Rochester Youth Development Study
Hindelang Criminal Justice Research Center
The University at Albany
Albany, NY 12222

Technical Report No. 1
(revised)

Sampling Design and Implementation

Margaret Farnworth
Terence P. Thornberry
Alan J. Lizotte
Marvin D. Krohn

August 1991

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The central purpose of the Rochester Youth Development Study (RYDS) is to develop a comprehensive, reciprocal causal model of factors associated with the initiation and maintenance of serious delinquency. The theoretical model for this overall goal highlights the processual nature of delinquency and identifies the developmental process in which it is embedded. The analysis of these processes necessitates a panel design in which a single cohort is followed over time.

The panel for the Rochester Youth Development Study was designed to include both males and females, members of different racial and ethnic groups, and subjects from different socioeconomic levels. Seventh and eighth graders were targeted for the sample with the premise that a youth cohort with ages ranging from 11 to 14 at the outset (as of December 31, 1987) would be studied for the four years of data collection. The respondents' ages will range from eleven to seventeen years old over the life of the longitudinal study.

The first wave of data collection was initiated in the Spring of 1988. The purpose of this report is to describe the sampling plan for the study, document its implementation in the first wave of data collection, and describe the Wave 1 sample.

II. THE SAMPLING PLAN

A sampling plan was devised to yield a sample meeting two somewhat competing research criteria. The first goal was to provide enough variation in class, race, and other demographic characteristics to support an analysis of the factors leading to delinquency. The second was to include in the sample a reasonably high number of serious, chronic delinquents. The assurance of a sample that allows for the examination of chronic delinquent careers is crucial for two reasons. First, a proportionately small number of chronic offenders account for a disproportionate share of juvenile offenses, especially for serious and violent types. Second, it is important to have a substantial number of serious offenders in the analysis as a point of contrast

to the much larger number of juveniles involved in trivial offenses, so that important causal effects can be identified.

The first of these goals (to provide demographic diversity) guided decisions about the research site, the sampling frame, and the size of the planned sample. The second goal (to ensure a sufficient number of serious young offenders) was achieved by site selection and by identifying at that site a "high risk" sample stratified on two dimensions: residence and sex.

A. Study Site

The city of Rochester, New York was selected as the study site. Rochester is a medium-large city (with a population of 250,000) characterized by a demographic diversity suited to the purposes of the planned research project. The racial, ethnic, and economic diversity of the city is reflected in the characteristics of students in the public school system in Rochester. At the time the study began, the total public school enrollment in Rochester was 33,133. The most recent statistics available indicated that students in the ninth grade were about fifty-one percent black and ten percent Hispanic. The remainder were white or of other racial or ethnic origin (New York State Education Department, 1985).

The geographic site for the study also increased the probability that a good proportion of the sample drawn was at a high statistical risk for serious offending. The city of Rochester had a crime rate of 9,420 per 100,000 population in 1984, considerably above the national rate (5,031), that of New York State (5,577) and even that of New York City (8,375)(FBI, 1984). Rochester is also characterized by a well-defined inner city likely to exhibit particularly high crime rates.

B. Sampling Frame

The sampling frame for this study was the population of seventh and eighth grade students enrolled in the Rochester Public School System when the study began, with exclusions described below as "ineligible cases." The targeted sample was limited to the public schools within the city limits of Rochester, since suburban and/or private school students are less likely to exhibit serious and chronic offending behavior. This does not imply that delinquency and drug use are characteristics exclusive to students in city school systems. On the contrary, data

show clearly that these behaviors are distributed across students enrolled in all types of school systems. Delinquency is, however, more highly centralized in urban areas -- hence the decision to focus on students enrolled in city schools.

C. Ineligible Cases

Approximately four thousand (4,013) seventh and eighth grade students were enrolled in the Rochester Public School System in the Spring semester of 1988. Students with severe mental or physical disabilities were not included on the school system's roster or in this number. The list of students provided by the Rochester Public School System was revised to exclude cases defined as ineligible for the sampling frame. Cases were defined as ineligible if they displayed any one of the following characteristics:

(1) *Students who lived outside of or moved out of Monroe County, which encompasses the city of Rochester, before Wave 1 cases were fielded*: A small number of students listed as seventh or eighth grade students in the Rochester Public Schools moved from Monroe County before the field interviews began. Since they no longer met the residency requirement, they were replaced in the sampling frame. (Appendix A describes strategies for replacing cases.) Note, however, that if the subject moved out of Monroe County after participating in Wave 1 interviews, he or she was kept in the panel. Attempts were made to track the subject and conduct home interviews in the new geographical location.

(2) *Students from families where the language spoken in the home was neither English nor Spanish*: This was necessary because personal interviews were conducted with a parent of each subject in the student sample. It was not practical to translate a rather long parent interview schedule into many languages, and to hire many different native-language interviewers. The sizeable number of students from Spanish-speaking families and the importance of studying delinquency among Hispanic youths warranted a Spanish version of the parent interview schedule to retain this ethnic group in the sampling pool.

(3) *Those with a sibling already in the sample pool*: Home and family characteristics, measured with responses from both students and parents, constitute important variables in

several of the planned analyses. Since siblings share these characteristics, assumptions of independence in their measurement would be violated if more than one student from the same household were included in the study. If some parents were required to complete more than one rather lengthy interview at each of seven waves of data collection, it would also increase the likelihood of respondent fatigue and attrition in the parent sample. The loss of one parent respondent with two offspring in the student sample would also increase the likelihood of losing both students members of the sample.

(4) *Students whose ages lay outside those defined for the age/grade cohort*: It was found in early stages of data collection that 123 of the students selected from the seventh and eighth grades were aged 15 or older. The inclusion of these older students in the sample was inconsistent with the cohort design underlying the study. Students who were 15 or older as of December 31, 1986 were therefore dropped from the sample. This cutoff date was used since it defined the "normal" age of students enrolled in the eighth grade for Rochester schools. Males who were ineligible because of age were replaced by new cases drawn with certainty from census tracts with the highest arrest rates. Beginning with the census tracts with the highest arrest rates, and moving down the list of eligible males in each census tract ranked by arrest rate, 103 males were added to the sample as replacements for those ineligible because of age. Female replacement cases were selected to match excluded cases on census tract residence and grade in school.

(5) *Nonminority females*: Disproportionate stratification of the sample by both census tract and sex (described below) yielded an overrepresentation of minority youths. Of these, the overwhelming majority (over 200) of the 250 targeted females were either black or Hispanic. Tract-by-sex stratification resulted in a very small subsample of white females. This small number was considered inadequate for the planned separate subgroup study, especially in view of the typically low rates of delinquency associated with young, white females. Nonminority females were therefore excluded from the pool of potential subjects for Wave 1. To address this imbalance, replacement cases in subsequent waves added white females.

Using these five standards for subject participation in the study, 641 of the 4,013 seventh and eighth grade students listed on the school's roster were defined as ineligible for the sampling frame. An eligible pool of 3,372 students remained from which to draw a sample of approximately 1,000 subjects.

