THE ACCURACY OF SELF-REPORTED ARREST DATA

Jennifer Schore, Mathematica Policy Research
Rebecca Maynard, Mathematica Policy Research
Irving Piliavin, University of Wisconsin

During the past decade, increasing emphasis has been placed on conducting social experiments and demonstrations to determine the economic and social effects of various public policies and programs. In such experiments, social science research has had to rely heavily on data collected through personal interviews; consequently, it has become imperative that researchers be concerned with the veracity of the data. Furthermore, researchers must determine not only the overall accuracy of the data, but also the factors that significantly influence data quality. In particular, knowledge of the determinants of response error can be critical in evaluations of experimental programs: if participation in a program is itself a significant influence on response error, and if adjustment for this fact is not made, then the conclusions based on comparisons between experimental and control group members may be incorrect.

The causes of response error vary. It may stem from faulty recall, misunderstanding, or a desire on the part of the respondent to give the interviewer the "right", but not necessarily accurate, responses. When the interview involves sensitive material, such as questions concerning criminal activity, the respondent might intentionally misreport information because of embarrassment, for purposes of self-aggrandizement, because he or she mistrusts the interviewers, or due to concern about the eventual destination of the interview data. The problem may be further complicated if individuals are enrolled in an experimental program. Participants, relative to controls, may be more prone to under-report criminal activity for fear of being asked to leave the program if they admit such activity or because they have more of a sense of what responses are "expected" of them. On the other hand, participants may report more accurately because their participation in the program has given them more self-confidence and a more stable life-style.

One way of verifying certain types of self-reported data is by comparison with official records. Most such comparisons have focused on data such as earnings, welfare receipt, and medical histories, rather than the more sensitive fields of criminal behavior. However, while studies of the quality of self-reports of criminal behavior have identified response error as a problem, only one has addressed the combined issues of the magnitude and the sources of error,[1] and only one has analyzed the effect of program participation on response error.[2]

In the current study, we have accomplished three objectives: (1) measured the extent of response error in self-reports of arrests; (2) identified sources of variability in response error; and (3) assessed whether participation in a demonstration employment program affects the accuracy of responses. The next section describes the sample and data used in the analysis. Section III discusses the extent and sources of response error, and Section IV summarizes our conclusions.

II. THE SAMPLE AND DATA

This study uses data from two sources: interviews administered to a subsample of experimentals and controls in the national Supported Work demonstration[3] and their official police arrest records. The sample includes 434 individuals enrolled in the ex-offender and ex-addict target groups of the Oakland and San Francisco Supported Work samples, and 340 individuals enrolled in the ex-offender target group of the Hartford sample. This sample was found to be generally representative of ex-addicts and ex-offenders enrolled in the national Supported Work demonstration, with the exception that blacks were slightly under-represented.

An interview was administered to experimentals and controls when they enrolled in the demonstration. This interview included questions on basic demographic characteristics, employment experiences, prior drug use, and criminal histories. These enrollment data were updated by means of follow-up interviews, administered at nine-month intervals for up to three years after enrollment. Among the questions asked in these later interviews were the number of arrests and the types of charges for the three most recent arrests in the previous nine months. By the time the current study was conducted, at least one, and up to three, post-enrollment interviews were available for 545 of the 774 individuals for whom we had collected arrest records.

The official arrest records for the Oakland and San Francisco respondents were gathered in August 1977 by a California Department of Corrections employee, while those for the Hartford sample were collected in October 1978 by an MPR employee. In both cases, the number of pre-enrollment and post-enrollment arrests, the arrest dates, and associated charges for post-enrollment arrests were recorded on specially prepared coding forms. Official records were located for all but 48 sample members, for whom the completed coding form, therefore, indicated that there were no recorded arrests for either the pre- or post-enrollment period.[4]

III. THE EXTENT AND SOURCES OF RESPONSE ERROR

Estimates of response error and its sources were obtained through the use of ordinary least squares regression. The results are reported in Table 1. The specific models estimated assumed that response error was a linear function of demographic characteristics, prior criminal history, the true number of arrests, and program status (i.e., whether enrolled in the experimental or the control group of the Supported Work demonstration sample). To facilitate the analysis, three measures of arrest activity for both the pre- and

