal record--such as number of previous probation/parole revocations, current supervision status, and history of incarcerations -- play a role that varies across the sites. A criminal record does not play any strong, consistent role in predicting failure to appear for court dates.

The seriousness of the instant offense (the initial arrest charge) has no affect on the probability of a rearrest. We find some evidence, however, that the probability of a failure to appear decreases with the seriousness of the instant offense. This is hard to explain. Defendants who are charged with serious crimes would seem to be most motivated to abscond.

Data about an arrestee's community ties were often missing and could not be entered into the analysis. When the data were available, employment and school attendance were useful when predicting rearrest and failure to appear. Marital status and length of time living in the community seemed to play no strong role when predicting misconduct.

Post-release drug-testing supervision had no apparent effect on being arrested during the pretrial period. This conclusion also holds for failure to appear, except in two sites (Dade County and Prince George's County) where failure to appear was higher for defendants placed under some forms of pretrial drug-testing supervision. Such findings are hard to interpret. Although post-release drug-testing supervision is designed to reduce pretrial misconduct, defendants under supervision may do worse than others for two reasons. First, the worst risks may be placed under supervision, and second, supervision may increase the visibility of the defendant's misconduct to criminal justice authorities. We strongly caution against interpreting these findings as a test of the efficacy of pretrial drug-testing supervision.

Drug use among those arrested for the first time. Criminal records appear to be more useful than drug test results for predicting rearrests. What about first-time arrestees? Is a drug test useful when predicting misconduct among those who have no recorded criminal history?

The evidence is mixed. We replicated the analysis reported above in five settings where the sample sizes were sufficiently large to support an analysis of pretrial rearrest rates for first-time arrestees.[58] We found essentially no effect in three sites: Dade County, Washington, D.C. (juveniles), and Manhattan. We found statistically significant *negative* effects in Washington, D.C. (adults, 1984); that is, first-time arrestees who tested positive for cocaine or heroin in Washington were less likely to be rearrested than first-time arrestees who tested negative for drugs. However, we reached just the opposite conclusions when analyzing the 1989-1990 D.C. data: first-time arrestees who tested positive for cocaine were more likely to be rearrested.[59] Even when a drug test improves the predictions for first-time arrestees, the substantive significance may be small. Knowledge of a drug test result does not necessarily increase the probability of misconduct to a level that is higher than that of repeat offenders. That is, first-time drug users who test positive for illicit substances are still better risks for release than are repeat offenders.[60]

Comparing these findings to other studies. Table 19 summarizes our findings from analyzing rearrests. Table 2 summarizes the findings of researchers who previously analyzed these data. Similarly,

---

[58] There were 206 first-time arrestees in Dade County. The parameter estimates had t-scores of 1.98 for marijuana, 0.29 for cocaine, and -0.74 for both. There were 1414 first-time juvenile arrestees in the District of Columbia. The t-scores were 0.90 for cocaine, 0.31 for marijuana, 0.34 for heroin, 0.38 for PCP, and 0.04 for multiple drugs. There were 2268 first-time adult arrestees in the District of Columbia (adults, 1984). The t-scores were: cocaine (-0.10), heroin (-2.99), methadone (-1.11), and PCP (-2.90).

[59] The t-score was 1.93. Those who tested positive for marijuana were less likely to be rearrested (t=-1.44), as were those who tested positive for other drugs (t=-1.69).

[60] Smith and Polsenberg (1992), who analyzed the 1989–90 data for District of Columbia adults, report that a positive test for recent cocaine use is highly predictive of being rearrested, especially when the defendant has no previous criminal record. Nevertheless, they report that arrestees who had no prior criminal record and tested positive for drug use still had rearrest rates that were no higher than those for arrestees who tested negative but had prior criminal records. (We are indebted to Jan Chaiken for making this observation.)

table 20 summarizes our findings from analyzing failure to appear, and table 3 summarizes the findings of earlier researchers.

We conclude that a positive test for cocaine probably does not help predict rearrest, although the evidence is equivocal. Prior analysis supports this conclusion.[61] We conclude that a positive test for opiates helps predict rearrest. Prior analyses do not support a strong conclusion, but evidence from Manhattan and Washington, D.C. seems consistent. For reasons explained later, we could not reproduce the strong effects associated with positive tests for PCP in Manhattan and in Washington, D.C.

We conclude that a positive test for recent cocaine use helps predict failure to appear. Prior analysis supports this conclusion, but note that our findings differ from earlier findings in Milwaukee and Prince George's County. Three explanations are possible. First, the original researchers used only baseline data in their analysis (see appendix), while we combined the baseline and the experimental data. Second, the original researcher used a model specification that was likely to yield imprecise parameter estimates,[62] while our model specification avoids this problem. Third, the original researcher used fewer control variables than we included in our model. Note also that we find a much weaker effect in Washington, D.C. than was reported by the original researchers. Three explanations are possible. The first is that the original researcher selected a sample of arrestees that differed from the sample that we

[61] Our approach to analyzing the Smith and Polsenberg data differed from that used by those two researchers. They analyzed data from 1990; we combined data from 1989 and 1990. They analyzed rearrests during a follow-up period that ended at a common date in 1992; we analyzed rearrests during a follow-up period that ended with the case's disposition (or, at a common date in 1992 when the case was still open). They used a probit model; we used a failure-time model, and our specification was more inclusive than theirs. Finally, they focused on a positive test for cocaine, including other drugs in their regressions only through the variables "any positive test" and "number of positive tests." We retained our standard approach of including multiple drugs (cocaine, heroin, PCP, marijuana, other, and poly-drug) in the regression's specification. Despite these differences, we also found that a positive test for cocaine predicted rearrest.

[62] See note 11.

used.[63] The second is that the original researchers used a model specification that omitted several variables that we found to be strong predictors of failure to appear.[64] The third explanation is that the original researchers used specialized computing procedures (written especially for their study) which may have affected findings.

We found no evidence that a positive test for opiates could be used to predict failure to appear. Again, our findings disagree with those reported in Washington, D.C.,[65] and to a lesser extent in Manhattan. The Washington, D.C. results are understandable, for reasons already discussed. The Manhattan differences are not troubling because the original researcher report findings that are weak, and their statistical model includes variables that were not used in our analysis.[66]

How good are the predictions? Whether any of the variables included in our models, taken alone or in combination, are adequate for predicting pretrial misconduct is an open question. Answering this

---

[63] The report by Toborg et al. (1989) indicates that 7883 observations entered their analysis. (See Table A-1, NOB=7883.) Only 5689 observations entered into our analysis. We discovered that we could nearly replicate the Toborg sample size and statistical results by including observations that lacked a urine test and classifying them as having a "negative urine test." Thus, we presume that Toborg's decision to treat missing drug test results as equivalent to negative test results largely accounts for the difference between our findings and hers.

[64] Toborg et al. (1989) used the following control variables: EMPLOYED (equal to 1 if employed and 0 otherwise), PPP (equal to the number of pending cases, parole, and probation), and EX-CON (the number of prior convictions). We used variations of these variables, as well as Felony (equal to 1 if the instant offense was a felony and 0 otherwise: t-score = -3.4), In school (equal to 1 if the defendant was in school and 0 otherwise: t-score = -3.3), Married (equal to 1 if married or common law and 0 otherwise: t-score = -1.1), $Age/_{10}$ (t-score = -5.7) and $Age^2/_{1000}$ (4.5), Offense Seriousness (t-score = -3.5), Open case pending (t-score = -2.9), and Number of parole terms (t-score = -3.2).

[65] A positive test for opiates had a t-score of only 1.84. However, a positive test for opiates and cocaine combined has a t-score of 2.62.

[66] The original researchers included the variables "recommended for release" (t-score = -8.35), "qualified recommendation" (t-score = -5.29), "scheduled court appearances" (t-score = 2.83), and dummy variables for the instant offense (t-score ranging from -5.43 to -1.48). We chose not to use these variables for two reasons. The first is that the variables were not available in all data sets, and we sought to have consistent model specifications across sites. The second is that inclusion of the release recommendation can obscure the effect of other variables included in the regression.

question would require at least two steps not taken here. The first is to more carefully develop the structural model to consider alternative specifications, such as by introducing other variables and interaction terms. The second is to conduct test-retest analyses to determine by how much predictions in a calibration sample will deteriorate in a validation sample.

Putting these two issues aside, how good are the statistics at predicting misconduct within the calibration sample? Figures 4 through 10 provide summaries. Though we discuss the predictions from Dade County, predictions from other settings are similar.

In Dade County, we ranked all arrestees into five groups ranging from the fifth who were predicted to be at lowest risk to the fifth who were predicted to be at highest risk of pretrial misconduct. The figure distinguishes rearrest from failure to appear.[67] Roughly 5 percent of the low-risk group failed to appear for a court date during the follow-up period, compared with roughly 25 percent of the high-risk group. Roughly 4 percent of the low-risk group was rearrested during the follow-up period compared with about 29 percent of the high-risk group.

