A Background Report on the Estimation Procedures Developed for the Juvenile Court Statistics Series

Howard N. Snyder

National Center for Juvenile Justice

March, 1988

A Background Report on the Estimation Procedures Developed for the Juvenile Court Statistics Series

Howard N. Snyder

National Center for Juvenile Justice

March, 1988

This report was prepared by the National Center for Juvenile Justice, the research division of the National Council of Juvenile and Family Court Judges, Pittsburgh, Pennsylvania, and was supported by grant #85-JN-CX-0012 from the Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice. Points of view or opinions expressed in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice.

A BACKGROUND REPORT ON THE ESTIMATION PROCEDURES DEVELOPED FOR THE JUVENILE COURT STATISTICS SERIES

An Overview of Relevant Estimation Issues

Estimation is a process in which an inference about a population is made by extrapolating from a subset, or sample, of the entire population. There would be no need for estimation if each member of a defined population were individually measured, if a census were conducted. But often it is impractical to conduct a census. For example, the outcome of a presidential election could be known precisely a day before the actual election by asking everyone who plans to vote how they will cast their ballots, assuming that no one lies or nothing occurs between the time of the polling and the election to change public opinion. But the cost of such a data collection effort far outweighs the time-limited value of the information. Therefore, to estimate how the entire population will behave, a sample of individuals who plan to vote could be polled and their responses used to generate an estimate of the entire population's voting behavior.

A sample is that part of the population selected to represent the whole. Estimates of population parameters with a known degree of uncertainty can be developed by extrapolating from sample statistics; however, the sampling procedure must be sound and the sampling design properly implemented. How can one be certain that a given sample can be used to accurately predict population characteristics? Returning to our voting example, there are many factors that have been shown to influence the nature of a voter's preference: age, sex, race, political attitudes, income, employment status, geographical location, religious preference, and voting history. By knowing each of these factors about a large set of individuals one should be able to generally predict how a large number, maybe even a majority,

of them would vote. There will, however, still be a number of incorrect predictions because voting behavior is a more complex process than any prediction model incorporating the above factors could replicate. This is because other unknown or unmeasurable factors also influence voting behavior. So, how can one know that a sample of individuals is representative of the whole when the sample can not be compared to the entire population on the multitude of factors that result in a specific lever being pulled in a voting booth?

Probability sampling procedures are commonly utilized to identify a representative sample. In a probability sample the chance of any individual being selected into the sample is known and from these probabilities, population characteristics can be extrapolated. There are many types of probability samples. Random samples are constructed when each member of a population has an equal chance of being included in the sample. If a true random sample can be constructed (and generally this is a very difficult task), it is assumed that all the dynamics that influence voting behavior in a population are represented by the individuals in the sample and population estimates can be made with confidence.

Another sampling strategy incorporating randomness is the stratified random sampling approach. The statistical theory supporting the notion of random sampling is actually based on the premise of repeated sampling. That is, a researcher can be certain that if an infinite number of random samples are drawn from a population, the distribution of the average of the sample means will be equal to the population mean. More to the point, sampling theory states that the characteristics of any single random sample may be very different from the population. To protect to some extent against the possibility of a nonrepresentative random sample, the stratified random sampling approach first divides the population into defined groups (or strata) based on a variety of dimensions that are related to the population attribute under study and then randomly samples from

within each strata to generate the estimation sample. Stratified random sampling is valuable in situations where the nature of the population attribute under study varies greatly across subgroups with differing representation in the overall population. Researchers employ a stratified sampling approach when they want to be certain that all significant population subgroups are represented in the sample. For example, one candidate may have an extremely loyal following in one geographical area. A pure random sampling procedure may miss or severely underrepresent this voting block. By stratifying the sample by geographic area such misrepresentation is less likely.

What if probability sampling is impractical? What if the only information that exists comes from a nonprobability sample? By definition probability sampling assumes that the sampling design controls the data collection effort. If the members of the sample (e.g., persons or courts) are limited to a subset controlled by factors other than a probability sampling design, the sample's ability to represent the entire population is questionable. Even if it can be shown that the sample is representative of the population on major dimensions related to the population attribute under study, one can never be certain that the sample can compensate for all of the unidentified factors that influence the level and character of the population attribute. Therefore, unless data collection is controlled by a sound sampling design, the possibility that the sample is biased exists and can never really be dismissed.

An example may help to demonstrate the impact of a biased sample. A classic biased sample occurred in 1948 during the presidential election between Harry Truman and Thomas Dewey. Immediately prior to the election a national poll was conducted which showed that Dewey would win by a large margin. Based on these results and with the hope of being the first on the street with the election results, newspapers published the headline 'DEWEY BEATS TRUMAN.' But, of

course, the prediction was wrong. The source of the error was in the design of the sampling procedure. The pollsters had obtained the names and telephone numbers of all persons who subscribed to a national literary magazine. A random sample of subscribers was polled and the results tabulated. The subscribers, however, did not represent the voting population in the country. They tended to have more education, a higher socio-economic level and to be more partial to the Republican party than the nation as a whole. The carefully designed poll yielded an accurate estimate of the magazine subscribers' voting behavior, but the national prediction was grossly in error.

If the pollsters had collected additional demographic information on the members of their sample, they probably would have discovered that their sample was biased. With this information they may have decided to ignore the results altogether; all they would have lost would have been a headline. Or, if the prediction was important to them, they could have adjusted their estimation procedure to compensate for the known biases in the sample, hoping that these adjustments would overcome the sample's biases.