D. Disproportionate Stratified Sampling

The sampling plan was not devised to achieve a proportionately representative sample since designs that have done so (e.g., Elliott et al., 1983) tend to yield low base rates of serious, repeat offending. The objective of the design was to oversample subjects likely to become involved in frequent and serious delinquency over the life of the longitudinal study. Students from high crime areas were therefore oversampled, with the premise that higher crime rates and delinquent opportunities are rather localized and that subjects residing in high crime rate areas are at a greater risk for offending.

To identify high crime areas, each census tract in the city of Rochester was assigned a resident arrest rate (RAR) based on police data for 1986. It should be noted that sampling was not proportionate to the *crime* rate (the number of crimes committed per population), but to the *arrest* rate (the number of residents in each tract who were arrested per population). The tracts were then rank ordered according to RAR level.

The probability of an individual's selection into the sample was proportionate to the RAR of his or her residential area. For this purpose, the last known address of each eligible student, as recorded in school records, was used to place each student in a census tract of residence. Subjects were then drawn with a probability of selection proportionate to the rate of known offenders living in their respective tracts of residence.

The 1980 census defined 91 census tracts in the city of Rochester. Of these, three tracts were excluded from consideration. Two tracts were excluded because they were nonresidential. Tract #89 consists of the local airport and the surrounding area; tract #3802 encompasses the University of Rochester. Rochester police arrest data for those tracts were not available. A third tract (#9403) was excluded because of an absence of age-appropriate residents for the study.

The latest census data indicated that only one child in that tract would fall within the age range for the study; and none of the students on the Rochester seventh and eighth grade rosters actually resided in that tract. Thus, none of these tracts contained residences of eligible students for the study, so no subjects were lost from the sampling frame when these tracts were excluded.

The remaining 88 tracts were rank ordered according to their 1986 arrest rates. The following sampling fraction was used to select cases from each census tract:

SAMPLING FRACTION = RAR #1/RAR #2, where

RAR #1 = the resident arrest rate for each census tract;

RAR #2 = the resident arrest rate for all tracts.

The probability of selection for each tract is the expected sampling fraction. Arrest rates for the census tracts and the calculation of sampling fractions for each census tract are listed in Appendix B. The sampling fractions range from three percent in the tracts with arrest rates of approximately seven per 100 residents, to near-zero in tracts with arrest rates near zero per 100 residents. The expected sampling fraction generated the number of cases to be selected from each tract.

The selection of cases according to the sampling fraction for each census tract was adjusted by two factors. The first was to allow the subsample of 127 students selected for the pretest phase of the study to be included in the final panel. This was accomplished by first allocating all of the pretest sample to their census tracts at the time of the Wave 1 sample. Pretest sample members, who were selected with virtually the same strategies as members of the Wave 1 sample (see Appendix C), are therefore included in the calculation of the proportions of the sample selected from each census tract according to the sampling fraction for that tract.

The second adjustment stems from the goal of overrepresenting in the panel youths at high risk for serious delinquency. Because of this, students residing in the highest arrest rate tracts were selected with certainty. That is, *all* students from these tracts, rather than a sample of them, are included in the panel provided they and their guardians agreed to participate. (Refusal rates are discussed below.) The cut-off for defining these "highest" arrest rate tracts was

determined by examining the distribution of resident arrest rates; there is an observable break in the distribution of tracts above and below the rate of three arrests per 100 residents. Students in tracts above that figure were selected with "certainty" while students in tracts below that were selected according to the tract's sampling fraction. In the lower arrest rate tracts, if the sampling fraction was three percent, then three percent of the sample (30 cases) were drawn from that tract.

In selecting the appropriate number from each tract, fifty percent were drawn from the seventh grade and fifty percent from the eighth grade. The sample was disproportionately stratified by sex at a ratio of three males to every female: from each tract, 75 percent of the cases were male and 25 percent female. This was done since the rate of serious, chronic offending among females is typically quite low (Tracy and Figlio, undated: 6).

E. Disproportionate Sampling and Weighting

Disproportionate stratified sampling implies that students selected into the sample had different nonzero probabilities of selection according to the arrest rates of the census tract where they resided, and according to their gender. Since controlled probability sampling procedures were used, however, the sample can be weighted to be representative of the sampling frame from which it was drawn. For example, the probability of an individual being selected was proportionate to the arrest rate of persons living in his or her tract of residence. The "true" probability of a youth living in a particular census tract and the sampling fraction for that tract are known. It is also known that males were oversampled at a three-to-one ratio to females. With this information, appropriate weights equal to the inverse of the probability of selection were calculated and assigned to approximate a random sample representative of the sampling frame or population of interest. Appendix D presents the weights assigned for this purpose, within five categories of census tracts ranked by RAR's; and for analyses using the full sample or a particular subsample of respondents.

III. SUBJECT PARTICIPATION

A. Assuring Respondent Cooperation

Once the target sample was identified, letters from the Principal Investigator were sent to all selected children and their parents (see Appendix E: Supporting Letters and Forms). The letters provided a general description of the project and an appeal for cooperation with the research effort. These letters were followed by interviewers' visits to the home to describe the project in greater detail, answer questions, explain parent payments and student incentives, and assure informed consent from the parents agreeing to participate.

B. Incentives to Participate

To encourage participation and continuance in the study, cash payments of ten dollars per interview were made to the parents of subjects in the study. A choice from a variety of "prizes" worth approximately ten dollars each, carefully selected as age-appropriate and attractive to the student sample, were provided to students as a reward for each interview completed in Wave 1. A periodical RYDS newsletter was also mailed to study subjects to remind them of the importance of their continued participation, and to maintain their interest in and enthusiasm for the research project (see Appendix E).

C. Participation

A sample of approximately one thousand was considered necessary to support the planned causal analyses across different demographic subsamples. A beginning sample of 1,334 was selected as potential subjects, based on an estimated nonparticipation rate of approximately twenty-five percent (Elliott et al., 1983). Of the 1,334 cases, parents could not be located in 42 cases. An additional 248 parents were contacted but refused to participate. Forty-five parents neither refused nor consented within the time available for defining the Wave 1 sample, and twelve cases were not fielded before the deadline for drawing Wave 1 cases.

Refusal rates, calculated as the number of refusals divided by the number of completed cases plus the number of refusals, were examined by racial-ethnic group, by grade, and by sex (Table 1). For the total sampling frame the refusal rate was 20 percent. Among blacks, 20

percent refused; among whites the refusal rate was 24 percent; and the refusal rate for Hispanics was 14 percent. The refusal rate for females was 18 percent compared with 21 percent for males. Nineteen percent of the seventh graders and 21 percent of the eight graders chose not to participate. Overall, refusal rates did not differ greatly for different demographic groups, nor did they differ in a patterned way that was disruptive to the study's sampling design.