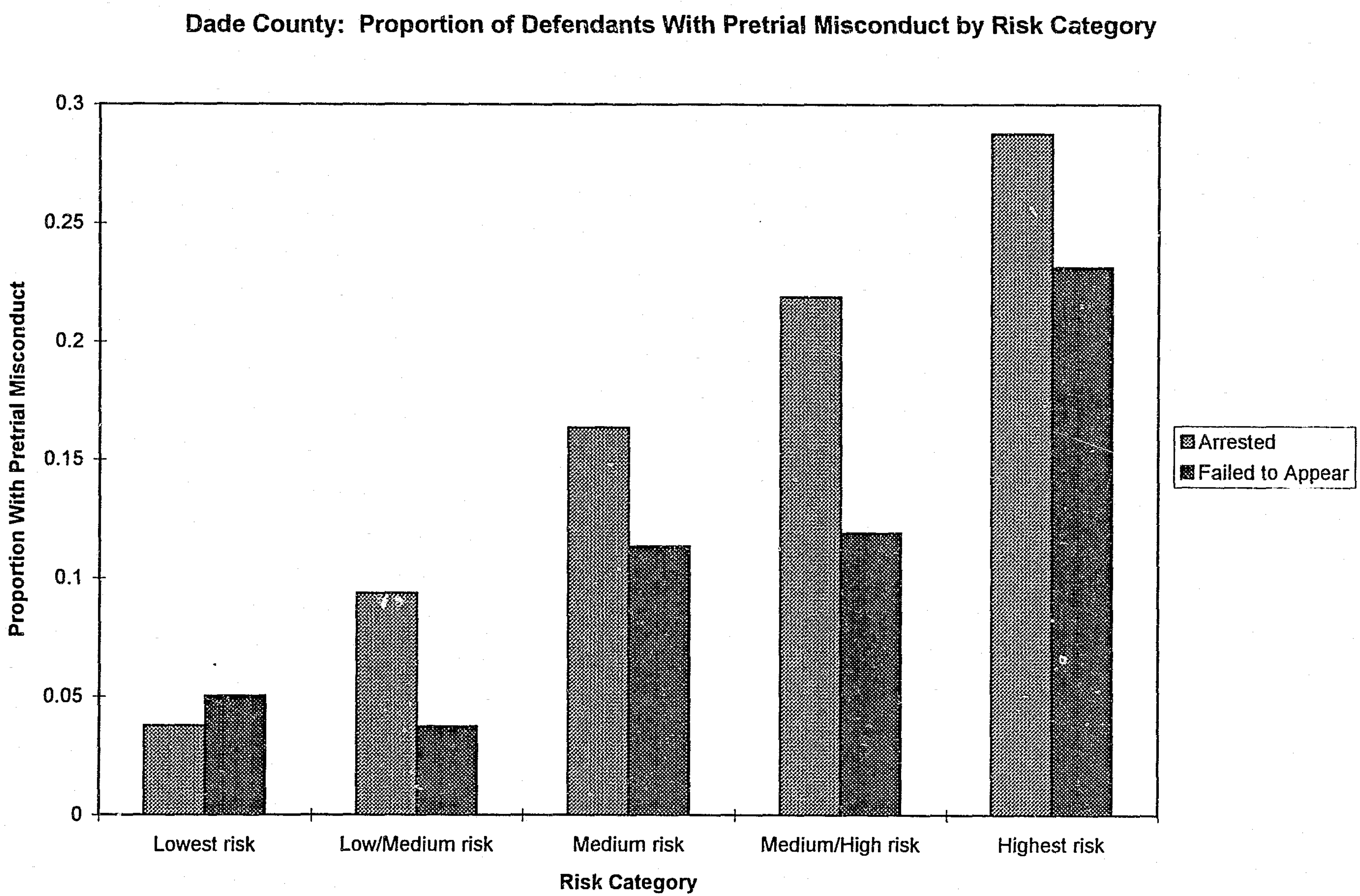
The predictions would not do so well in practice, because of shrinkage. Even taking shrinkage into account, however, the statistical analysis seems to have considerable power to distinguish between arrestees who would or who would not be arrested or fail to appear for a court date. We leave it to the reader's discretion to judge whether or not the predictive power is sufficient to advocate the use of such predictive instruments for making pretrial release decisions.

Summary and Interpretation

The results of our study show that arrestees who test positive for recent drug use are more likely to be rearrested during pretrial release than those who test negative. They are also more likely to fail to appear for scheduled court dates. Once other factors (especially criminal history and community ties)

[67] The low-risk group for failure to appear is not the same as the low-risk group for pretrial arrest. The same is true of the high risk group and the other three groups in the figure.

Figure 4



Dade County: Proportion of Defendants With Pretrial Misconduct by Risk Category
Proportion With Pretrial Misconduct
0.3
0.25
0.2
0.15
0.1
0.05
0
Lowest risk
Low/Medium risk
Medium risk
Medium/High risk
Highest risk
Risk Category
Arrested
Failed to Appear

Prince George's County: Proportion of Defendants With Pretrial Misconduct by Risk Category

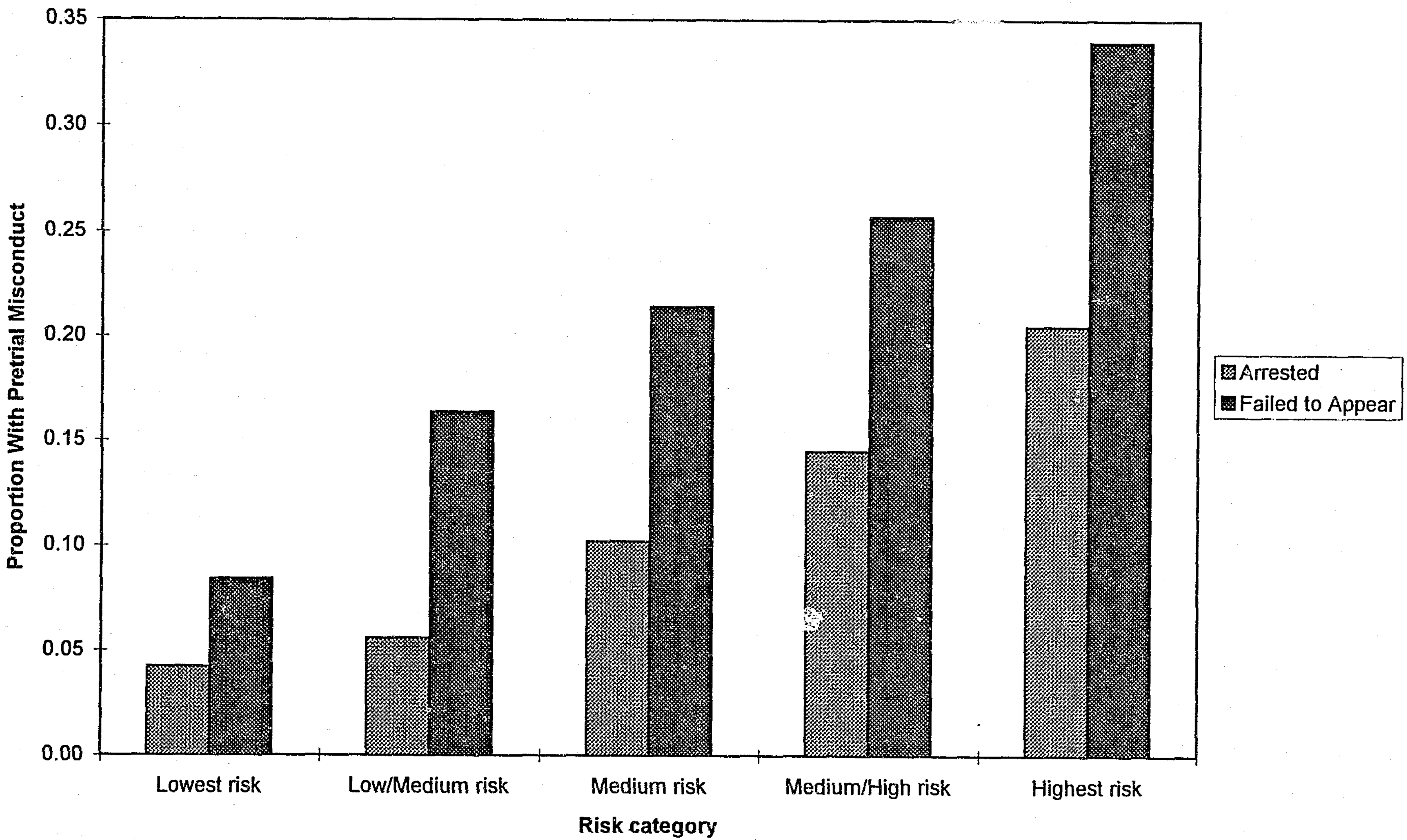


Figure 6

Maricopa County: Proportion of Defendants With Pretrial Misconduct by Risk Category