Adjusting a biased sample is not an easy task and never guarantees that the results produced will be unbiased. The adjustment process assumes that the available sample is not directly representative of the population. However, if the necessary collateral information exists, adjustments can be made to improve the sample's representativeness and predictive ability. In the 1948 election example, it is likely that the sample underrepresented blacks, persons living in rural areas, and the unemployed. If the pollsters had collected, for example, data on the race of those who responded to the survey, they could have made independent estimates on the nature of the black vote. If they also had an estimate of the number of blacks who would vote in the presidential election, they may have been better able to control for the racial bias in their sample. As might be expected, the data necessary

to adjust for some sample biases often do not exist. It is, of course, impossible to control for biases related to unknown causal factors that may be under or overrepresented in the sample.

No one would attempt to develop estimates from a nonprobability or a potentially biased sample unless there was a pressing need for the information and a belief that a better sample could not easily be identified and compiled. This is the situation in which we currently find ourselves when dealing with a national description of juvenile court activity.

Estimation Procedures Based on Nonprobability Samples

Estimates based on nonprobability samples are relatively uncommon because they require a user to accept many assumptions. Unlike the traditional survey research method of designing a sampling strategy and then collecting the data, estimation procedures based on nonprobability samples generally have little control of the data used. The estimates found in the FBI's *Crime in the United States* and in our *Juvenile Court Statistics* series, for example, begin with data made available to the projects by cooperating justice agencies. Even though adjustment techniques have been developed in an attempt to address some of the concerns surrounding the use of nonprobability samples, the lack of control over data collection logically leads to questions about the representativeness of the sample.

The techniques used to compensate for a nonprobability sample mimic, to some extent, the procedures used in stratified random sampling procedures. When probability sampling is not possible, the sample is often stratified into what are believed to be homogeneous subgroups in hopes that this will overcome some of the potential bias in the data. To the extent it is possible to construct subgroups which are alike on key variables which have been shown, or are assumed, to lead to similarities in the behavior under study, then it is reasonable to believe that such

grouping leads to the construction of homogeneous groups. This logic is reminiscent of the previous discussion on stratified random sampling, only in reverse and without the luxury of randomized data collection within strata. The sources of potential error are clear: (1) the stratification dimensions may not yield homogeneous subgroups and (2) the reporting units within each stratum were not randomly selected and consequently may not represent that stratum. The acceptance of estimates that result from a stratified nonprobability sample centers around a reviewer's belief that all the key dimensions are utilized and there is no inherent bias in the available data within the strata or clusters to invalidate the cluster estimates.

The FBI produces national estimates from their nonprobability sample by clustering police agencies into one of eight strata based on the size of the population under the agency's jurisdiction. For example, police departments servicing cities with a population of over 250,000 are placed in the same cluster, as are all police departments that service suburban areas, and all police departments that service rural counties. In 1984 police departments servicing 76 percent of the nation's population reported standardized aggregate data to the FBI. The assumptions underlying the FBI's estimation procedure are (1) the nature and rate of crime is related solely to type of community and (2) the large number of police departments that do report within each stratum are representative of those that do not report.

Each of these appear to be logical assumptions. However, the critical reviewer can never be assured that the subsamples are not biased in other ways. For example, perhaps the reporting police departments in a stratum have a higher proportion of minorities in their general population than does the stratum as a whole. This would probably result in an overestimate of minority involvement in index violent crime. The FBI could test for racial differences but other potential

biases would exist, some of which can not be tested because data to develop the comparisons do not exist. For example, some critics might argue that the reporting police agencies might tend to have more aggressive juvenile units and consequently would tend to have a higher rate of juvenile arrests. Since it is unlikely that the aggressiveness of each police department's juvenile unit has been assessed, this potential bias can not be tested or dismissed.

The Basis of the *Juvenile Court Statistics 1985* Estimation Procedures

This project is faced with many of the same problems as the FBI's Uniform Crime Reporting Program. The juvenile court data form a nonprobability sample. A court is included in the estimation sample if its available data meet the quality-control criteria established by the project. Unlike the FBI's data, most of the information received by the National Juvenile Court Data Archive (NJCDA) is originally collected to meet the needs of the local court. Consequently, the data are important to the data collectors and they are more careful encoding the information than they would be if simply completing a form to meet national reporting requirements. In fact, in many courts the automated information system is an essential tool. For example, many of these systems write petitions, generate dockets and maintain the daily rosters of their detention centers. The data in these systems must be accurate and errors are quickly identified.

Unlike the FBI's Uniform Crime Reporting Program, the juvenile court data collected by each information system is unique. A major task of NJCDA staff is to recode the information into a standardized national format. An intimate understanding of the development, structure, and content of each data set is required. This learning process consumes more time and energy than any other aspect of the project. Every attempt is made to insure that only compatible information from the various data sets is merged into the national file. Recoding

decisions are made through the study of documentation, discussions with data suppliers, diagnostic analyses of the data files and a detailed understanding of each information system's characteristics.

The major concern in the development of national estimates is the wide variation in the nature of court processing rather than the accuracy or uniformity of the reported data. The great heterogeneity across courts raises questions about the ability of available data to represent nonreporting courts. Experience has shown that courts across, and even within states, handle cases very differently. Some of the state differences are related to differences in legislation, the structure of the juvenile justice system, the responsibilities of the various system components and in court rules. For example, in some states, all 16- and 17-year-olds are under the original jurisdiction of the juvenile court (e.g., Virginia), while in other states all 16- and 17-year-olds are classified as adults and are under the jurisdiction of the criminal court (e.g., New York). Variations occurring within states are often related to differences in local policies and programs.