TABLE 1

ESTIMATED REGRESSION COEFFICIENTS

(MEAN VALUES OF INDEPENDENT VARIABLES IN PARENTHESES)

Independent Variable	Dependent Variable			
	Response Error[a] in Number of Pre-enrollment Arrests		Response Error[a] in Number of Post-enrollment Arrests	
Experimental Status	-.057	(.46)	-.150	(.48)
Ex-offender Target Group	-.869	(.88)	-.005	(.90)
Oakland Sample	-.407	(.36)	-.093	(.35)
Hartford Sample	-.293	(.41)	.135	(.41)
Age	-.151**	(26.03)	.002	(25.89)
Male	1.520	(.92)	-.220	(.93)
White	-.903	(.15)	-.292**	(.13)
Black	-.515	(.66)	-.044	(.69)
Regular Use of Heroin	-2.922**	(.35)	-.041	(.35)
Probation/Parole Status	-.651	(.75)	.074	(.75)
Weeks in Jail Prior 2 Years	.007	(49.98)	.000	(48.74)
2 or 3 Follow-up Interviews			-.103	(.33)
Number of Pre-enrollment Arrests from Records	.524**	(9.38)	-.010	(9.30)
Number of Pre-enrollment Arrests: Records minus Interviews			.003	(3.45)
Number of Post-enrollment Arrests from Records			.661**	(1.15)
≥ 2 Follow-up Interviews * Number of Post-enrollment Arrests			-.050	(.56)
Number of Post-enrollment Arrests from Records with a Main Charge of:				
Robbery			.140*	(.12)
Other Property			-.020	(.19)
Burglary			-.161	(.07)
Murder or Aggravated Assault			.031	(.06)
Other Crimes Against Persons			-.027	(.04)
Drug Offense			.295**	(.12)
Number of Arrests Expunged from Record			.377**	(.17)
Interviewed in Prison			-.382**	(.14)
Interactions of Experimental Status with:				
Number of Pre-enrollment Arrests from Records	.073	(4.24)	-.002	(4.34)
Ex-offender Target Group	-.370	(.39)	.165	(.41)
Oakland Sample	.579	(.19)	.242	(.19)
Hartford Sample	.534	(.15)	-.122	(.16)
Weeks in Jail Prior 2 Years	-.020	(23.48)		
Number of Arrests Expunged			.199	(.05)
Dependent Variable Mean	3.411		.549	
Sample Size	710		545	
$\bar{R}^2$	.201		.709	

*Statistically significant at the .10 level; **statistically significant at the .05 level.

[a] The number of arrests on the records minus the number of arrests reported on the interviews.

the post-enrollment periods were constructed: the number of arrests on the police records (the "true number"), the number of self-reported arrests, and the difference between the official and the self-reported number. It should be noted that for the post-enrollment variable, the number of arrests from the records includes only those arrests that occurred during the time periods covered by the available post-enrollment interview data.

With reference to the pre-enrollment period, we observed that 95 percent of the sample reported having been arrested at least once while only 90 percent had such an arrest in their official record. The apparent "over-reporting" occurred only among the San Francisco and Oakland samples and is most likely attributable to arrests having occurred outside the jurisdiction from which records data were obtained. Despite this tendency to "over-report" the occurrence of any arrest, the number of arrests was substantially under-reported: the average number of pre-enrollment arrests indicated on the official records was nine while the average reported on the interviews was only six.

Less than one-quarter of the discrepancy between the records and interview data was explained by the regression. However, three factors did seem to be significantly related to response error. Not surprisingly, response error increased significantly with the true number of arrests. This result is undoubtedly due in part to the fact that the error is bounded by the actual number of arrests. The other factors that related significantly to the extent of error were the respondent's age and whether or not he or she used heroin regularly prior to enrollment. Controlling for other factors, older individuals tended to report their arrest history more accurately than did younger persons. Similarly, interview data were more accurate than average among prior heroin users. Other factors, including program status, site, sex, race and parole/probation status, were not significantly related to response error on pre-enrollment arrests.