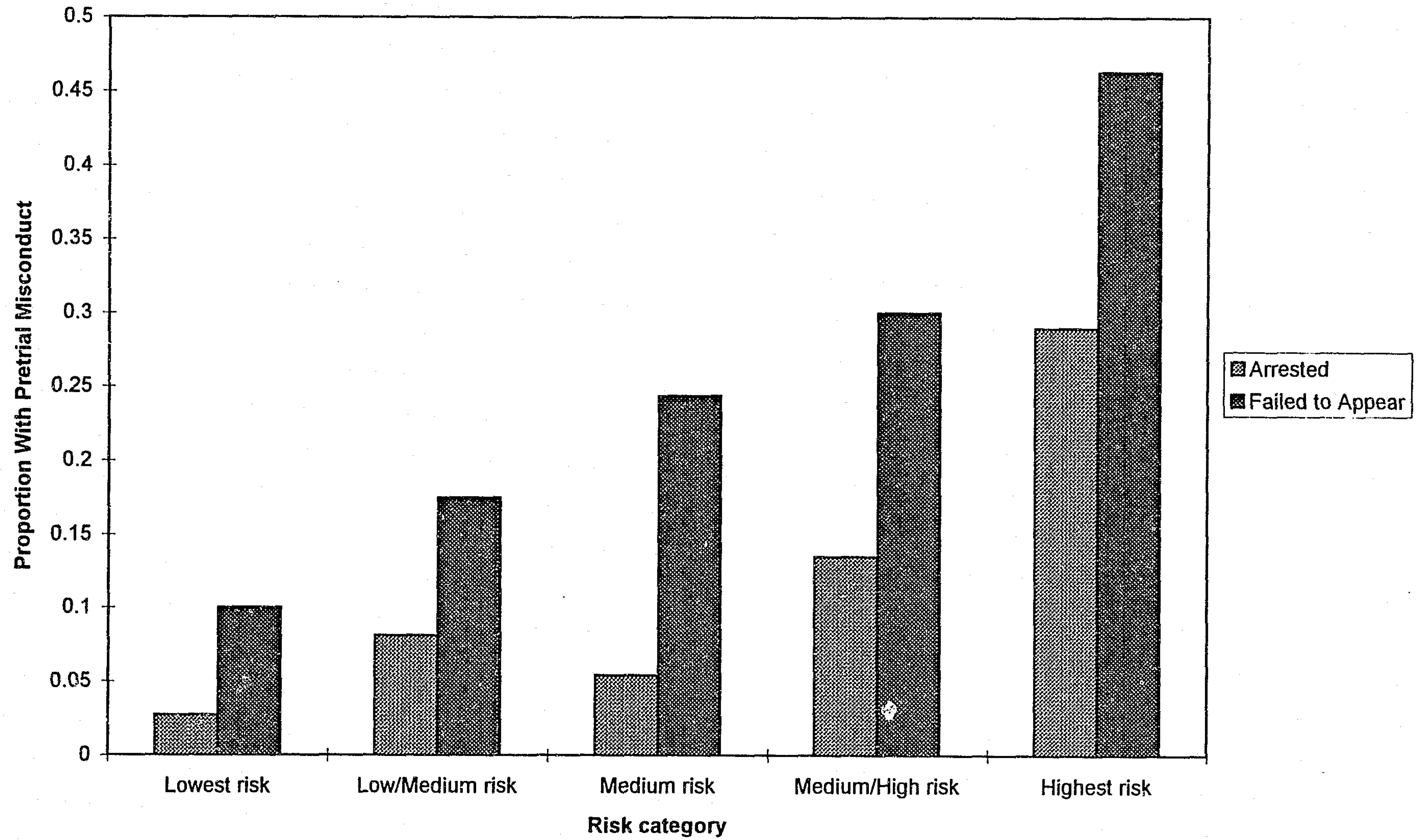


Figure 7

Milwaukee County: Proportion of Defendants With Pretrial Misconduct by Risk Category

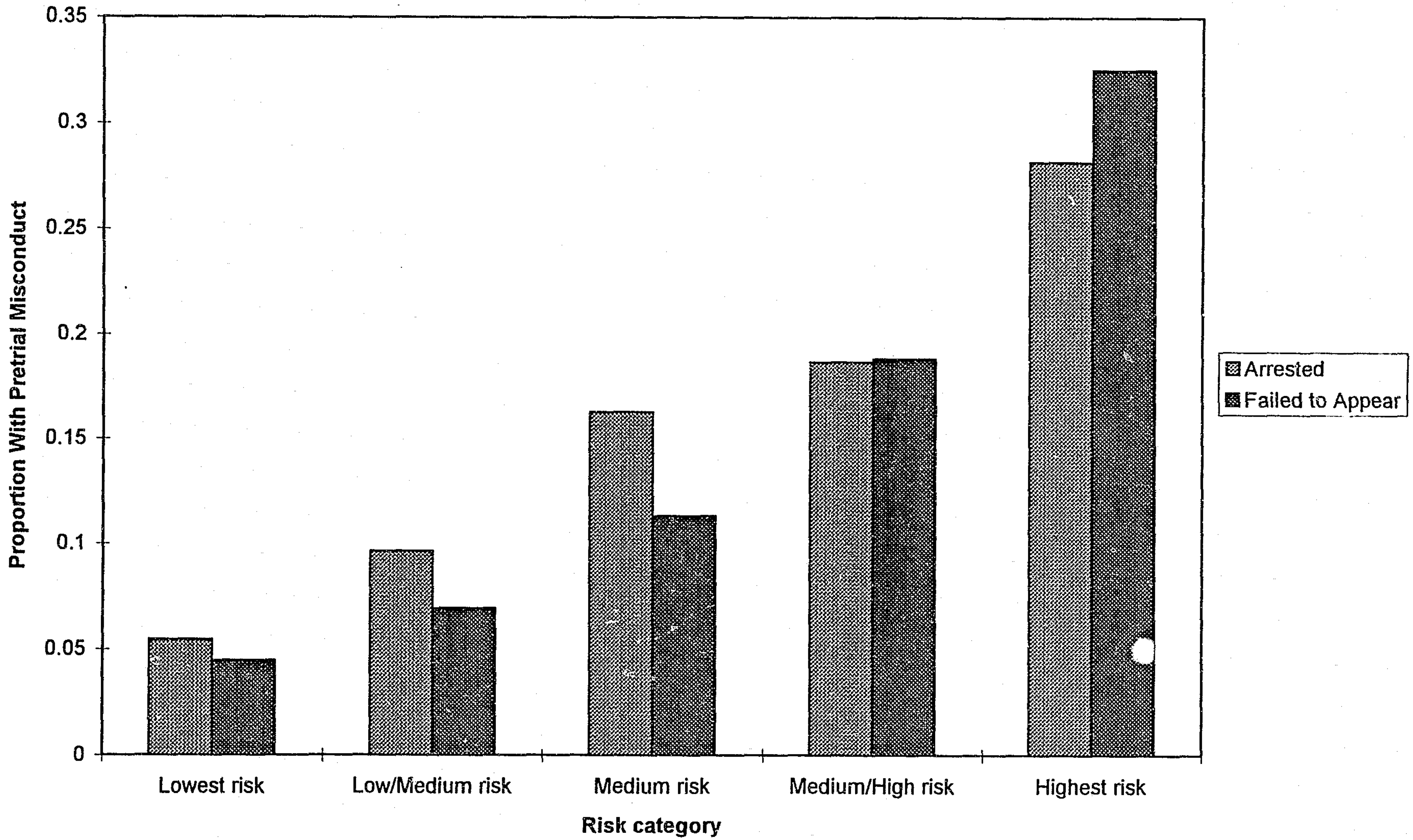


Figure 8

Washington, D.C. (Adults, 1984): Proportion of Defendants With Pretrial Misconduct by Risk Category