In many ways the available juvenile court data are far superior to the data collected by the FBI. The greatest advantage is that the juvenile court data are primarily individual case records and not simply aggregate counts. Along with having much more information on the activities of its phase of the justice system, the juvenile court data can be disaggregated into homogeneous subgroups of cases based on the youth's age, sex and race. Also the juvenile court data are aggregated at the county-level (which has yet to be done for the FBI's arrest data) enabling the combination of court data with the county-level demographic, economic and sociological descriptive information that is produced by the U.S. Bureau of the Census. With these Census Bureau data the project has searched for the underlying factors that influence the rate of court processing so that more homogeneous subgroups of counties could be identified to develop adjustments for sample biases.

Stratification Dimensions

To address, in part, the heterogeneity of court processing and potential bias in the sample of available data, analyses were conducted to determine on which dimensions the sample should be stratified. In other words, the analyses were conducted to determine which demographic, economic and sociological factors were related to variations in case counts and rates and to court processing. In 1984 the U.S. Bureau of the Census released its *County Statistics File* (COSTAT II). This data file provides over 1,100 county-level items collected by the Bureau of the Census and other federal agencies including the Bureau of Economic Analysis, the Bureau of Labor Statistics, the FBI, and the Social Security Administration, as well as several private agencies. Over one hundred of these county-level descriptive variables were pulled from the COSTAT II file for further study. The variables selected were those items which correlated significantly with the number and/or rate of juvenile court cases. Regression analyses were then conducted using the COSTAT II data as independent variables to predict both the number and rate of juvenile court cases. The population variables (e.g., population ages 10 through 17) dominated the prediction of the volume of cases handled. In fact, the simple Pearson correlation between youth population at risk in a county and its number of juvenile court cases was 0.92. The single variable most predictive of case rate was "Percent of single family households in 1980" with a correlation with case rate of 0.44.

Of course, this large set of variables contained a high degree of multicollinearity, so the set was factor analyzed to minimize redundancy. Eight factors resulted with eigenvalues greater than one. These factors were labeled: Total Population, Income/Employment, Urbanness, Black Population, Growth, Youth Population, Hispanic Population, and Population Density. Regression

analyses were then conducted using the factor scores as the independent variables. The results were rather straightforward. The number of cases referred to court was, as would be expected, highly correlated with the Total Population factor which explained over 80 percent of the variance in case counts. Case rates, which are actually measures of the variation in the number of cases referred after controlling for population effects, were related to the Urbanness factor (which contained high factor weights for percent urban population in 1980, percent single parent head of households in 1982, retail sales in 1982, and serious crime rates in 1983) and the Income/Employment factor. However, these relationships were much weaker than the Population-Case Count finding. Combined, they explained 25 percent of the variance in case rates.

Arguably the relationships between urbanness, serious crime rates, income and employment levels and juvenile court case rates are much greater than shown by these data. Problems lie in the timeliness of the various data elements and their completeness. For example, many of the economic measures found on the COSTAT II tape are compiled once or maybe twice in a decade, so that these measures are not concurrent with the court data. In addition, many are generated by sampling selected areas across the nation, so for many counties the value of the measure is unknown. Much of the information also represents an undercount. For example, nearly 1,500 of the more than 3,000 county-level serious crime rates were flagged on the data file as undercounts, indicating that some unknown proportion of law enforcement agencies within the county did not report their crime statistics to the FBI for that particular year. With better measures of urbanness, serious crime rates, income and employment levels, no doubt stronger relationships would be found with court referral rates.

Variations in the nature of court processing were studied by comparing case rates and proportions across various strata. Table 1 presents the analyses which

show the relationships of age, sex, race of youth referred and county size with the rate and nature of juvenile court processing. Age, sex and race of the youth referred were each related to the rate of court referral and the nature of the cases referred. For example, nonwhite case rates were nearly double white rates. Nonwhite cases were also far more likely to involve a person offense. As is obvious to all who work in a juvenile court setting, the rate of offending and the nature of delinquent behavior changes with age. Younger juveniles are referred at lower rates and their offense profiles contain a larger share of status offenses. These data also imply that the variations in the upper ages of juvenile court jurisdiction across the country should be strongly related to the nature of the problems seen by juvenile courts. Therefore, the demographic composition of a county and the upper age of juvenile court jurisdiction clearly should be incorporated into any estimate of juvenile court activity.

The size of the county was also related to variations in case rate and offense profile (which are also related to the nature of court processing). For this analysis the nation's 3,081 counties (actually county-equivalents) were divided into four clusters based on the size of their youth population 10 through 17 years of age. The counties in each cluster contained as close as possible to 25 percent of the nation's youth population in 1984. The first cluster contained the 2,516 smallest counties in the country, the second the next 411 counties, the third next 117 counties, and the fourth contained the 37 largest counties in the country. The analyses showed that the larger counties had higher case rates and serviced a larger proportion of person offense cases. In fact, the offense profiles across counties indicate major differences in the types of cases courts are asked to handle and, as prior research has shown, the types of dispositions they render.

Table 1

Variations Across Strata in Petitioned Case Rates and Characteristics

Strata	Case Rate (Cases per 1,000 Population at Risk)
Male	35.2
Female	8.4
White	20.2
Nonwhite	39.3
10-15 years of age	16.5
16 years of age	46.3
17 years of age	54.3
Population Quartile 1	18.2
Population Quartile 2	22.2
Population Quartile 3	27.6
Population Quartile 4	24.7

Offense Profiles

Strata	Percent of Cases in Offense Category				
	Person	Property	Drugs	Public Order	Status
Male	12.1	49.0	5.3	18.4	15.8
Female	9.5	33.0	3.8	15.8	38.0
White	12.0	49.9	6.2	21.6	10.2
Nonwhite	25.3	48.7	4.4	16.7	5.0
< 16 years of age	15.0	49.3	3.0	16.0	16.6
16 years of age	15.2	47.4	6.5	20.8	10.1
> 16 years of age	16.0	44.6	8.6	22.9	7.9
Quartile 1	8.9	52.6	2.9	15.5	20.1
Quartile 2	12.9	49.8	3.8	18.7	14.8
Quartile 3	14.9	47.9	5.1	20.4	11.8
Quartile 4	19.8	44.3	6.6	18.3	11.0