During the post-enrollment period, there was general under-reporting of both the occurrence of any arrest and the number of arrests. On average, individuals under-reported the number of their post-enrollment arrests by .55, or 48 percent of the official number. Over 70 percent of the response error for post-enrollment arrests was explained by the regression. In particular, there were strong significant relationships between the error and the official number of arrests (particularly arrests for drug offenses), race, whether the respondent received a follow-up interview while in prison, and the number of expunged arrests on a respondent's record.[5] As with pre-enrollment response error, a greater number of official arrests were associated with greater under-reporting. On average, white respondents under-reported only one-half as much as black and Spanish-speaking respondents. Those with arrests in which drug offenses were the main charge exhibited still greater under-reporting, possibly because respondents regarded drug arrests as relatively minor and, therefore, failed to report them. Similarly, respondents with a higher number of expunged arrests under-reported by more

than the average, suggesting that at least some erasures were excluded from self-reports. Finally, individuals to whom at least one follow-up interview was administered while they were in prison under-reported by .22 arrests, on average, while those who received no interviews in prison under-reported by an average of .60 arrests.

Factors that had no significant impact on post-enrollment response error include program status. Neither the coefficient on the status variable nor the adjusted mean difference in the error between experimentals and controls was statistically significant.[6] Similarly, site, target group, age, prior drug use, parole/probation status at enrollment and prior incarceration experiences were unrelated to post-enrollment response error.

IV. CONCLUSION

The main conclusions from this study are (1) that individuals will substantially under-report the number of times they have been arrested; (2) that the extent of under-reporting increases significantly as the true number of arrests increases; and (3) that participation in a program such as Supported Work does not significantly affect the response errors to questions concerning arrests. In terms of our particular interest in obtaining valid estimates of the impact of Supported Work on arrests, these results suggest that tests of whether Supported Work significantly influenced participants' frequency of arrest are valid, but that the estimated experimental-control differences underestimate the true differences. However, by setting up a predicting equation for official arrests as a function of self-reported arrests and other pre-enrollment characteristics, we can obtain rough estimates of true experimental-control differences.

As privacy legislation becomes more widespread and official records more difficult to obtain, researchers will have to rely increasingly on self-reported data, despite their possible inaccuracies. In particular, social scientists trying to examine the effects of a treatment program will be faced with the question of whether participation in the treatment itself is a source of response error. The results of this study are encouraging in that they indicate no differential under-reporting by treatment status and they provide us with useful information on both the size and determinants of response error.

[1]Wyner, G., "Report on Methodological Study With Vera," Mathematica Policy Research Internal Memorandum #SWP 1053, Princeton, N.J.: Mathematica Policy Research, Inc., February 1976.

[2]Vera Institute of Justice, Third Annual Research Report on Supported Employment, New York, N.Y.: The Vera Institute of Justice, 1975.

[3]For more information on the national Supported Work demonstration, see: MDRC, First Annual Report on the Supported Work Demonstration, New York, N.Y.: Manpower Demonstration Research Corporation, 1976.

[4] We recognize that some of these individuals may have been arrested outside the jurisdictions canvassed. This is especially likely among the ex-offenders for whom a condition for eligibility in the Supported Work demonstration is that they had been incarcerated within the last six months. (Forty-three of the 48 for whom no record was located were ex-offenders.)

[5] It is possible that an arrest can be expunged from the official records as the result of a successful probation period. The Connecticut police files maintained a record of whether a particular post-enrollment arrest had been expunged; the California files did not. At the onset of the analysis, there was a question as to whether expunged arrests should be included in the official numbers. Primarily because preliminary cross-tabulations did not suggest a clear exclusion of expunged arrests in respondents' reports, the erasures were included in the official counts. Thus, the number of erasures was included as a control variable in post-enrollment equations.

[6] We also regressed both official arrests and self-reported arrests on status and various demographic characteristics. The results indicated that program status had no statistically significant effect on either official or self-reported arrests, nor were these nonsignificant effects different from each other.