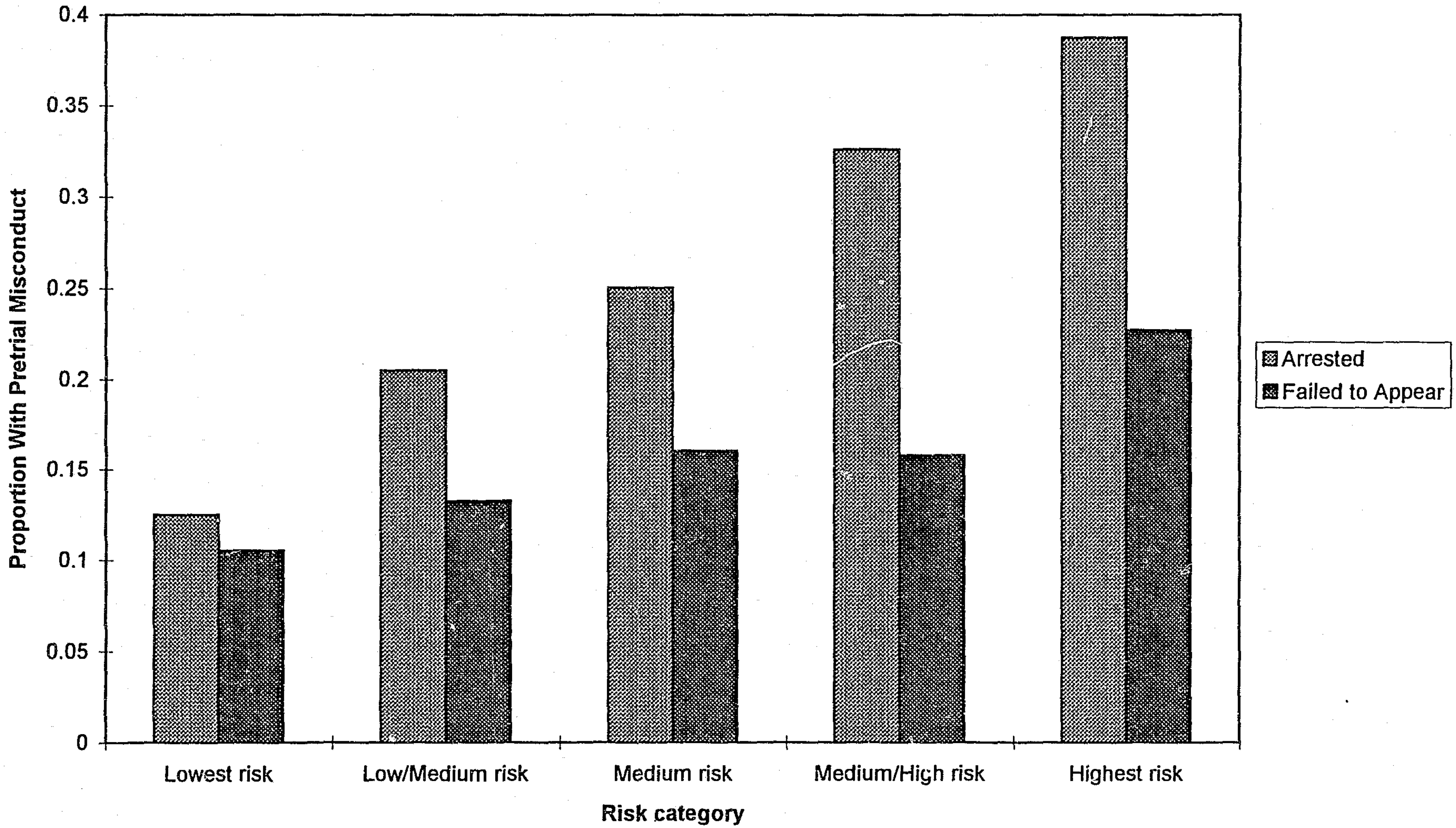


Figure 9

Manhattan: Proportion of Defendants With Misconduct by Risk Category

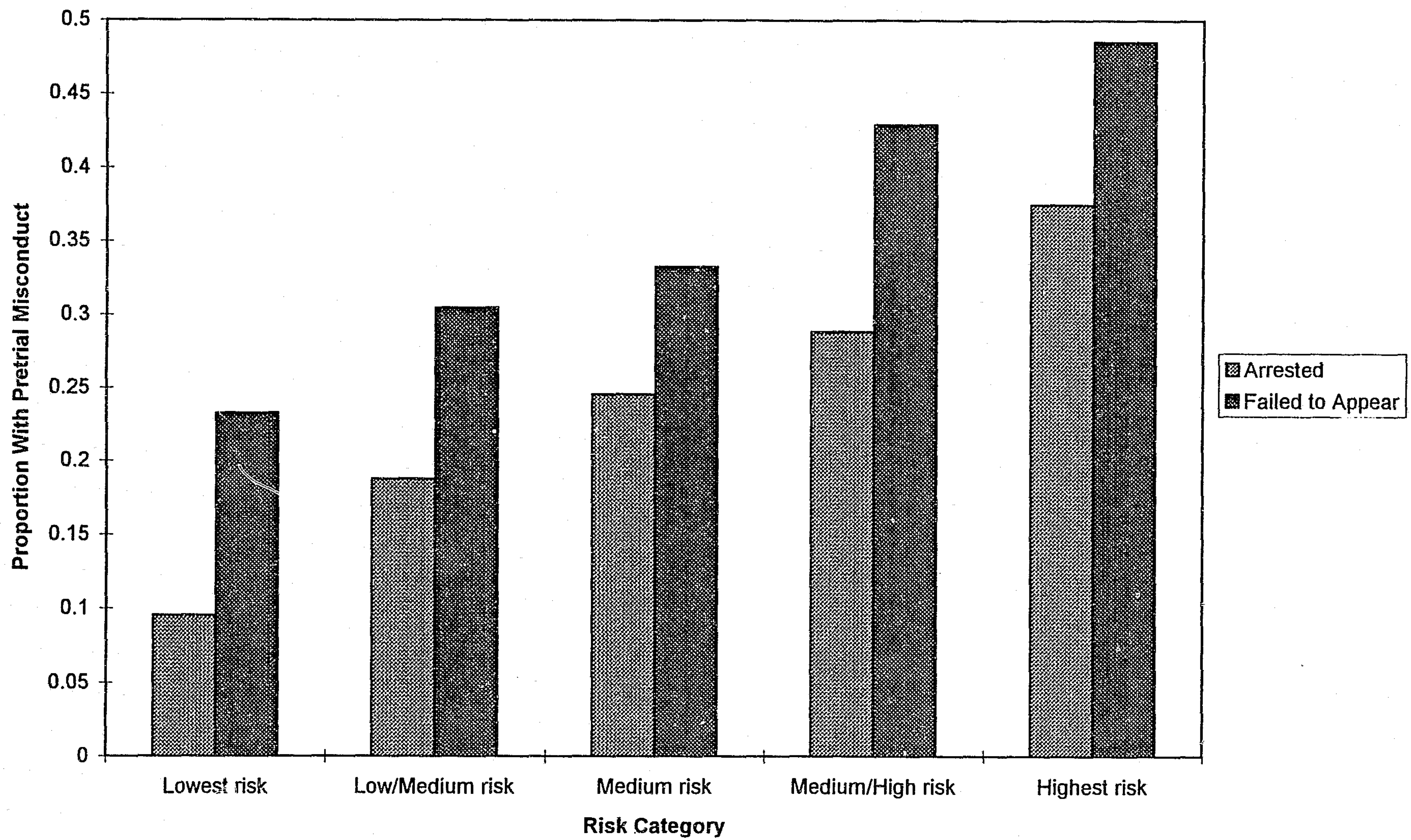
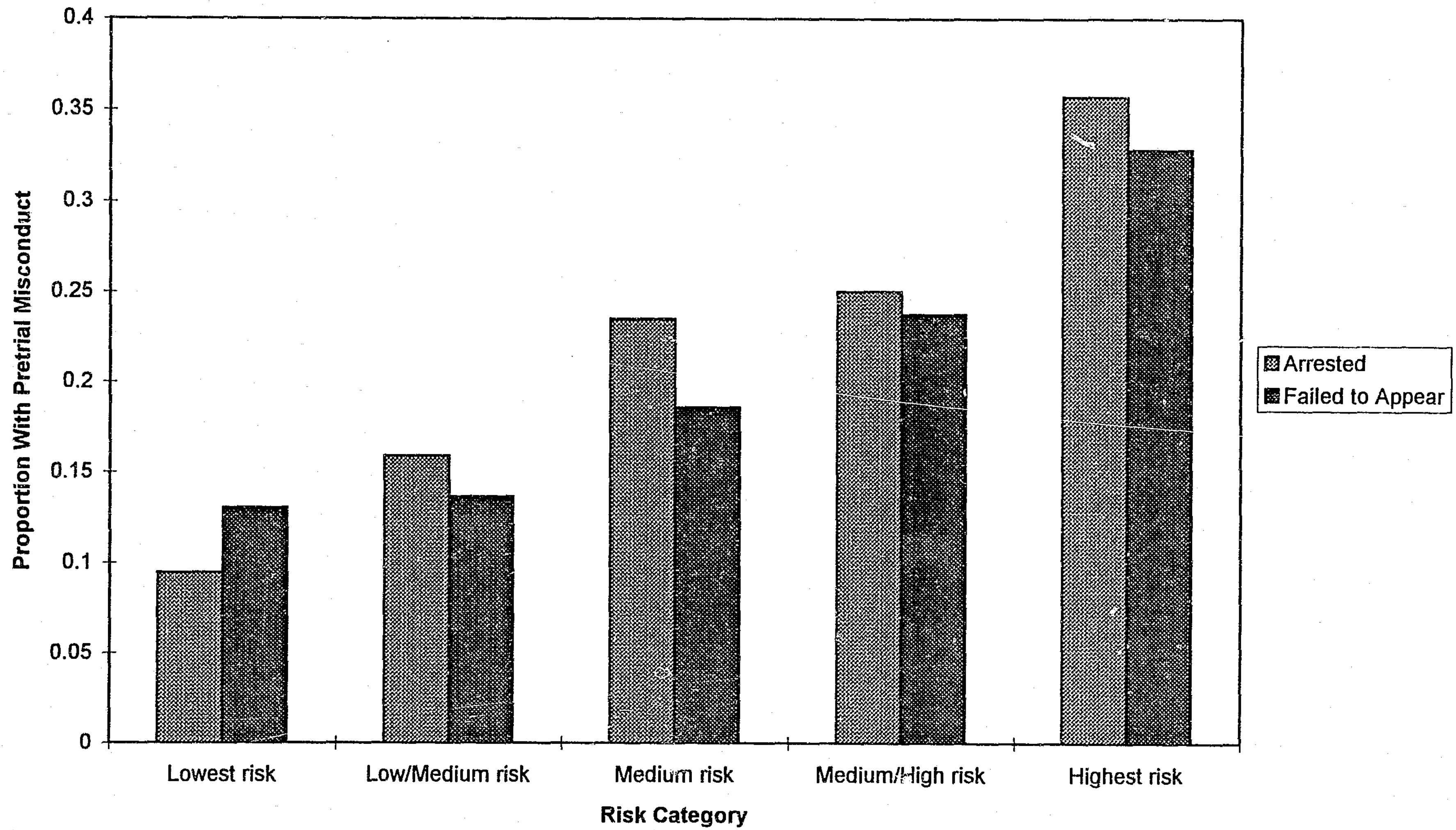


Figure 10

Washington, D.C. (Adults, 1989-1990): Proportion of Defendants With Pretrial Misconduct by Risk Category



are taken into account, however, no evidence indicates that arrestees who test positive pose greater risks of pretrial misconduct. Nevertheless, evidence supports the conclusion that those who test positive for heroin are more likely to be rearrested during pretrial release, and that those who test positive for cocaine are more likely to fail to appear for scheduled court dates.