The Representativeness of the Sample

With this background, the project was faced with developing an estimation procedure that incorporated the use of a nonprobability sample and observed variations in court processing. The goal was to construct homogeneous strata on which to base estimates. The dimensions that could define the strata are found in the analyses above. As mentioned previously, the volume of cases in any court is strongly tied to the size of the court's youth population at risk. In addition, there were variations in the rate and nature of juvenile court processing based on the age distribution (i.e., the court's upper age of court jurisdiction), the race and sex composition of the population at risk and the size of the community.

Analyses were performed to determine if the reporting sample was representative of the nation along these demographic dimensions. In fact, two samples were used in the development of juvenile court estimates. The first contained aggregate case counts of the number of male and female, petitioned and nonpetitioned cases handled annually by a court. The second sample contained an individual case record for each case processed from courts with automated information or reporting systems. Aggregate counts of the number of cases handled in 1984 were available from 1,600 courts with jurisdiction over 61 percent of the nation's youth population at risk. Detailed case information was available from 1,040 courts with jurisdiction over 44 percent of the nation's youth population at risk. Analyses were conducted to compare the demographic characteristics of the nation with those of the two 1984 reporting samples.

The first comparison explored the relative mix of small and large counties in the two samples with that of the nation. The proportions of the youth population at risk in each of the four county clusters in 1984 are presented in Table 2. Nationally, by construction, each population cluster contained approximately one-quarter of the youth population at risk. In comparison, both estimation samples contained an

Table 2

Representativeness of Sample Characteristics

	U.S. Profile	Aggregate Data Sample	Detailed Data Sample
County Clusters			
1 (small)	25.3%	20.1%	17.4%
2 (medium)	24.8	22.5	20.7
3 (large)	25.4	23.5	25.9
4 (largest)	24.5	33.9	35.9
Age			
<16	79.4%	79.8%	78.5%
16	12.0	11.8	11.0
>16	8.6	8.4	10.5
Race			
White	81.8%	80.2%	81.3%
Nonwhite	18.2	19.8	18.7
Sex			
Male	51.2%	51.2%	51.2%
Female	48.8	48.8	48.8

overrepresentation of the larger counties. Therefore, both samples were biased toward larger courts. The court samples also differ slightly from the nation in their age and race compositions, but had identical male/female components. As Table 2 shows, 79 percent of the nation's youth population at risk in 1984 was below 16 years of age, 12 percent was 16 years of age and 9 percent was greater than 16 years of age. The summary data sample had nearly an identical distribution across age groups, while the detailed automated case record sample had a slightly greater representation of the older youth. Nationally, 82 percent of the youth population at risk nationally was white and 18 percent was nonwhite. Both estimation samples had a slightly larger proportion of their population classified as nonwhite. Finally, the male/female composition of each sample was the same as the nation's. Therefore, strictly in terms of demographic characteristics, the 1984 estimation

samples were slightly overrepresentative of larger jurisdictions and nonwhite populations, and the detailed case record sample was somewhat biased toward older youth.

Variations in the rate of court referral were also found to be related to measures of county urbanness, rate of reported crime and unemployment rate. Unfortunately, measures of these county characteristics are not currently available on a yearly basis. To base a 1985 estimate of court activity on the county's percentage of single parent household in 1980, or percentage of its urban population in 1980, or retail sales figures from 1982 is tenuous. Even the FBI's yearly serious crime rate is unreliable given that over half of the counties in the nation provided only partial data. So, even though these county characteristics could both be important predictors of juvenile court activity and important dimensions on which to assess the representativeness of the sample, useful measures of these characteristics are not available to the project at this time. This lesson has been learned through experience.

The project attempted to incorporate similar data into regression based estimation procedures in the early years of the work. Regression equations to predict court workloads were independently developed using (1) the significantly correlated economic and sociological variables and (2) factor scores based on a reduction of this variable set. The characteristics of these equations changed dramatically with the different yearly samples of court data. It was assumed that the underlying dynamics were not changing (e.g., there was a definite relationship between retail sales - a measure of urbanness - and juvenile court caseloads). But the economic and sociological measures described the counties for time periods different than those of the court data. The relationships between the court and the sociological measures were thus greatly weakened and distorted.

The Proposed *Juvenile Court Statistics 1985* Estimation Procedures

As a result of these and similar analyses conducted over the history of the project, several strata or clustering dimensions have been identified to develop more homogeneous subsamples from which to develop national estimates. These stratification factors include (1) the size of the county based on the number of youth in the county's population, (2) the number of youth in each county in the 10-15, 16, and 17 years of age groups who were under the original jurisdiction of their juvenile court and (3) the number of youth in each age group who were white and nonwhite. It was assumed that each of these strata or subgroups (e.g., 16-year-old whites from the largest counties in the nation) would exhibit similar delinquent behavior and would be processed by their juvenile courts in a similar manner. Knowing the number and types of cases generated by the youth in the reporting courts in each subgroup and assuming homogeneity within the subgroup, it was possible to extrapolate from the reporting sample within the subgroup to the entire subgroup and from these subgroups to the nation.