The subtraction of the non-participating cases from the targeted sample resulted in a drawn sample of 987 student households, or 74 percent of the targeted sample. This number constitutes the total base panel for the longitudinal study. Of this number, Wave 1 interviews were completed with 956 students (97 percent of the base panel) and with 980 parents (99 percent of the base panel). There were 949 "matched" household cases, in which both the parent and the child were interviewed in Wave 1.

In three percent of the base panel cases (31 cases), Wave 1 parent interviews were conducted and parental consent for participation granted, but student interviews were not completed by the ending date for data collection. These students could not be contacted during the school year and were difficult to locate for home interviews. They were retained in the base panel and priority was given to locating and interviewing them in Wave 2 and to retaining them in the panel.

IV. THE WAVE 1 SAMPLE

A. Description of the Wave 1 Sample

Table 2 describes the distribution of the sample by sex, race, grade, and age. The first column in Table 2 is based on information about the sample gathered from school records before the interviews were conducted, and represents the total base sample of 987. Overall, 62 percent of the sample are black, 14 percent are white, 16 percent are Hispanic, and 8 percent are of other races. Of the females, 85 percent are black and 15 percent are Hispanic. Of the males, 63 percent are black, 19 percent are white, and 17 percent Hispanic. These proportions are quite

Table 1.   Participation and Refusal Rates, Wave 1: Total Sample and Within Groups Differentiated by Sex, Grade in School, and Race\Ethnicity

	% Participation	% Refusal Rate
Total Targeted Sample	80	20
Sex:		
Males	79	21
Females	82	18
Grade in School:		
7th Grade	81	19
8th Grade	79	21
Race/Ethnicity:		
White	76	24
Black	80	20
Hispanic	86	14

Table 2. Demographic Characteristics, Wave 1 Sample

	Base Panel (N=987)		Student Interviews (N=956)	
	N	(%)	N	(%)
SEX:				
Male	731	(74.1)	708	(74.1)
Female	256	(25.9)	248	(25.9)
RACE/ETHNICITY:				
Black	610	(61.9)	589	(61.7)
Hispanic	156	(15.8)	152	(15.9)
White	135	(13.7)	129	(13.5)
Other	85	(8.6)	85	(8.9)
GRADE IN SCHOOL:				
Seventh	553	(56.3)	535	(56.0)
Eighth	431	(43.7)	421	(44.0)
CENSUS TRACT SECTION:				
1	328	(33.2)	316	(33.1)
2	319	(32.3)	307	(32.1)
3	176	(17.8)	173	(18.1)
4	97	(9.8)	94	(9.8)
5	51	(5.2)	50	(5.2)
6	86	(1.6)	16	(1.7)

Note: Lower census tract section indicates higher RAR's (resident arrest rates).

close to what was expected given the population characteristics of the Rochester Schools and the decision to oversample high-risk youth.

The second column in Table 2 is based on student interview responses (N=956). The smaller sample of students who were interviewed have demographic characteristics that are virtually identical to the base panel sample which includes 31 students not contacted in Wave 1.

Table 3 presents social background information provided by parent respondents in cases for which both parent and student Wave 1 data were available. The distribution of these variables indicates that this sample, selected to overrepresent youth from high crime areas, also overrepresents youth from lower-class households. Fifteen percent of the households were classified as poverty level. Almost half of the households (42 percent) received welfare payments at the time of the interviews, and fully 34 percent of the principal wage earners were unemployed. Almost half (46 percent) of the principal wage earners had fewer than twelve years of formal schooling (Farnworth et al., 1990).

B. Delinquency in the Wave 1 Sample

The prevalence of delinquency in the Wave 1 sample of students interviewed (N=956) is depicted in Figure 1. Table 4 describes the component items in each of the delinquency measures charted in Figure 1. More than three-quarters (77 percent) of the sample, even at this early age, self-reported some type of delinquency. As might be expected in a youthful sample, the most common types were the less serious ones. Fifty-five percent reported status offenses. Almost half the sample, however, reported involvement in some kind of offense against persons (47 percent) or against property (42 percent). The prevalence of illegal drug use is relatively low; only 12 percent reported that they had ever used drugs, and most of this was marijuana use. One-quarter reported ever-involvement in street crimes, which selectively crosscuts personal and property offenses and also includes such behaviors as drug sales. Seventeen percent engaged in gang fights. Fourteen percent were involved in "white collar" delinquency reflecting economic motivation and nonviolence (Farnworth et al., 1990).

Table 3. Social Background Characteristics of the Wave 1 Sample (N=949)

I. Household of Residence

Variable	Categories	N	(%)
Income	1=<average (<$4545)	72	(7.6)
	2=average ($4546-25672)	745	(78.5)
	3=>average (>$25642)	132	(13.9)
Poverty Level	0=above poverty level	807	(85.0)
	1=poverty level	142	(15.0)
Welfare Status	0=household on welfare	402	(42.4)
	1=no welfare	547	(57.6)

II. Principal Wage Earner in Household of Residence

Variable	Categories	N	(%)
Unemployment	0=employed	628	(66.2)
	1=unemployed	321	(33.8)
Occupational SEI (employed only)	1=low average (<28)	428	(68.1)
	2=average (28-42)	109	(17.4)
	3=>average (>42)	91	(14.5)
Years of Schooling	6 or <6	43	(4.3)
	7	22	(2.3)
	8	37	(3.9)
	9	75	(7.9)
	10	118	(12.4)
	11	145	(15.3)
	12	329	(34.7)
	13 or >13	180	(19.0)

Note: N for Wave 1 sample for whom both student and parent data are available.