Interpreting these findings is complicated by the absence of any strong theoretical grounding for the empirical analysis. Drug users do not comprise a homogeneous group (Chaiken and Johnson 1988). Some use compulsively; others use occasionally. Some undoubtedly steal or deal drugs to support their drug use; others pay with incomes gained legitimately. This lack of a theoretical grounding leaves us groping, for example, to explain why recent cocaine users fail to appear for trial, while recent heroin users commit new crimes, and recent drug users in general pose no special risk either to be rearrested or to fail to appear for scheduled court dates.

This lack of a theoretical grounding forces us to treat as equivalent results from different time periods and places. This equivalence is questionable. For example, only 18 percent of D.C. adult defendants tested positive for cocaine in 1984, while 73 percent of Dade County defendants tested positive in 1987. Are cocaine users in D.C. equivalent to those in Dade County? In 1984, the D.C. users were a relatively select group who were probably using powdered cocaine. Cocaine use was ubiquitous among arrestees in Dade County during 1987, and crack use probably predominated. Yet these two user groups are lumped together when we assume that a positive urine test for recent cocaine use is an equivalent marker for pretrial misconduct in both sites.

The validity of this assumption is doubtful. Some measure of drug use that is finer than a urine test, coupled with a more sophisticated understanding of the role that drug use plays in the lives of arrestees, might greatly advance the utility of drug testing at the time of pretrial release. In the meantime, what can we say about drug testing?

Urine Test Results Are an Imprecise Screen

Approximately 60 percent of all arrestees tested positive for at least one illicit substance.[68] Also, approximately 19 percent of the arrestees in our data were rearrested, and about 21 percent failed to appear for a court date. Suppose that pretrial misconduct is observed exclusively among those who test positive for recent drug use. Even in this extreme case only one of three would be rearrested. Only one in three would fail to appear for a court date. Thus, illicit drug use is so widespread among arrestees that a urine screen could not possibly provide an accurate predictor of pretrial misconduct.

Urine testing may be an insufficient screen for identifying defendants who become involved in pretrial misconduct. Part of the problem may be that urine testing does not reveal the intensity of drug use. That is, frequent users are indistinguishable from infrequent ones, and infrequent users may be an appreciable proportion of all arrestees who test positive for illicit substances.

Tabulations from DUF data make this point. From the 1990 DUF data we selected all subjects who both tested positive for cocaine and admitted cocaine use within 72 hours of the interview. (We assumed that they would be truthful about their overall drug use if they admitted recent use.) Although 40 percent of these subjects reported using cocaine on more than 20 days per month, 18 percent used cocaine on between 11 and 20 days, and 42 percent used cocaine on 10 or fewer days. Not all those who test positive are heavy users of cocaine.

Thus, many arrestees who test positive for illicit substances actually use illicit substances at what might be considered moderate rates.[69] By moderate, we mean that this level of usage would not necessarily compel the user to commit crimes to support his or her consumption. We mean also that the

---

[68] This percentage is based on a tabulation of DUF data for 1990. The percentage varies widely across the nation. About 75 tested positive in New York. Only 48 percent tested positive in San Antonio.

[69] Wish, Cuadrado, and Magura (1988) found about 36 percent of those who tested positive to self-report a need for treatment. Although the researchers consider this percentage an underestimate of those who need treatment, the estimate nevertheless suggests that many of those who test positive in jail are able to moderate their consumption.

frequency of use is sufficiently low that whatever aberrant behavior that follows from drug use (and which might bring the users to the attention of the criminal justice system) is also relatively infrequent.

A urine test is unable to distinguish between those users who are high intensity users and those who are not. This may be one reason why drug test results have limited utility when predicting pretrial misconduct. Almost all arrestees use illicit drugs at some time. A urine test simply does not help to separate those whose drug use is moderate from those whose drug use is heavy.

Are Heroin Users Different?

Heroin users may be different from cocaine users in important ways. Almost 74 percent of subjects who test positive for opiates claimed to use heroin on more than 20 days per month. In short, the majority were regular users, and most were probably addicted. This is in contrast to arrestees who test positive for cocaine, many of whom use cocaine less frequently.

This difference between cocaine users and heroin users may explain why drug tests that are positive for opiates help predict rearrests. Most heroin users are relatively old, having started using heroin in the 1970s or earlier (National Institute on Drug Abuse 1992) and they are not being replaced by new users (Hunt and Rhodes 1993). As documented in an extensive literature, criminality is exceptionally high among this group (Ball et al. 1981; Speckart, Anglin, and Deschenes 1989). Of course, given the lengthy drug use and criminal careers of these addicts, many are already known by criminal justice authorities. Alternative methods of identification may be practicable.

Other Information May Provide a Better Screen.

Drug use and crime may still be highly correlated even if urine tests do not predict pretrial misconduct. Repeat offenders leave a trail in the form of criminal records and unstable life-styles. This trail appears to provide the best prediction of future criminal behavior.[70] Tests for opiates excepted, the results from urine tests appear to tell us little more than what the offender's record already reveals.

An analogy is useful. A bank might screen a loan applicant by asking how much money she currently has in her pocket. Good loan risks probably tend to carry more money than those who are bad risks. The bank could use this information when considering the loan.

However, better information probably takes the form of past credit records. Bad loan risks have demonstrated prior difficulties in or failure to repay loans. The money-in-the-pocket test, although highly correlated with ability to pay a loan, probably tells the bank little about the risk of default when past credit history is taken into account.

This may be the case with criminal records and drug use in predicting pretrial misconduct. The criminal record in the criminal justice context is the counterpart of a bad credit history in the bank loan context. Results from a drug test are the counterpart of money in the pocket. As a bad loan-risk self-reveals through her past credit history, so too does a bad pretrial release risk through her criminal record. Additional information is almost superfluous.

Speculation: The Future of Drug Testing.

Pretrial drug testing has its advocates and opponents. Its advocates argue that drug tests help predict which arrestees will engage in pretrial misconduct. Its opponents argue that drug tests do not help predict, or if they do, the marginal improvement in prediction is not worth the cost.

[70] A panel of the National Research Council identified four classification scales that have been proposed or used to identify offenders who are likely to recidivate. The first scale (INSLAW) was based on four variables that pertain to criminal record, two variables that pertain to self-reports of drug use, and one variable that pertains to employment. The second scale (U.S. Parole Commission) relied on four criminal history variables, a self-report of heroin dependence, and age. A third scale (State of Iowa) used a criminal history score, a substance abuse score, and several other variables. The fourth scale (Rand) was based on four criminal history variables, information about the offense of conviction, and indications of heavy alcohol use and heroin use (Blumstein et al. 1986).

The analysis presented here cannot settle this argument, but it provides additional data for the debate. People who test positive for heroin use appear most likely to be rearrested; people who test positive for cocaine use appear most likely to fail to appear for scheduled court dates. These findings imply that drug test results help predict pretrial misconduct, but the evidence is not overwhelming.

Proponents and opponents can sift this evidence as they see fit. But, in our view, this sifting of old evidence runs the risk of becoming a sterile exercise. Considering how drug test results might be improved may be more fruitful.

In our view, much of the ambiguity about the utility of drug testing derives from the fact that drug tests cannot separate high-rate and low-rate users. We speculate that there are, or there are likely to be, ways to overcome this problem.

One way to overcome this problem may be to use more than one urine test to make predictions. This solution has two forms. The first form is to use drug test results from two or more sequential arrests to establish that the arrestee is a "problem user." Of course, this procedure would not work for first-time arrestees, but our analysis suggests that drug use tests are not especially useful for first-time arrestees anyway. Although this procedure would not work for drug-testing programs that have just been set up, many programs have long histories. For established programs, reconstructing drug test histories would be practical with computer assistance.