An example will help to clarify the nature of the estimation process. One stratum or subgroup was defined as 16-year-old white youth who lived in one of the largest counties in the nation. The Bureau of the Census has developed county-level age and race population estimates for 1984. From these data it is possible to know the number of white youth living in each county in 1984 who were 16 years of age. Counties in the largest population stratum were identified and the number of 16-year-old whites noted. However in many states, due to upper age of jurisdiction legislation, no 16-year-olds were under juvenile court jurisdiction. In such counties, therefore, their population of 16-year-olds was set to zero. Consequently, a count was developed of the number of white 16-year-olds under juvenile court jurisdiction living in the largest counties in the nation.

A large proportion of these counties provided a detailed record on each case, including cases disposed in 1984 involving 16-year-old white youth. The underlying assumptions of the estimation procedure are that (1) the youth in this subgroup will have a similar probability of interacting with a juvenile court and (2) those youth who are referred to court will be referred for similar reasons and experience similar processing. If these assumptions are true, then if courts containing half of the youth in the stratum generated X number of cases, it can be estimated that the entire stratum generated $2X$ cases. Similarly if 15 percent of the reported cases were for a violent offense, then it can be estimated that a similar percentage of the estimated number of cases involved a violent offense. In practical terms, if it is determined that the case records in the reporting sample represent only half of the actual number of cases estimated in their stratum, then each case record is given a weight of 2 and is interpreted in all analyses as representing two case records.

Construction and Validation Samples

Paul Tracy in his review of *Delinquency in the United States 1982* and James Lynch in his review of OJJDP's national data collection efforts both recommended a similar approach to study the stability of the national estimates. They recommended that the estimation sample be divided randomly in half into construction and validation samples. Each sample should then be used to develop national estimates using the same procedure and the estimates compared to determine the internal consistency of the nonprobability sample.

The 1,600 reporting jurisdictions in 1984 were randomly divided into construction and validation samples and national estimates were generated from each sample. The samples' strata characteristics are found to be very similar (see Table 3). The estimation procedure was applied to each sample and national

Table 3

Strata Characteristics of Construction and Validation Samples (Case Rates)

	Construction Sample (796 Counties)	Validation Sample (804 Counties)
Overall	21.4	21.4
White	18.7	17.9
<16 years of age	13.0	12.2
16 years of age	36.0	35.0
>16 years of age	47.2	47.8
Nonwhite	35.9	35.8
<16 years of age	26.4	25.8
16 years of age	69.5	70.6
>16 years of age	82.6	87.2

estimates were generated independently. Overall, the national estimates of petitioned cases differed by less than 2 percent (see Table 4). The differences in the national estimates across demographic and processing characteristics were also very small. The construction sample estimated that 70 percent of the youth petitioned were white compared to 69 percent for the validation sample. Both samples estimated that 5 percent of cases petitioned involved a drug offense, 33 percent of petitioned cases were securely detained and 2 percent of petitioned cases were waived in 1984. Overall, the differences in the two estimates are relatively small and show strong support of the internal consistency and stability of the samples as well as the stability of the estimation procedures.

The Rolling Sample

Another approach to test the stability of the estimates was suggested by the advisory committee. It could be argued that those courts which can provide data to NJCDA are different on at least one dimension from those which do not and,

Table 4

Comparison of Construction and Validation Sample Estimates

	Construction Sample	Validation Sample
Estimated number of petitioned cases:	572,000	561,000
Estimated percentage of cases		
Less than 16 years of age	56	55
16 years of age	24	23
More than 16 years of age	21	22
White	70	69
Nonwhite	30	31
White & <16 years of age	54	53
White & 16 years of age	24	24
White & >16 years of age	22	23
Nonwhite & <16 years of age	59	59
Nonwhite & 16 years of age	23	22
Nonwhite & >16 years of age	18	18
Male	80	81
Female	20	19
Person offenses	14	15
Property offenses	49	49
Drug offenses	5	5
Public order offenses	19	18
Status offenses	14	14
Securely detained	33	33
Not securely detained	67	67
Person offense securely detained	41	40
Property offense securely detained	33	30
Drug offense securely detained	38	37
Public order offense securely detained	39	40
Status offense securely detained	20	23
Waived	2	2
Dismiss	26	24
Referred	3	3
Probation	40	43
Out of home placement	18	19
Other	11	9

consequently, may differ on others. This dimension is the existence of an automated reporting system. In fact as previous analyses have shown, the reporting sample contains an overrepresentation of large metropolitan courts, those which would be most likely to need automated systems to handle their case flow. The advisors pointed out that if this selection bias were true, the courts which most recently entered the project's reporting sample might be more like the nonreporting courts than would the original members of the sample. If this were true, it would also be true that the early contributors to the Archive would differ from those which most recently installed information systems.

To test this hypothesis, the 20 information systems which provided their data to the 1984 sample were rank ordered in terms of the age of their information system. For example, Utah was placed first of the list because their system began in 1968, while Minnesota and Wisconsin which installed their systems in 1984 were placed at the bottom of the list. The first ten states on the list were selected as the base sample and national estimates for 1984 were developed using only courts from these jurisdictions. Then the eleventh court on the list was added to the base sample and estimates were recalculated. This process of 'rolling' courts into the sample was continued adding each data set in chronological order until the estimation sample contained all 20 contributing states. In summary, the point of the rolling sample analysis is that the direction of the observed differences in the national estimates would be indicative of the changes that would be seen if the nonreporting courts could be miraculously added to the 1984 sample, assuming that the newest additions to the 1984 sample are most like the nonreporting courts.