Fig. 1 Prevalence Rates,
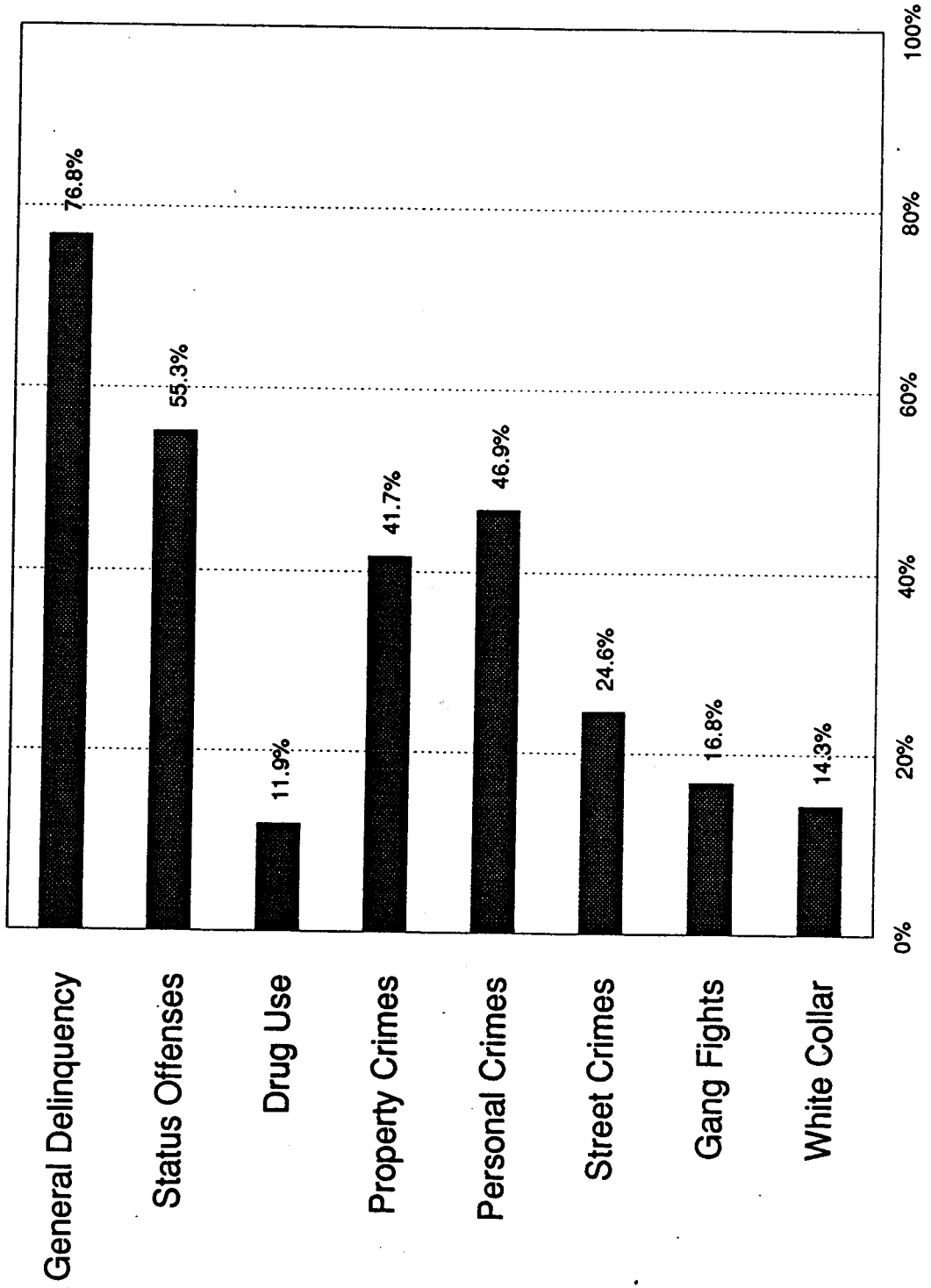Self-Reported Delinquency (by Type)

Table 4.   Delinquency Scales: Items Included in Composite Measures in Figure 1

GENERAL DELINQUENCY
    Runaway
    Skipping class
    Lying about age for illegal purposes
    Hitchhiking with strangers
    Carrying a hidden weapon
    Rowdiness in public
    Begging from strangers
    Public drunkenness
    Property damage/destruction
    Arson
    Breaking and entering
    Theft $5 or less
    Theft $5-50
    Theft $50-100
    Theft >$100
    Shoplifting
    Purse snatching, picking pockets
    Theft from cars
    Buying/selling stolen goods
    Joyriding
    Motor vehicle theft
    Forgery/fake money
    Illegal credit/bankcard use
    Fraud
    Attack with weapon
    Other assault
    Gang fights
    Throwing objects at people
    Robbery
    Marijuana sale
    Other drug sale

"WHITE COLLAR" DELINQUENCY
    Forgery
    Illegal credit/bankcard use
    Fraud

STATUS OFFENSES
    Runaway
    Skipping class
    Lying about age for illegal purposes
    Underage drinking: beer/wine
    Underage drinking: hard liquor

DRUG USE
    Marijuana
    Other drugs

CRIMES AGAINST PROPERTY
    Property damage/destruction
    Arson
    Breaking and entering
    Theft $5 or less
    Theft $5-50
    Theft $50-100
    Theft >$100
    Shoplifting
    Buying/selling stolen goods
    Joyriding
    Motor vehicle theft
    Fraud

CRIMES AGAINST PERSONS
    Attack with weapon
    Other assault
    Gang fights
    Throwing objects at people
    Robbery

STREET CRIMES
    Purse snatching/picking pockets
    Theft from car
    Buying/selling stolen goods
    Motor vehicle theft
    Robbery
    Assault with weapon
    Gang fights
    Sale of marijuana
    Sale of other drugs
    Breaking and entering

Figure 2 presents prevalence rates of official delinquency in the sample. As of 1989, 28 percent of the panel members had experienced at least one contact with the Rochester police. Males had an official prevalence rate of 30 percent compared with 22 percent for females. Black sample members had the highest rate of contact (35 percent), Hispanics the lowest rate (18 percent), and the official prevalence rate for whites was 27 percent. The average number of official contacts among official offenders was 2.11.

C. Representivity of the Sample

Consistent with the sampling plan, males are overrepresented at a ratio of three-to-one to females: 74 percent of sample are males, and 26 percent are females (Table 2). Seventh grade students are overrepresented (56 percent) compared with eighth graders (44 percent). The strategy of oversampling students from high crime areas of the city resulted in a sample that overrepresents minorities relative to white respondents. Black students comprise 62 percent of the sample, 16 percent are Hispanic, and 14 percent are white.

One way to examine the fit between the expected and observed samples is to compare the number of cases that would be expected with the number actually obtained in census tracts of varying arrest rates. These data are presented in Table 5. Starting with the high arrest rate census tracts (those tracts in which certainty sampling was used), the difference between the expected and obtained number of cases is quite small. Overall, the sampling plan called for 72.6% of the cases to be drawn from these tracts and 72.2% were obtained. The comparisons differ slightly for male and female subjects. For males the expected percentage was 72.5% and 70.6% was obtained. For females, the expected percentage was 72.8% and 77.1% was obtained. The data for the medium and low arrest rate tracts also indicate that the obtained sample matches the expected sample quite well. This holds for the total sample and for both males and females.

In sum, this analysis suggests that the sampling plan designed for the RYDS was successful in meeting its objectives. It has provided a panel that overrepresents adolescents who are at high risk for delinquency while at the same time yielding a sample that can be generalized