Our analysis provides no direct evidence that sequential drug testing would improve predictions of pretrial misconduct. However, evidence from the District of Columbia is supportive. Toborg et al. (1989) and Visher (1990) report that defendants who fail multiple drug tests during the period of pretrial release are most likely to engage in pretrial misconduct. Retrospective tests of drug use may be equally useful.

Evidence from Toborg and Visher also points toward the second form. Drug testing can be extended into the pretrial period so that sequential testing is prospective rather than retrospective. Not surprisingly, most programs that test at arrest give the judge the option to continue testing during pretrial release supervision. According to the Bureau of Justice Assistance, seven sites that replicated the program from the District of Columbia tested defendants designated by the courts an average of ten times during the pretrial period (BJA, 8). Of course, such prospective screens would be much less expensive if they could be limited to defendants who were identified by a criterion that was more selective than a single urine test.

Thus, the first way to overcome the problem of the imprecision of urine tests may be to use sequential retrospective and prospective tests. A second way to overcome this problem is to adopt alternatives to urine tests, such as hair testing,[71] for identifying drug users.

Chemical assay based on hair samples can measure long-term drug consumption. Much as a cross-section of a tree trunk reveals a season's rainfall by the width of the trunk's rings, hair tests show the use of drugs that accumulate as the follicle grows. Hair tests can show the timing and intensity of drug consumption. Drawbacks to the hair test stem from its newness: the technology is expensive, and not much experience has accumulated so far (Wish and Gropper 1990, Mieczkowski et al. 1993).

A third way to overcome this problem is to use information other than, or perhaps in addition to, a urine test. A history of drug treatment, or an admission of a need for treatment, may be suitable.

Wish, Cuadrado, and Magura (1988) provide supportive evidence. They classified a sample of 2500 Manhattan arrestees by arrest history, pretrial release recommendation,[72] drug test result, and self-report of drug dependency. Using the latter two variables, they grouped arrestees as "nonusers" (no positive urine test), "users" (positive urine test but no admission of dependency), and "dependent" (positive urine test and admission of dependency). Those deemed dependent had weighted failure to appear rates that were 0.114 probability points higher than those who were users and 0.151 probability

[71] Hair testing has been lauded for its ability to identify drug users who are missed by urine testing methods. This is the wrong virtue to extol if we are correct that the problem with urine tests is that they identify too many low-intensity users.

[72] The New York City Criminal Justice Agency had classified defendants as being recommended for release (with qualifications or without qualifications) or not recommended for release (because of insufficient ties or outstanding warrants/other problems).

points higher than those who were nonusers (Wish, Cuadrado, and Magura 1988, table 19).[73] Given an overall probability of failure equal to 0.391, these differences are impressive.

Of course, evidence is compelling that many arrestees will deny use of illicit substances,[74] and even if they admit use, those in need of treatment often deny that need. Such evidence is sometimes cited as a reason for using urine testing instead of self-reports.

We have no quibble with this evidence. It may miss the point, however. Our view is that urine tests are already too comprehensive. Self-reports of heavy drug use seem to predict misconduct (Rhodes 1985), and if the criminal justice response is not heavily punitive, arrestees who need and could benefit from special treatment might be willing to self-report that need. Many heavy users would probably be misclassified, of course, but small loss seems to accrue from replacing comprehensive but ineffective programs with selective but effective ones.

Finally, as we asserted earlier, the use of drug tests to predict pretrial misconduct lacks strong theoretical grounding. Why do those who test positive for recent drug use pose elevated risks for pretrial

---

[73] The data supported eight comparisons. Arrestees who were deemed dependent on drugs had the highest failure to appear rates in all comparisons except for persons with an arrest in the prior two years who also received an unqualified recommendation for release. For this exception, 34 percent of those deemed dependent failed to appear compared with 35 percent of those who were users and 31 percent of those who were nonusers. The weighted average reported in the text includes this exception.

[74] People may provide inaccurate responses to questions about substance use for several reasons (Rouse, Kozel, and Richards 1985). Because illicit drug use has legal consequences, respondents may lie to avoid punishment, promises of confidentiality notwithstanding. Putting legal concerns aside, respondents may deny or understate their consumption because of social disapproval, or they may overstate their use because overstating consumption may help them gain entry to treatment. Even when willing to report their drug use, users frequently seem unable to provide accurate reports (Feucht, undated). Finally, the user's memory may be impaired as a result of drug abuse.

Numerous researchers have compared respondents' reported drug use with independent, objective measures, usually urine tests. Although most older studies have indicated that respondents will report 70 to 90 percent of their drug consumption, recent studies seem to show higher levels of underreporting (Elliot, Huizinga, and Menard 1989; Hubbard et al. 1989; Mieczkowski 1990; Hser et al. 1992; Falck et al. 1992).

Most validation studies have been conducted with populations who are involved with the criminal justice system. For example, arrestees appear to report current cocaine use only 50 percent of the time when questioned in a jail or lockup (Wish and O'Neil 1989; Hubbard et al. 1989; Harrison 1990; Mieczkowski 1992) and when interviewed to inform the judge at a bail hearing (Toborg et al. 1989). Offenders serving probation appear even less likely to be truthful (Wish, Cuadrado, and Martorana 1986).

misconduct? Is that risk conditional: that is, does it apply to drug users under some conditions but not under other conditions? If so, can those conditions be identified, and can some remedy be provided by the criminal justice system? Such questions cannot be answered with extant data, but continued behavioral research may provide a key to improving release decisions.

Appendix: Data

The data for this study came from six geographically and demographically diverse sites: Washington, D.C., (adults and juveniles), suburban D.C., Manhattan, Milwaukee, Phoenix, and Miami. The District of Columbia pretrial release program served as a prototype for many of the other programs.

District of Columbia (Adults, 1984)[75]

The D.C. Pretrial Services Agency (PSA) implemented the first research-oriented pretrial urine-testing program designed to interview all arrestees, make release recommendations to the court, and monitor compliance with pretrial release conditions for those placed under supervised release.

All arrestees except those arrested for Federal offenses or "relatively minor crimes" were included in the program. At the time of arrest, the PSA tested and interviewed adult arrestees who agreed to voluntary drug testing for opiates, cocaine, PCP, amphetamines and methadone. At a bail hearing (an adversarial hearing with arguments from the defense attorney and the prosecutor), PSA staff made recommendations to the hearing commissioner, who then made a decision about release. Arrestees were either released on nonfinancial conditions [restrictions on travel, association, or behavior (including drug use)], released on financial conditions (including bond, cash, or surety), released to a third party for supervision, or detained. PSA monitored all those under nonfinancial release conditions.

During the period when the data were collected (the study followed those arrested from June 1984 to January 1985), the PSA ran a monitoring experiment. Individuals were eligible for participation in the experiment if they met four conditions: they tested positive or admitted drug use

[75] Toborg et al. 1989, p. vi.

when interviewed; they were not in drug treatment; they did not request referral to treatment at the time of the PSA interview; and they were released under drug-related release conditions to PSA.

Eligible arrestees were assigned randomly to three groups. The first group of 650 arrestees received periodic drug testing prior to disposition. The second group of 1,109 arrestees was referred to drug abuse treatment. The third group of 394 arrestees was a control group that was not subject to any special conditions. Arrestees in these groups were followed to final disposition, which was often as long as six to eight months after arrest.

District of Columbia Juveniles[76]

The District of Columbia Pretrial Services Administration (PSA) implemented an experimental pretrial drug testing program for juveniles who were processed through lockups between October 1986 and January 1988. The program had three components. The first was testing urine samples of arrestees for four illicit substances: marijuana, PCP, cocaine, and opiates. The tests were conducted in lockup for those juveniles who were processed through lockup, and on the day of the first court hearing for other juveniles. The second component was a randomized experiment in which those juveniles who tested positive were assigned to one of four groups whose members received varying degrees of urine testing: weekly, bimonthly, monthly, and never. A fifth group, consisting of juveniles whose urine tests were negative, was tested monthly. The consequence of a failure generally involved the imposition of more frequent testing. Treatment was prescribed for some juveniles whose urine tested positive. The third component provided for urine testing following adjudication.

As reported by Yezer (1990), about 37 percent of the youth were between 16 and 17, another 42 percent were between 14 and 15, and the rest were younger. Roughly 64 percent of the youths who were arrested during the 15-month window period had never been arrested, 16 percent had one

[76] Rhodes 1991.

prior arrest, and 20 percent had multiple prior arrests. During the 15-month follow-up period, 53 percent of the youths avoided an arrest, but 22 percent were arrested once, 11 percent were arrested twice, and 14 percent were arrested more than twice.

District of Columbia (Adults, 1989-1990)[77]

Smith and Polsenberg developed a third data set from the District of Columbia by merging data from the Drug Use Forecasting System (DUF) with data from the District of Columbia Pretrial Services Agency for 1989 and 1990. Defendants were tested for ten drugs (five by the PSA and ten by DUF). Because DUF is not a random sample, these data are not necessarily representative of arrestees in the District of Columbia.