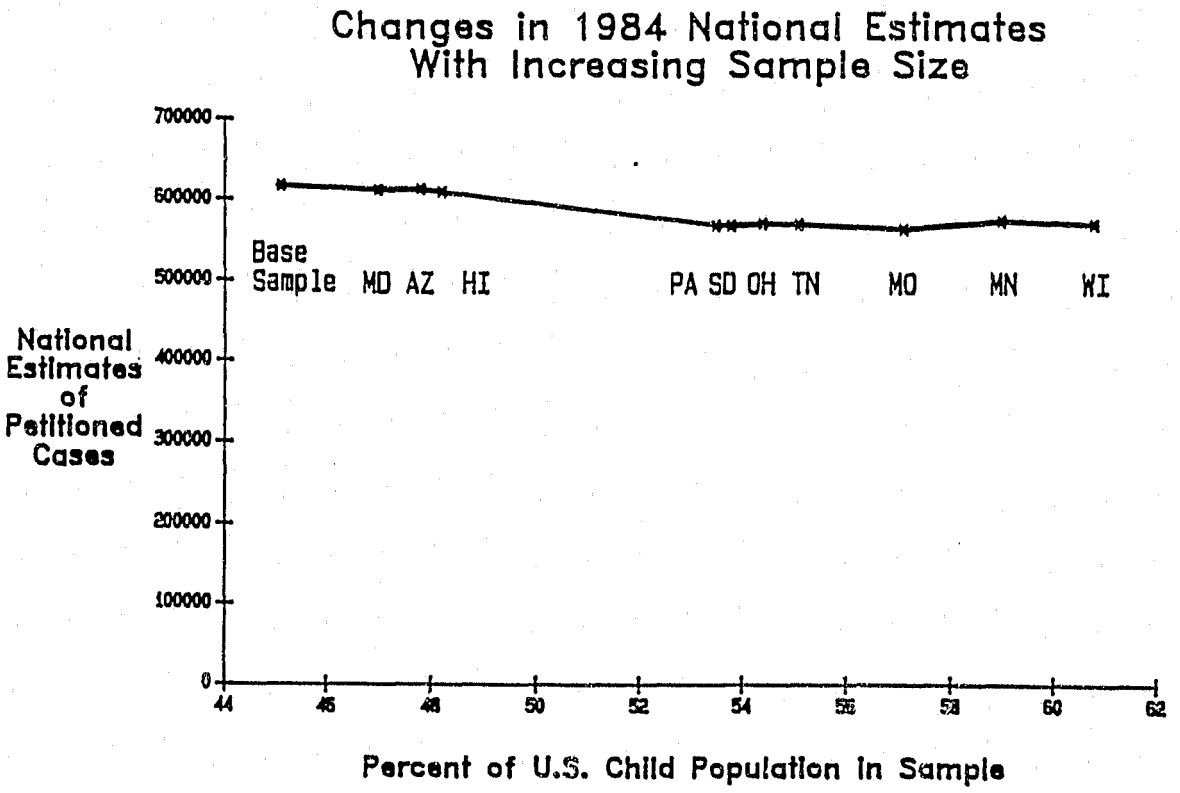
Table 5 and Figure 1 display the results of these analyses. As the figure shows, the national estimate remained relatively constant until Pennsylvania was added to the estimating sample. The effect of Pennsylvania was to reduce the national estimate by 7 percent. Subsequent growth of the sample had little impact

Table 5

Results of the "Rolling" Sample

	10 States	15 States	20 States
Estimated number of petitioned cases:	617,000	568,000	570,000
Estimated percentage of cases			
Less than 16 years of age	55	56	55
16 years of age	25	25	23
More than 16 years of age	20	20	21
White	74	71	69
Nonwhite	26	29	31
White & <16 years of age	53	54	54
White & 16 years of age	26	25	24
White & >16 years of age	21	21	23
Nonwhite & <16 years of age	61	60	59
Nonwhite & 16 years of age	23	23	22
Nonwhite & >16 years of age	16	17	18
Male	80	81	81
Female	20	19	19
Person offenses	13	15	15
Property offenses	47	48	48
Drug offenses	5	5	5
Public order offenses	21	19	18
Status offenses	14	13	14
Securely detained	32	33	34
Not securely detained	68	67	66
Person offense securely detained	40	41	41
Property offense securely detained	31	31	32
Drug offense securely detained	37	38	38
Public order offense securely detained	39	38	39
Status offense securely detained	18	19	21
Waived	2	2	2
Dismiss	24	24	23
Referred	3	3	3
Probation	41	42	42
Out of home placement	18	18	19
Other	13	11	11

Figure 1



on the national estimate. Estimates of demographic and case characteristics were influenced somewhat by the changing sample. The national estimate of white cases declined from 74 to 69 percent of the total juvenile court workload when the sample was increased from 10 to 20 states. The estimated proportion of youth over 16 years of age decreased from 25 to 23 percent. Most other predictions varied only slightly with increasing sample size.

Two tentative conclusions can be drawn from a study of the impact of the natural increasing sample on the national estimates. First, except for the addition of Pennsylvania, the national estimates are relatively stable as the sample expands. Second, the addition of Pennsylvania's data did have a noticeable impact on the estimated volume of cases. Pennsylvania is a large state with a rather unique court processing style. For example, its petitioned case rate is one-third of the average of the other states in the sample. Other large states which are not in the 1984 sample

(e.g., Michigan) may also have some unique characteristics which can not be compensated for by the estimation procedure. This is the danger faced when attempting to estimate for a system that is potentially controlled by a number of locally-unique factors. For instance, the nature of court processing in one county could vary from year-to-year as a result of a change in administrative policy or changes in key personnel, all of which would not be predicted by data from other courts across the country or even within the state. Thus, estimates based on a nonprobability or probability sample are always threatened by the uncommon practices of some large nonreporting (or outlier) courts.[1]

The Validity of the Estimates

National estimates based on a nonprobability sample are always open to doubt and criticism. By their nature nonprobability samples can not generate defined confidence intervals around estimates. The relative accuracy of such estimates can only be assessed by comparing them with other statistics from independent sources. For this work the FBI's Uniform Crime Reports (UCR) *Crime in the United States 1984* provides one point of contact, since both efforts provide estimates of law enforcement referrals to juvenile court.