Fig.2 Prevalence Rates, Official
Delinquency (by Sex and Race)
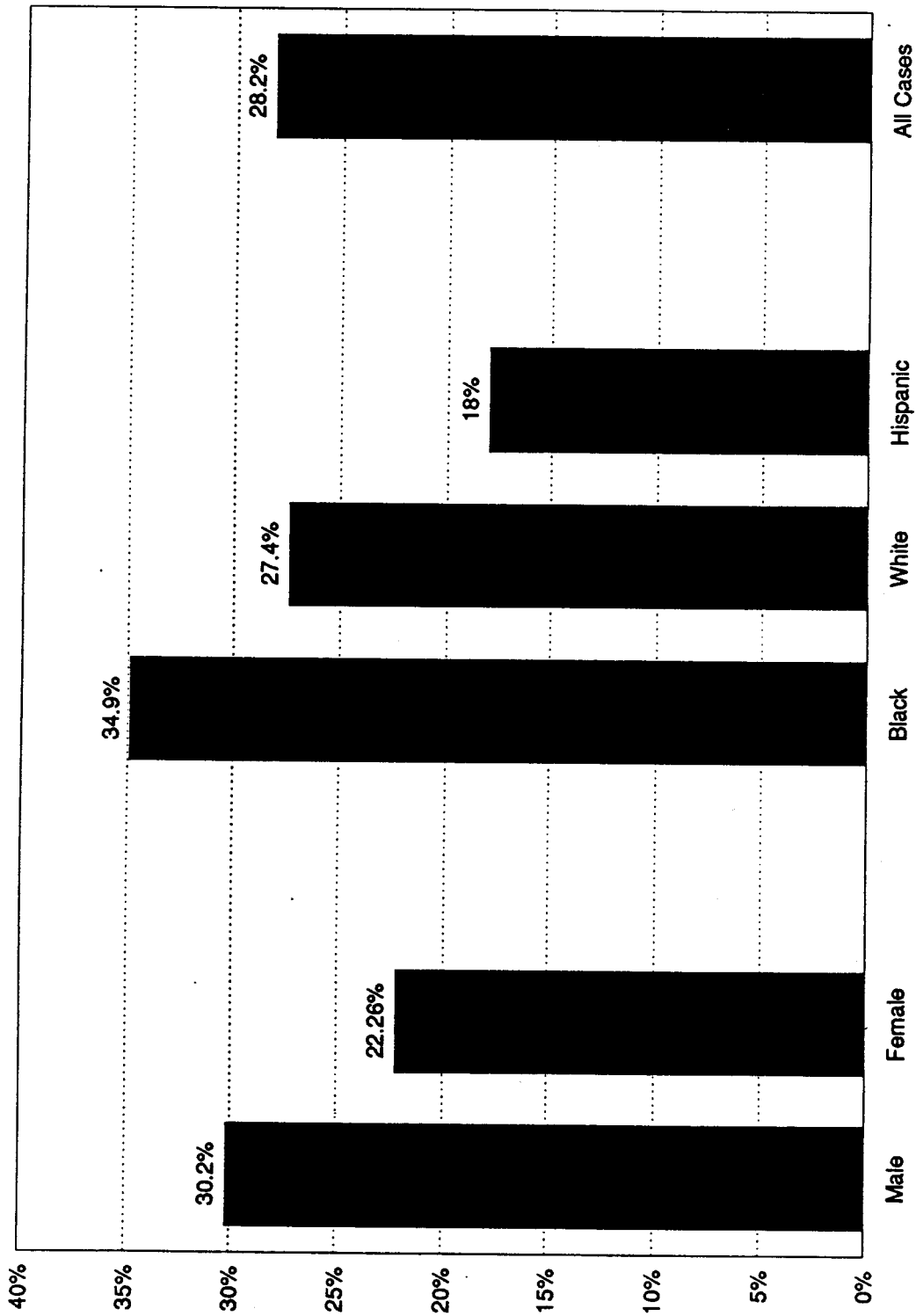
* Mean # of offenses per offender=2.11

Table 5. Expected and Obtained Samples from High, Low, and Medium Resident Arrest Rate Census Tracts

	Males		Females		Total	
	Expected	Obtained	Expected	Obtained	Expected	Obtained
High RAR	72.5%	70.6%	72.8%	77.1%	72.6%	72.2%
Medium RAR	20.9	22.6	20.7	17.3	20.1	21.3
Low RAR	6.6	6.7	6.5	5.5	6.6	6.5

Note: High RAR = census tract resident arrest rate of >3 per 100 population;
Medium RAR = rate <3 and >1;
Low RAR = rate <1 per 100 population.

to the school population. The obtained sample matches the expected one quite closely and adolescents from high arrest rate areas are represented as planned.

V. REFERENCES AND RELATED RYDS REPORTS

A. References

Center for Government Research, Inc., 1983. "Community Profiles--Community Analysis Model: Rochester Region." Unpublished Report.

Elliott, D. S., S.S. Ageton, D. H. Huizinga, B. A. Knowles and R. J. Canter, 1983. The Prevalence and Incidence of Delinquent Behavior 1976-1980. National Youth Survey Report No. 26. Boulder, CO: Behavioral Research Institute.

Farnworth, Margaret, Terence P. Thornberry, Alan J. Lizotte, and Marvin D. Krohn, 1989. "Social background and the early onset of delinquency: Exploring the utility of various indicators of social class background." Paper presented at the Annual Meetings of the American Society of Criminology (Reno).

Federal Bureau of Investigation, 1984. Uniform Crime Reports. Washington, D.C.: U.S. Department of Justice.

New York State Department of Education, 1985. School Summary, Information Center on Education Ethnic Census. Albany, New York.

Thornberry, Terence P., Margaret Farnworth and Alan J. Lizotte, 1986. A Panel Study of a Reciprocal Causal Model of Delinquency. Proposal submitted to the Office of Juvenile Justice and Delinquency Prevention.

Tracy, Paul E. and Robert M. Figlio (undated). "Chronic recidivism in the 1958 birth cohort." Center for Studies in Criminology and Criminal Law, University of Pennsylvania.

B. RYDS Reports and Proposals

Cass, Elizabeth, 1988. "An Account of the Sampling Process for the RYDS," Interim Report, Rochester Youth Development Study (August).

Chard, Deborah J., 1990. "An Update of the Documentation of the Rochester Youth Development Study Sampling Process," Interim Report (March).

Thornberry, Terence P., 1990. "Panel Study of a Reciprocal Causal Model of Delinquency," Proposal for Year Five Funding, submitted to OJJDP, U.S. Dept. of Justice (May).

Thornberry, Terence P., Marvin D. Krohn, and Alan J. Lizotte, 1990. "Delinquency and Criminal Careers: Phase II of the Rochester Youth Development Study," submitted to OJJDP, U. S. Dept. of Justice.

Appendix A

Procedures for Replacing Cases

PROCEDURES FOR REPLACING CASES

Criteria and procedures for replacing cases dropped or lost from the sampling pool were established during the pretest stage of the project (see Appendix C).

<u>Reasons for Replacement in the Pretest Stage</u>

A preliminary sampling frame of 220 cases was defined for the pretest. Of this number, 32 cases were dropped because a brother or sister of that student was already included in the preliminary sampling frame. Of the remaining 188 cases, 61 were eliminated for one of the following reasons:

(1) Forty of the parents refused to participate when contacted by the interviewing staff;

(2) Twelve of the students' families could not be located at the most recent address available; and

(3) Interviewers could not determine in four cases if the targeted subject and his or her family lived at the most recent available address.

An additional five cases were dropped for miscellaneous reasons.