By the time that Smith and Polsenberg collected these data, the District no longer conducted an experimental program of assigning a random sample of drug-involved arrestees to urine testing. However, post-release drug testing had been institutionalized by 1989, and the data show whether an arrestee was placed on urine testing, or was recommended for drug treatment, or both.

Prince George's County, Maryland[78]

The program in Prince George's County was based on the Washington, D.C. adult pretrial services program. The county is part of suburban Washington, D.C. Most of its criminal justice population is seen as "spill-over" from the District.

Following arrest and booking, arrestees were arraigned before a bail commissioner for an initial bail hearing. Some were released under financial or nonfinancial conditions, to supervision under the Pretrial Release Division, or both. Others were detained in the county correctional facility pending a prebond review (conducted by a district court judge) and drug testing. Some of the

---

[77] Smith and Polsenberg 1992.

[78] Goldkamp et al. 1990b, Goldkamp et al. 1990c.

arrestees who were ineligible for drug-test monitoring were released on their own recognizance, released on financial conditions only, or detained.

The study design required four samples. The first was a baseline, which was collected before the experiment's implementation. Arrestees were tested for drugs on a voluntary basis, but test results were concealed from the court. Drug tests included cocaine, opiates, marijuana, and PCP. Once the baseline sample was completed, a new group of arrestees released under the supervision of the Pretrial Release Division were tested for recent drug use, and the results were provided to the court. Arrestees who were eligible for the study (by having an initial positive drug test and being released to the Pretrial Release Division) were randomly assigned to either an experimental or a control group. The experimental group participated in drug monitoring in addition to standard pretrial supervision procedures. Those in the control group received standard pretrial supervision but no drug monitoring. A sample of arrestees who were ineligible for the experimental and control groups (because they tested negative for drugs or did not meet other criteria) were included as a fourth, comparison group.

Baseline data was collected from July 15, 1988 to August 26, 1988, yielding a sample of 506. Experimental, control group, and comparison group data were collected from August 26, 1988 to February 15, 1989. The experimental group comprised 298 arrestees, the control group 289, and the group of noneligible arrestees 251 individuals. All participants were followed for 120 days or until disposition, whichever came first.[79]

Milwaukee County, Wisconsin[80]

Following arrest and booking, arrestees were interviewed by the Wisconsin Correctional Services bail evaluation staff. At this point, any special requests from the judge, district attorney, and

---

[79] The pretrial period for most arrestees extended beyond 120 days.

[80] Goldkamp, Jones, and Gottfredson 1990c.

public defender were taken into account. At the prebail hearing, arrestees booked on felonies, serious misdemeanors (including misdemeanors involving crimes against persons, and weapons or drug charges), "high-risk" misdemeanors (as determined by WCS bail guidelines), and those with outstanding bench warrants were asked to submit voluntarily to drug testing. At the initial bail hearing, the WCS would request that those arrestees who tested positive for drugs and whom the judge planned to release be released to WCS supervision. Arrestees who were detained or who were released on their own recognizance or under financial conditions were not eligible for the study.

The Milwaukee site's research design is nearly identical to that for Prince George's County. Drugs tested for were cocaine, opiates, amphetamines and benzodiazepines. Four sample groups were studied: a baseline sample was tested for drugs but the information was not relayed to the court for decision making. Following the baseline study, arrestees who tested positive for illegal drugs were divided randomly into either an experimental group or a control group. Experimental group members were tested regularly for drug use as well as receiving standard pretrial supervision. The control group members were tested, but they did not receive any special monitoring for drug use during the pretrial period, although they did participate in the standard pretrial supervision program. A comparison group of arrestees who tested negative for drug use was also studied.

All participants were followed for 90 days or until disposition of criminal charges, whichever occurred first.[81] Baseline data was collected from February 7, 1989 to March 20, 1989, yielding a sample size of 908. Experimental, control, and comparison group data were collected from March 20, 1989 to December 31, 1989. The experimental group comprised 389 arrestees, the control group 348, and the group of noneligible arrestees 351 individuals.

[81] About half the arrestees had pretrial periods that were longer than the study period.

Maricopa County, Arizona

Maricopa County includes the city of Phoenix. This study was begun in the summer of 1988. Approximately 500 felony arrestees who tested positive for drugs at booking were subsequently tested periodically for 90 days of the pretrial period to determine whether they were still using drugs. Drug tests included cocaine, amphetamines, and other drugs. Failure to appear and subsequent arrests were also tracked.

Manhattan, New York[82]

The study includes a sample of felony arrestees who were arrested for offenses that excluded possession and sales of drugs. Arrestees are brought to Manhattan Central Booking for identification and processing prior to arraignment. After a pretrial interview, arrestees were asked to participate in a voluntary confidential research interview and to provide a voluntary urine test for opiates, cocaine, PCP, and methadone. Drug test results were not made available to the court.

Male nondrug felony arrestees who were arrested from April to October 1984 were included in the study. Only those arrestees who were eligible for release and whose cases were continued at arraignment were eligible for study. Study subjects were followed until disposition or the end of the two-year study, whichever occurred first. A total of 2,606 arrestees were ultimately included in the study.

Dade County (Miami), Florida[83]

This study was not implemented as part of a service program, and drug test results were not made available to judges for release decisionmaking. Rather, judges were asked what difference the information would have made in their decisionmaking had they been made available. The study

[82] Smith, Wish, and Jarjoura 1989.

[83] Goldkamp, Gottfredson, and Weiland 1990a.

included felony cases likely to reach the first judicial stage (bond hearing) in Dade County's Circuit Court, excluding those arrestees who made bond within a few hours of arrest, those who were charged with capital offenses, offenses punishable by life imprisonment, and other nonbondable crimes under Florida law. Voluntarily submitted urine samples were tested for cocaine, marijuana, and alcohol.

Sampling occurred in June and July 1987. Arrestees who were released and included in the study were monitored for failure to appear, rearrest and rearrest for serious crimes against persons (including assaults, kidnapping, rape, robbery, murder, manslaughter, and arson with personal harm) either for 90 days or until adjudication of their cases, whichever occurred first. Drug use during the pretrial period was not monitored.

References

Austin, J.; Krisberg, B.; and Litsky, P. 1985. "The Effectiveness of Supervised Pretrial Release." *Crime & Delinquency*, 31: 4, 519-537.

Ball, J.; Rosen, L.; Flueck, J.A.; and Nurco, D.N. 1981. "The Criminality of Heroin Addicts When Addicted and When Off Opiates." In *The Drugs-Crime Connection*. Edited by J. Inciardi. Beverly Hills: Sage Publications.

Belenko, S.; Mara-Drita, I.; and McElroy, J. 1992. *Pre-Arraignment Drug Tests in the Pretrial Release Decision: Predicting Defendant Failure to Appear*. Brief Report Series. New York City Criminal Justice Agency, Inc.

Blumstein, A.; Cohen, J.; Roth, J.; and Visher, C.; eds. 1986. *Criminal Careers and Career Criminals*. Volume I. Washington, D.C.: National Academy of Sciences.

Boone, H. 1992. "The Deterrent Effect of Drug Testing in the Criminal Justice System: A Report From a National Systematic Evaluation." Unpublished paper.

Bureau of Justice Assistance, 1989. "Estimating the Cost of Drug Testing for a Pretrial Services Program." Washington, D.C.: Bureau of Justice Assistance.

Bureau of Justice Statistics, U.S. Department of Justice. 1992. *Compendium of Federal Justice Statistics, 1989*. Washington, D.C.: Government Printing Office, NCJ-134730.

Chaiken, M. and Johnson, B. 1988. *Characteristics of Different Types of Drug-Involved Offenders*. Washington, DC: National Institute of Justice.

Chaiken, J. and Chaiken, M. 1982. *Varieties of Criminal Behavior*. Santa Monica, California: Rand Corporation. Rand Report R-2814-NIJ.

Chaiken, J. and Chaiken, M. 1990. "Drugs and Predatory Crime." In *Drugs and Crime*, edited by M. Tonry and J. Wilson. Chicago: The University of Chicago Press.

Clark, J. 1989. "Estimating the Costs of Drug-Testing for a Pretrial Services Program." Washington, D.C.: Bureau of Justice Assistance, U.S. Department of Justice.

Elliott, D.S.; Huizinga, D.; and Menard, S. 1989. *Multiple Problem Youth: Delinquency, Substance Use, and Mental Health Problems*. New York: Springer-Verlag.

Falck, R.; Siegal, H.; Forney, M.; Wong, J.; and Carlson, R. 1992. "The Validity of Injection Drug Users Self-Reported Use of Opiates and Cocaine." *The Journal of Drug Issues* 22, no. 4: 823–832.

Feucht, T. Undated. "An Analysis of Drug Use among Female Arrestees in D.C.: The Use of Cocaine Among Prostitutes." Department of Sociology, Cleveland State University, Cleveland, OH. Unpublished paper.

Flinn, C. and Heckman, J. 1982. "New Methods for Analyzing Individual Event Histories." In *Sociological Methodology 1982*. Edited by S. Leinhardt. San Francisco: Jossey-Bass.

Gandossy, R.; Williams, J.; Cohen, J.; and Harwood, H. 1980. *Drugs and Crime: A Survey and Analysis of the Literature.* Washington, D.C.: National Institute of Justice.