The UCR statistics are based on information from a nonprobability sample of reporting law enforcement agencies. Despite relying on a nonprobability sample, FBI UCR data are commonly considered to be an accurate reflection of law enforcement activities in this country. Among other things, the FBI reports the number of persons above and below 18 years of age known to have been arrested during the year. In addition they provide national estimates of the total number of

1 This is precisely the reason why the project will attempt in the next year to collect detailed aggregate statistics from the large counties that do not currently provide automated case records to the project. Of the 100 largest counties in the country, 35 do not provide automated case records to the National Juvenile Court Data Archive and these courts will be the targets of the supplementary data collection effort.

persons arrested for various offenses by estimating for nonreporting law enforcement agencies (*Crime in the United States 1984*, Table 24, p. 163). Also reported are the number of arrested persons under age 18 known to have been referred to juvenile courts, probation offices, and related agencies (*Crime in the United States 1984*, Table 68, p. 238).

In 1984 the FBI estimated there were 11,564,000 arrests for various offenses. To estimate number of arrests of persons under age 18 the proportions of all reported arrests in the index violent, index property and the nonindex offense categories which involved persons under age 18 were applied to the total arrests estimates. As a result it was estimated that there were 1,989,000 arrests of persons under age 18 in 1984 (or 17 percent of all arrests that year). The FBI further reported that 60 percent of "under 18 arrests" were referred by law enforcement agencies to juvenile court. Therefore, from FBI data it is estimated that 1,193,000 arrests of persons under age 18 were referred by law enforcement agencies to juvenile court.

Juvenile Court Statistics 1984 estimated 980,000 cases were referred to juvenile courts by law enforcement agencies. The FBI estimate differs from the court data estimate by 22 percent. However, the estimates are employing different units of count as well as a somewhat different definition of juvenile court. First, the FBI counts *arrests* referred to juvenile court, while *Juvenile Court Statistics* counts *cases* referred. If a juvenile were arrested multiple times in a relatively short period of time and then referred to court, the court would probably combine the set of arrests into a single case. Thus, what the FBI counted as two or more referrals to juvenile court would be counted in the court data as only one referral from law enforcement. Second, the FBI counts those arrests that are sent to a prosecutors' offices or other agencies that may be related to the juvenile court system as arrests referred to juvenile court. These agencies may routinely divert cases away from the

juvenile justice system (e.g., prosecutor may not pass all referrals on to the juvenile court). Thus, a law enforcement agency may count as *referred to juvenile court* some arrests which were never in fact received by court intake.

Critical reviewers must decide for themselves whether, given these explanations, the 22 percent difference in these estimates is acceptable. The magnitude and direction of the difference are certainly consistent with an understanding of the juvenile justice system and the nature of the two data collection efforts. However, as with most work based on nonprobability samples, the validity of the results will always be open to debate.

How Could the Estimates be Improved?

In the best of all possible worlds, a national description of delinquency and status offense cases would be based on a census of all such cases handled by every court in the nation with juvenile jurisdiction. In reality the cost of conducting such a census at this time would be prohibitive. At a minimum all courts would have to agree to report information on each case handled. Most juvenile courts with automated information systems would probably be able to provide the requested information. However, those courts without information systems or those which collect incompatible information would either have to install new data collection systems or open their confidential records to data collectors. Needless to say, a national census is a very expensive option.

Another possibility is to mount a new data collection effort based on a probability sample of juvenile court cases or more likely juvenile courts. Of the more than 2,000 courts with juvenile delinquency jurisdiction nationally, possibly a 10 or 20 percent stratified random sample could be selected, stratified along dimensions related to the incidence and nature of juvenile court activity. Each court selected could be asked to provide a complete listing of all the cases handled

in a calendar year. Once again, those with automated systems would probably comply. Those without information systems would have to install a new reporting system in order to comply. If the local court were forced to bear the cost of the new information system, their inclusion into the reporting sample would be seen as an unlucky additional burden and they would be reluctant to cooperate. Even if the federal government could tie formula grant payments to compliance with this data collection request, the quality of data from a mandated effort would be questionable because the content of the data would be of little relevance to the court itself. If the cost were covered by the federal government, OJJDP's budget for research would have to increase substantially.

The fact is that just such a stratified random sampling approach was attempted between 1957 and 1969 by the Department of Health, Education and Welfare when they had the responsibility for producing the *Juvenile Court Statistics* report. The country was divided into about 250 strata based on geographic location. One unit was randomly selected from each stratum to represent the entire stratum. This unit (or court) was asked to report the total number of delinquency, status offense and dependency cases they handled each year. No attempt was made to collect more detailed information. The reported statistics where then weighted by the total population of each stratum to generate stratum estimates and these estimates were summed to produce the national estimates. However, the integrity of the sampling design could not be maintained. Each year a proportion of the courts identified as members of the reporting sample were unable or unwilling to provide the requested information. Other courts were used in their places when possible, violating the stratified random sampling design. Finally in 1969, HEW decided that they could no longer maintain the sampling strategy and turned to the use of available data, or in other words a nonprobability sample.[2]

2 In fact for 1969 HEW used both their probability sample and the sample of all available data to generate national estimates. The results differed by less than one-tenth of one percent.

Given the level of available funds and local cooperation, it is difficult to believe that a census of courts or even a probability sampling approach could be installed to develop national estimates of juvenile court activity. So, what can be done?