<u>Replacement</u>

To maintain sample size, new cases were substituted for those that were dropped. New cases were matched as closely as possible to lost cases on gender, grade in school, and census tract to retain the desired characteristics of the sample. The strategy of oversampling in certain high-arrest census tracts led in some instances, however, to a depletion of eligible subjects in those tracts. In these instances, comparable census tracts were identified as the best substitutes for tracts that could no longer provide suitable subjects for the study. Comparability was defined in the following order:

(1) The first choice of a substitute census tract was a census tract in the same "neighborhood" as the original tract. Definitions of neighborhoods were provided by the Center for Government Research in Rochester.

(2) If (1) above was also depleted of appropriate replacement cases, a census tract in a different neighborhood but adjacent to the original tract, and with a similar arrest rate as the original tract, was substituted.

(3) If neither (1) nor (2) were possible, then replacement cases were drawn from a census tract not adjacent to the original one, but with a similar arrest rate.

Modifications in Wave 1

In Wave 1, practical and theoretical considerations led to the following modifications to the replacement plan developed in the Pretest:

(1) Advance replacements: In order for the field office to keep a readily-available pool of case numbers on hand for replacement purposes as the need arose, 105 "advance replacements" were defined. These cases were drawn from census tracts with an arrest rate above the median for all census tracts.

(2) Sampling with certainty: Beginning in March of 1988, the replacement pool for potential male sample members was drawn with certainty from census tracts with the highest arrest rate in which there were any males suitable for the sampling frame.

As of April 15, 1988, all eligible females from the two highest arrest rate tracts (N=41) were assigned to the replacement pool. On April 29, 1988, all eligible students still available in census tracts with the highest arrest rates were added to the reserve pool of replacement cases (N=50) in a male-to-female ratio of three to one.

As of May 14, 1988, male cases for replacement purposes were drawn with certainty from census tracts with a minimum arrest rate of three per one thousand population. Female replacement cases were drawn from tracts with a minimum arrest rate of 7.5 per one thousand.

Appendix B

Arrest Rates and Sampling Fractions

Table B.1. Drawing the Pretest Sample of 150: Computation of Sampling Fractions

Census Tract	Arrest Rate	Sampling Fraction	Unrounded		Rounded	
			Boys	Girls	Boys	Girls
1400	7.87	.034	25.75	8.58	26	8
6500	7.56	.033	25.72	8.57	25	8
9301	7.44	.032	24.34	8.11	24	8
5900	7.32	.032	23.95	7.98	24	8
700	6.86	.032	22.45	7.48	22	8
9100	6.84	.029	22.35	7.45	22	8
6400	6.75	.029	22.12	7.37	22	7
4300	6.72	.029	21.97	7.32	22	7
2700	6.63	.029	21.87	7.22	22	7
5200	5.91	.025	19.27	6.42	19	6
9601	5.81	.024	18.97	6.32	19	6
4100	5.47	.023	17.85	5.95	18	6
1600	5.26	.022	17.17	5.72	17	6
1500	5.15	.021	16.30	5.60	17	6
6600	4.97	.020	16.20	5.40	16	5
5300	4.75	.020	15.52	5.17	15	5
5500	4.66	.019	15.22	5.07	15	5
5600	4.60	.019	15.00	5.00	15	5
3200	4.60	.019	15.00	5.00	15	5
5700	4.51	.019	14.70	4.90	15	5
9602	4.50	.019	14.70	4.90	15	5
5000	4.17	.018	13.57	4.52	14	5
5100	3.84	.017	12.52	4.17	13	4

(Table B.1 -- continued)

Census Tract	Arrest Rate	Sampling Fraction	Unrounded		Rounded	
			Boys	Girls	Boys	Girls
9604	3.80	.016	12.37	4.12	12	4
9402	3.66	.016	11.92	3.97	12	4
6900	3.63	.016	11.85	3.95	12	4
9200	3.51	.015	11.47	3.82	11	4
6300	3.47	.015	11.32	3.77	11	4
1700	3.40	.015	11.10	3.70	11	4
6700	3.16	.014	10.27	3.42	10	3
1300	3.12	.013	10.20	3.40	10	3
7500	3.08	.013	10.05	3.35	10	3
4000	3.05	.013	9.97	3.32	10	3
4900	2.77	.012	9.00	3.00	9	3
9403	2.73	.012	8.85	2.95	9	3
7900	2.72	.012	8.40	2.80	8	3
2300	2.58	.011	8.32	2.77	8	3
3900	2.56	.011	8.25	2.75	8	3
9401	2.54	.011	7.72	2.57	8	3
2400	2.37	.010	7.27	2.42	7	2
9603	2.23	.009	7.05	2.35	7	2
3000	2.17	.009	7.05	2.35	7	2
8000	2.16	.009	6.82	2.27	7	2
9500	2.10	.009	6.75	2.25	7	2
9302	2.07	.009	6.37	2.12	6	2
5800	1.95	.008	5.55	1.85	6	2
3400	1.71	.007	5.55	1.85	6	2
4800	1.71	.007	5.25	1.75	5	2

(Table B.1 -- continued)

Census Tract	Arrest Rate	Sampling Fraction	Unrounded		Rounded	
			Boys	Girls	Boys	Girls
4602	1.62	.007	5.25	1.75	5	2
7000	1.52	.007	4.50	1.50	5	2
6200	1.47	.006	4.80	1.60	5	2
1000	1.39	.006	4.50	1.50	5	2
200	1.28	.006	4.12	1.37	4	1
8400	1.27	.005	4.12	1.37	4	1
3600	1.25	.005	4.05	1.35	4	1
8702	1.24	.005	4.05	1.35	4	1
4702	1.23	.005	3.97	1.32	4	1
1900	1.09	.005	3.52	1.17	4	1
6800	1.04	.004	3.37	1.12	3	1
3300	1.03	.004	3.30	1.10	3	1
2000	1.02	.004	3.30	1.10	3	1
8701	.99	.004	3.22	1.07	3	1
3700	.93	.004	3.00	1.00	3	1
8200	.93	.004	3.00	1.00	3	1
7100	.89	.004	2.85	.95	3	1
2200	.87	.004	2.77	.92	3	1
5400	.86	.004	2.77	.92	3	1
8800	.79	.003	2.55	.85	3	1
2100	.73	.003	2.32	.77	2	1
8100	.73	.003	2.32	.77	2	1
9000	.72	.003	2.32	.77	2	1
6000	.71	.003	2.25	.75	2	1
8500	.69	.003	2.25	.75	2	1

(Table B.1 -- continued)

Census Tract	Arrest Rate	Sampling Fraction	Unrounded		Rounded	
			Boys	Girls	Boys	Girls
2900	.53	.002	1.72	.57	2	1
4701	.46	.002	1.50	.50	2	1
3804	.46	.002	1.50	.50	2	1
3801	.46	.002	1.50	.50	2	1
6100	.45	.002	1.42	.47	1	0
3500	.40	.002	1.27	.42	1	0
7700	.37	.002	1.20	.40	1	0
7600	.36	.002	1.12	.37	1	0
3803	.33	.001	1.05	.35	1	0
1800	.29	.001	.90	.30	1	0
8300	.28	.001	.90	.30	1	0
7801	.27	.001	.82	.27	1	0
8600	.24	.001	.75	.25	1	0
7802	.18	.001	.52	.17	1	0
3100	.12	.001	.37	.12	0	0
4601	.00	.000	.00	.00	0	0
10500	.00	.000	.00	.00	0	0
TOTAL:			747.5	249.2	746	246