Goldkamp, J.; Gottfredson, M.; and Weiland, D. 1988. "The Utility of Drug Testing in the Assessment of Defendant Risk at the Pretrial Release Decision." Draft report to the U.S. Department of Justice.

Goldkamp, J.; Gottfredson, M.; and Weiland, D. 1990a. "Pretrial Drug Testing and Arrestee Risk." *The Journal of Criminal Law and Criminology*, 81: 3, 585-652.

Goldkamp, J.; Jones, P.; Weiland, D.; and Gottfredson, M. 1990b. "Measuring the Impact of Drug Testing at the Pretrial Release Stage: Experimental Findings From Prince George's County and Milwaukee County." Final report to the Bureau of Justice Assistance, U.S. Department of Justice.

Goldkamp, J.; Jones, P.; and Gottfredson, M. 1990c. "Measuring the Impact of Drug Testing at the Pretrial Release Stage: Pretrial Drug Testing in Milwaukee County, New Castle County and Prince George's County." Preliminary report to the Bureau of Justice Assistance, U.S. Department of Justice.

Graham, M. 1987. *Controlling Drug Use and Crime: A Research Update.* Washington, D.C.: National Institute of Justice.

Harrison, L. 1990. *The Validity of Self-Reported Drug Use Among Arrestees.* (Washington, D.C.: Department of Justice, National Institute of Justice).

Harrison, L. and Gfroerer, J. 1992. "The Intersection of Drug Use and Criminal Behavior: Results From the National Household Survey on Drug Abuse." *Crime and Delinquency*, 38: 1, 422-443.

Hser, Y.; Anglin, D.; Wickens, T.; Brecht, M.; and Homer, J. 1992. *Techniques for the Estimation of Illicit Drug-Use Prevalence: An Overview of Relevant Issues.* Washington, D.C.: National Institute of Justice. NIJ Research Report No. NCJ 133786.

Hubbard, R.; Marsden, M.; Rachel, J.; Harwood, H.; Cavanaugh, E.; and Ginzburg, H. 1989. *Drug Abuse Treatment: A National Study of Effectiveness.* Chapel Hill: The University of North Carolina Press.

Hunt, D. and Rhodes, W. 1993. "Tracking the Incidence of Heroin Use." Washington, D.C.: Office of National Drug Control Policy.

Judge, G.; Griffiths, W.; Hill, R.; Lutkepohl, H.; and Lee, T. 1985. *The Theory and Practice of Econometrics*. Second Edition. Wiley: New York.

Kalbfleisch, J. and Prentice, R. 1980. *The Statistical Analysis of Failure Time Data*. New York: John Wiley and Sons, Inc.

Kapsch, S. and Sweeny, L. 1989, Revised 1990. "Multnomah County DMDA Project Implementation Report." Report to the Bureau of Justice Assistance, U.S. Department of Justice.

Kapsch, S. and Sweeny, L. 1990. "Multnomah County DMDA Program Evaluation Final Report." Report to the Bureau of Justice Assistance, U.S. Department of Justice.

Lancaster, T. 1990. *The Econometric Analysis of Transition Data.* New York: Cambridge University Press.

Leukefeld, C. and Tims, F., ed. 1988. *Compulsory Treatment of Drug Abuse: Research and Clinical Practice.* Washington, D.C.: National Institute on Drug Abuse, NIDA research monograph 86.

Maddala, J. 1983. *Limited-Dependent and Qualitative Dependent Variables in Econometrics.* Cambridge: Cambridge University Press.

Maltz, M. 1984. *Recidivism.* New York: Academic Press.

Mieczkowski, T. 1990. "The Accuracy of Self-Reported Drug Use: An Evaluation and Analysis of New Data." In R. Weisheit (ed.). *Drugs, Crime and the Criminal Justice System.* Cincinnati: Anderson Publishing.

Mieczkowski, T. 1992. "Immunochemical Hair Assays, Urinalysis, Self-Reported Use and the Measurement of Arrestee Cocaine and Marijuana Exposure in a Large Sample." Presented at the Annual Meetings, American Society of Criminology, New Orleans.

Mieczkowski, T.; Landress, H.; Newel, R.; and Coletti, S. 1993. *Testing Hair for Illicit Drug Use.* Washington, D.C.: National Institute of Justice.

National Institute on Drug Abuse. 1992. *Heroin Revisited: A Longitudinal Perspective on Heroin Abuse in the United States.* Washington, D.C.: National Institute on Drug Abuse.

Rhodes, W. 1985. "The Adequacy of Statistically Derived Prediction Instruments in the Face of Sample Selectivity." *Evaluation Review,* 9: 3, 369-382.

Rhodes, W. 1986. "A Survival Model With Dependent Competing Events and Right-Hand Censoring: Probation and Parole as an Illustration," Journal of Quantitative Criminology 2, no. 2 (June): 113–137.

Rhodes, W. 1989. "The Criminal Career: Estimates of the Duration and Frequency of Crime Commission." *Journal of Quantitative Criminology,* 5: 1, 3–32.

Rhodes, W. 1991. "Using Drug Testing at Arrest to Identify Juveniles Who Will Be Rearrested." Unpublished paper prepared under contract to Toborg Associates, Inc.

Rouse, B.; Kozel, N.; and Richards, L. 1985. *Self-Report Methods of Estimating Drug Use: Meeting Challenges to Validity.* Washington, D.C.: National Institute on Drug Abuse, NIDA research monograph 57.

Schmidt, P. and Witte, A. 1988. *Predicting Recidivism Using Survival Models.* New York: Springer-Verlag.

Segebarth, K. 1991. *Pretrial Services and Practices in the 1990s: Findings from the Enhanced Pretrial Services Project.* Washington, D.C.: National Association of Pretrial Services Agencies.

Smith, D.; Wish, E.; and Jarjoura, R. 1989. "Drug Use and Pretrial Misconduct in New York City." *Journal of Quantitative Criminology*, 5: 2, 101-126.

Smith, D. and C. Polsenberg. 1992. "Specifying the Relationship Between Arrestee Drug Test Results and Recidivism." *The Journal of Criminal Law and Criminology*, 83:2, 364-77.

Speckart, G.; Anglin, M.; and Deschenes, E. 1989. "Modeling the Longitudinal Impact of Legal Sanctions on Narcotics Use and Property Crime." *Journal of Quantitative Criminology*, 5:1, 33-56.

Spelman, W. 1994. *Criminal Incapacitation*. New York: Plenum Press.

Toborg, M.; Bellassai, J.; Yezer, A.; and Trost, R. 1989. *Assessment of Pretrial Urine Testing in the District of Columbia*. Washington, D.C.: National Institute of Justice.

Trussell, J. and Richards, T. 1985. "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure." *Sociological Methodology*.

Visher, C. 1992. "Pretrial Drug Testing: Panacea or Pandora's Box." *The Annals of the American Academy* 521: 112-131.

Visher, C. and Linster, R. 1990. "A Survival Model of Pretrial Failure." *Journal of Quantitative Criminology*, 6: 2, 153-84. N. Tuma, ed. San Francisco: Jossey-Boss.

Visher, C. 1990. "Using Drug Testing to Identify High-Risk Defendants on Release: A Study in the District of Columbia." *Journal of Criminal Justice*, 18: 321-32.

Visher, C. and McFadden, K. 1991. *A Comparison of Urinalysis Technologies for Drug Testing in Criminal Justice*. Washington, D.C.: National Institute of Justice.

Wish, E.; Cuadrado, M.; and Magura, S. 1988. "Drug Abuse as a Predictor of Pretrial Failure-to-Appear in Arrestees in Manhattan." Unpublished paper prepared under grant #83-IJ-CX-K048 to Narcotic and Drug Research Inc.

Wish, E.; Cuadrado, M.; and Martorana, J. 1986. "Estimates of Drug Use in Intensive Supervision Probationers: Results from a Pilot Study." *Federal Probation* 4: 4-16.

Wish, E., and Gropper, B. 1990. "Drug Testing by the Criminal Justice System." In *Drugs and Crime*, edited by M. Tonry and J. Wilson. Chicago: The University of Chicago Press.

Wish, E. and Johnson, B. 1986. "The Impact of Substance Abuse on Criminal Careers." In *Criminal Careers and Career Criminals*, edited by A. Blumstein, J. Cohen, J. Roth, and C. Visher. Volume II. Washington, D.C.: National Academy Press.

Wish, E. and O'Neil, J. 1989. *Drug Use Forecasting: January-May 1989*. Washington, D.C.: National Institute of Justice.

Wolf, F. 1986. *Meta-Analysis: Quantitative Methods for Research Synthesis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-001. Beverly Hills: Sage Publications.

Yamaguchi, K. 1986. "Alternative Approaches to Unobserved Heterogeneity in the Analysis of Repeatable Events." *Sociological Methodology* 1986, N. Tuma, ed. San Francisco: Jossey-Boss.

Yezer, T. 1990. "Characteristics of the Juvenile Urine-Testing Sample." Memo dated March 29, 1990.

END