Table B.2.   Drawing the Pretest Sample of 150: Computation of Sampling Fractions

Census Tract	Arrest Rate	Sampling Fraction	Unrounded		Rounded	
			Boys	Girls	Boys	Girls
1400	7.87	.034	3.842	1.258 (+1)	4	1
6500	7.56	.033	3.729	1.221	4	1
9301	7.44	.032	3.616	1.184	4	1
5900	7.32	.032	3.616	1.184	3	2
700	6.86	.032	3.277	1.073	3	1
9100	6.84	.029	3.277	1.073	3	1
6400	6.75	.029	3.277	1.073	3	1
4300	6.72	.029	3.277	1.073	3	1
2700	6.63	.029	3.277	1.073	3	1
5200	5.91	.025	2.825	.925	3	1
9601	5.81	.024	2.825	.925	3	1
4100	5.47	.023	2.712	.888	3	1
1600	5.26	.022	2.599	.851	3	0
1500	5.15	.021	2.486	.814	3	0 (+1)
6600	4.97	.020	2.373	.777	2	1
5300	4.75	.020	2.26	.74	2	1
5500	4.66	.019	2.26	.703	2	1
5600	4.60	.019	2.147	.703	3	0
3200	4.60	.019	2.147	.703	3	0
5700	4.51	.019	2.147	.703	3	0
9602	4.50	.019	2.147	.703	2	1
5000	4.17	.018	2.034	.666	2	1 (+1)
5100	3.84	.017	1.921	.629	2	0

(Table B.2 -- continued)

Census Tract	Arrest Rate	Sampling Fraction	Unrounded		Rounded		
			Boys	Girls	Boys	Girls	
9604	3.80	.016	1.808	.592	1	1	
9402	3.66	.016		.592	2	0	
6900	3.63	.016	1.808	.592	1	1	(+1)
9200	3.51	.015	1.69	.555	2	0	
6300	3.47	.015	1.69	.555	1	1	
1700	3.40	.015	1.69	.555	1	1	
6700	3.16	.014	1.582	.518	1	1	
1300	3.12	.013	1.469	.481	1	1	
7500	3.08	.013	1.469	.481	2	0	
4000	3.05	.013	1.469	.481	1	1	
4900	2.77	.012	1.356	.444	2	0	
9403	2.73	.012	1.356	.444	2	0	
7900	2.72	.012	1.356	.444	1	1	
2300	2.58	.011	1.243	.407	2	0	
3900	2.56	.011	1.243	.407	1	1	
9401	2.54	.011	1.243	.407	2	0	
2400	2.37	.010	1.13	.37	2	0	
9603	2.23	.009	1.017	.333	1	0	
3000	2.17	.009	1.017	.333	1	0	
8000	2.16	.009	1.017	.333	1	1	
9500	2.10	.009	1.017	.333	1	0	
9302	2.07	.009	1.017	.333	1	0	
5800	1.95	.008	.904	.296	0	1	
3400	1.71	.007	.791	.259	1	0	
4800	1.71	.007	.791	.259	1	0	

(Table B.2 -- continued)

Census Tract	Arrest Rate	Sampling Fraction	Unrounded		Rounded	
			Boys	Girls	Boys	Girls
4602	1.62	.007	.791	.259	1	0
7000	1.52	.007	.791	.222	0	1
6200	1.47	.006	.678	.222	1	0
1000	1.39	.006	.678	.222	1	0
200	1.28	.006	.678	.222	0	1
8400	1.27	.005	.565	.185	1	0
3600	1.25	.005	.565	.185	1	0
8702	1.24	.005	.565	.185	1	0
4702	1.23	.005	.565	.185	1	0
1900	1.09	.005	.565	.185	1	0
6800	1.04	.004	.452	.148	1	0
3300	1.03	.004	.452	.148	1	0
2000	1.02	.004	.452	.148	1	0
8701	.99	.004	.452	.148	1	0
3700	.93	.004	.452	.148	0	1
8200	.93	.004	.452	.148	1	0
7100	.89	.004	.452	.148	1	0
2200	.87	.004	.452	.148	1	0
5400	.86	.004	.452	.148	1	0
8800	.79	.003	.339	.111	0	1
2100	.73	.003	.339	.111	0	0
8100	.73	.003	.339	.111	1	0
9000	.72	.003	.339	.111	0	0
6000	.71	.003	.339	.111	1	0
8500	.69	.003	.339	.111	0	0

(Table B.2 -- continued)

Census Tract	Arrest Rate	Sampling Fraction	Unrounded Boys	Girls	Rounded Boys	Girls
2900	.53	.002	.226	.074	0	0
4701	.46	.002	.226	.074	0	0
3804	.46	.002	.226	.074	1	0
3801	.46	.002	.226	.074	0	0
6100	.45	.002	.226	.074	0	0
3500	.40	.002	.226	.074	1	0
7700	.37	.002	.226	.074	0	0
7600	.36	.002	.226	.074	0	0
3803	.33	.001	.113	.037	1	0
1800	.29	.001	.113	.037	0	0
8300	.28	.001	.113	.037	0	0
7801	.27	.001	.113	.037	0	0
8600	.24	.001	.113	.037	0	0
7802	.18	.001	.113	.037	0	0
3100	.12	.001	.113	.037	0	0
4601	.00	.000	.000	.000	0	0
10500	.00	.000	.000	.000	0	0
TOTAL:					119	33

Table B.3.   Sampling Fractions Expressed as a Percentage of Sample for Each Census Tract

Census Tract	Arrest Rate Per 100	Expected (ESF)	Obtained (OSF)	Difference Between OSF-ESF
1400	7.87	3.40%	4.02%	.62%
6500	7.56	3.30	4.22	.92
9301	7.44	3.20	2.21	-.99
5900	7.32	3.20	1.61	-1.59
700	6.35	3.00	2.21	-.79
9100	6.84	3.00	1.10	-1.90
6400	6.75	2.90	3.41	.51
4300	6.72	2.90	.70	-2.20
2700	6.63	2.90	2.41	-.49
5200	5.91	2.50	3.61	1.11
9601	5.81	2.50	2.31	-.19
4100	5.47	2.40	1.71	-.69
1600	5.26	2.30	1.20	-1.10
1500	5.15	2.30	1.81	-.49
6600	4.97	2.10	1.81	-.29
5300	4.75	2.00	2.91	.91
5500	4.66	2.00	3.01	1.01
5600	4.60	2.00	2.91	.91
3200	4.60	2.00	1.00	-1.00
5700	4.51	2.00	2.41	.41
9602	4.50	2.00	2.31	.31
5000	4.17	1.90	4.62	2.72
5100	3.84	1.70	2.31	.61

(Table B.3 -- continued)

Census Tract	Arrest Rate Per 100	Expected (ESF)	Obtained (OSF)	Difference Between OSF-ESF
9604	3.80	1.60	2.61	1.01
9402	3.66	1.60	.10	-1.50
6900	3.63	1.60	2.31	.71
9200	3.51	1.50	.50	-1.00
6300	3.47	1.50	2.81	1.31
1700	3.40	1.50	.80	-.70
6700	3.16	1.30	2.51	1.21
1300	3.12	1.30	1.81	.51
7500	3.08	1.30	2.31	1.01
4000	3.05	1.30	.60	-.70
4900	2.77	1.20	2.31	1.11
7900	2.72	1.20	.80	-.40
2300	2.58	1.10	1.61	.51
3900	2.56	1.10	1.41	.31
9401	2.54	1.10	.00	-1.10
2400	2.37	1.10	2.01	.91
9603	2.23	.90	.80	-.10
3000	2.17	.90	.50	-.40
8000	2.16	.90	1.31	.41
9500	2.10	.90	1.41	.51
9302	2.07	.90	.00	-.90
5800	1.95	.80	.70	-.10
3400	1.71	.80	1.20	.40
4800	1.71	.80	.80	.00

(Table B.3 -- continued)

Census Tract	Arrest Rate Per 100	Expected (ESF)	Obtained (OSF)	Difference Between OSF-ESF
4602	1.62	.70	.60	-.10
7000	1.52	.70	.50	-.20
6200	1.47	.70	.60	-.10
1000	1.39	.70	.40	-.30
200	1.28	.50	.00	-.50
8400	1.27	.50	.70	.20
3600	1.25	.50	1.00	.50
8702	1.24	.50	.80	.30
4702	1.23	.50	.70	.20
1900	1.09	.50	.40	-.10
6800	1.04	.40	.30	-.10
3300	1.03	.40	.10	-.30
2000	1.02	.40	.30	-.10
8701	.99	.40	.50	.10
3700	.93	.40	.60	.20
8200	.93	.40	.30	-.10
7100	.89	.40	.50	.10
2200	.87	.40	.40	.00
5400	.86	.40	.40	.00
8800	.79	.40	.40	.00
2100	.73	.30	.30	.00
8100	.73	.30	.20	-.10
9000	.72	.30	.00	-.30
6000	.71	.30	.40	.10
8500	.69	.30	.40	.10

(Table B.3 -- continued)

Census Tract	Arrest Rate Per 100	Expected (ESF)	Obtained (OSF)	Difference Between OSF-ESF
2900	.53	.30	.20	-.10
4701	.46	.30	.10	-.20
3804	.46	.30	.00	-.30
3801	.46	.30	.40	.10
6100	.45	.10	.10	.00
3500	.40	.10	.30	.20
7700	.37	.10	.10	.00
7600	.36	.10	.00	.10
3803	.33	.10	.00	-.10
1800	.29	.10	.20	.10
8300	.28	.10	.10	.00
7801	.27	.10	.30	.20
8600	.24	.10	.10	.00
7802	.18	.10	.10	.00
3100	.12	.00	.00	.00
4601	.00	.00	.00	.00
10500	.00	.00	.10	.10

Appendix C

The Pretest Sample

THE PRETEST SAMPLE

The first fifteen months of the project were spent in organization and planning, and in constructing and pretesting data collection schedules and procedures. A thirteen percent subsample of the projected Wave 1 sample was defined for this purpose and interviews initiated in the Fall of 1987 (November 1987 to January 1988). As in the projected Wave 1 sample, males were overrepresented relative to females. Students from Spanish-speaking households were not included in the pretest but were included in the Wave 1 panel.

The pretest subsample of 127 was retained for the Wave 1 sample, supplemented with new cases to sum to the desired panel size of approximately a thousand (N=987). In the Pretest and the first wave of data collection, parental interviews were conducted before the student interviews so that the project could be described to the parent and written informed consent obtained. In all subsequent waves, the order is reversed so that information about the identity and location of the primary adult caretaker can be obtained from the student.

Table C.1.   The Pretest Sample

Random Sample of 7th and 8th Grade Students	900
Less Ineligible Cases: Sampling Frame for Pretest	798
Sampling Pool	220
Interviews conducted	127

Appendix D

Weighting the Sample

WEIGHTING THE SAMPLE

Since the sample was drawn disproportionately from different census tracts according to the resident arrest rate (RAR) of each tract, the sample was weighted differently for each tract to approximate a random sample drawn from the sampling frame of eligible cases. To simplify the weighting procedure, census tracts were grouped into five sections with each section comprised of tracts that were similar in RAR level. Weights are grade-specific in order to assure representative numbers of subjects in both the seventh and eighth grades. Disproportionate sampling by gender was constant across census tract sections (75 percent male and 25 percent female); males and females in the sample were weighted accordingly.

The weights were calculated once the panel was defined at Wave 1. Replacement cases were drawn in some instances, however, after the weights were calculated. These replacement cases were drawn nonrandomly for specific purposes -- e.g., to increase the number of white females in the sample, or to assure the representation of particular types of students likely to drop out of the panel. When selected for these purposes, students chosen as replacements tended on average to live in neighborhoods that were undersampled within the original design, and therefore have relatively large weights. Thus a side effect of replacement is a slight artificial inflation in the number of cases in the weighted sample. Without correction, this inflated sample size could bias analyses by increasing the likelihood of significant findings. Although this inflation is quite small, it is corrected by multiplying all of the cases after weighting by a constant that operates to make the weighted sample size equivalent to the number of students interviewed. For analyzing boys and girls separately, this constant is (1000/1045.34); for analyzing boys and girls together, this constant is (500/507.67) for boys and (500/561.05) for girls.

Weights to analyze boys and girls separately are presented in Table D.1. Weights to generalize to a population that represents males and females equally, that is, when analyzing boys and girls in the same analysis are presented in Table D.2.

Table D.1. Weights for analyzing boys and girls separately

Census Tract Section	7th Grade Boys	8th Grade Boys	7th Grade Girls	8th Grade Girls
1 (High RAR)	.66007	.99489	.45918	.49745
2	.59311	.80357	.85140	1.02359
3	.66007	.97576	1.30101	1.01402
4	1.49234	2.22894	1.81759	1.72193
5 (Low RAR)	3.47255	2.44896	3.14730	3.82651

Source: Table 7 in Chard, 1990.

Table D.2. Weights for analyzing boys and girls in the same analysis

Census Tract Section	7th Grade Boys	8th Grade Boys	7th Grade Girls	8th Grade Girls
1 (High RAR)	.45305	.67958	.84663	.90901
2	.40381	.55154	1.56849	1.89823
3	.46290	.65988	2.35273	1.94279
4	1.04399	1.51673	3.29739	3.29739
5 (Low RAR)	2.38344	1.67432	5.85509	6.83540

Source: Table 9 in Chard, 1990.