A Supplemental Survey Focusing on Large Courts

The nonprobability sample will always make the national estimates open to criticism. Some will always say that no matter how accurate and detailed the data in the sample are, one can never be certain that they reflect the activities of the nonreporting courts. A probability sample of nonreporting courts could be identified and their case records reviewed manually to assess the validity of the predictions. However, the sample size needed to insure the validity of the estimates and the fact that sampling should probably be replicated at frequent intervals raises serious questions about the benefits of this approach relative to the costs.

Another tactic avoids the cost of reading records and takes advantage of data collection activities naturally existing within the nonreporting courts. To date, the work of the project has focused on statewide data collection. In some states such data collection is rather meager, even though some local courts within the state may be collecting more detailed information. The lesson learned from the rolling sample analysis shows that changes in national estimates will be caused by the introduction of data from large jurisdictions with unique case handling characteristics. The Archive currently receives data from 65 of the 100 largest counties in the nation. The 35 missing from this set are the primary source of potential error in the national estimates. For example, if the courts in Detroit or Chicago have a unique processing style, their large volume of cases could dramatically impact the true national statistics. Some of these courts may have automated information systems and be willing and able to contribute their data to

the Archive. The cost of collecting, analyzing, documenting and standardizing 35 new data sets in a year is, however, prohibitive. In addition, the project and the national estimates can not wait.

So, during the current project period the NJCDA staff is conducting a survey of the 35 largest counties in the nation that do not presently report case-level data to the Archive. We are studying the data they normally collect and will be asking them to provide us with a summary of court processing for 1986 in a form that can be incorporated into the national estimation procedure. In the past the project has received from courts either automated case records or simply aggregate counts of the number of cases processed. From these 35 courts we will be requesting data with a detail that falls in between these two extremes. For example, we will be asking the courts to provide a sex and offense profile of the cases disposed in 1986 and a description of their detention activities in terms of the number of males and females detained within various offense categories.

The key to this process is to use terminology familiar to the court and to ask only for information that they can provide with minimal effort. This will be accomplished by reviewing the statistical reports the courts produce on a routine basis and designing our survey to tap their normal reporting format. It is likely that some courts will not be able to provide all the information we would like. In these situations we have discussed the possibility of taking a Bayesian approach. For example, if a court does not collect information on the proportion of detentions that involve whites and nonwhites, it is still possible to ask a few court staff who work in the detention centers to give us their best guess on this measure. The multiple responses would give both a measure of the distribution and a measure of the uncertainty. It remains to be seen if the Bayesian approach is feasible. But between the survey and the Bayesian questioning it is likely that we will be able to collect enough information so that outlying courts will not go unnoticed.

Recommendations

Before the recommendations are presented, a practical example of the conflict between the demands for statistical rigor and the needs of decision makers seems appropriate. The newly created federal Sentencing Commission was recently directed by Congress to estimate the impact of the new sentencing guidelines on future prison populations. In a Commission report summarizing the guidelines (Supplementary Report on the Initial Sentencing Guidelines and Policy Statements, 1987, pages 53-54) problems encountered when making such estimations are discussed. The Commission recognized that "projecting future sentencing patterns and inmate populations is difficult under any circumstances."

The report points out that future prison populations will, of course, be substantially influenced by future levels of crime. Unfortunately, except for short-term projections, no reliable method exists for predicting future crime rates. Even if future crime rates could be accurately predicted, there are several other factors which might influence future prison populations. One factor to be considered is prosecutor discretion. Prosecutors may refer cases to other jurisdictions, or simply decline prosecution for various reasons. Enforcement priorities may change, as might plea negotiation practices. Changes in legislation and sentencing structure can also have a major impact on future prison populations. Career-offender and determinant sentencing statutes would dramatically alter prison populations. Judicial discretion must also be considered. Even where sentencing guidelines exist, some degree of judicial discretion usually takes place. "These highly speculative factors make [estimating future prison populations] especially difficult."

The report concludes:

> "... Against the backdrop of these problems, one might fear that any prison projections will be far too speculative to be trusted. But decision-makers do not have the luxury of adopting this view. Given

> the many years required to plan and construct new facilities, federal prison capacity at the end of this century will depend on choices made today. Failure to forecast as realistically as possible could convert an imprecise decision-making process into a wholly arbitrary one. Central to [this] forecasting effort...is the recognition that we proceed in the presence of many imponderables."

Decisions must be made. Without information, decisions rest on personal experience or bias. Most conscientious decision makers search for empirical evidence. It is the researcher's responsibility to provide decision makers with information they need. But we are faced with a dilemma. The basic researcher in us strongly argues that estimates based on nonprobability samples are unsupportable, while the policymaker in us asks for the best available empirical evidence to support decision making. These two points of view are often in conflict, as they are in selecting the best course for the *Juvenile Court Statistics* series.

What Should Be the Content of the Juvenile Court Statistics Reports?

There are four general options for the content of the *Juvenile Court Statistics* reports. The options are ordered from those based on a belief that the estimation procedure generate valuable national descriptions of juvenile court activity to those based on the belief that any extrapolation beyond the sample of reporting courts is improper.

Content Option 1 assumes that national estimates are most valuable and all detailed descriptions of juvenile court activity should be stated at the national level. Such *Juvenile Court Statistics* reports would include national estimates from the most general (e.g., the total number of delinquency and status offense cases handled in a year) to the most detailed characteristics (e.g., the number of 16-year-old white males referred to court for an aggravated assault and waived to criminal court).

Content Option 2 assumes that the national estimates are valuable, but the accuracy of the estimates are such that only general features of juvenile court activity should be presented. Such *Juvenile Court Statistics* reports would present

national estimates of the more general characteristics such as those found in Chapter 1 of *Juvenile Court Statistics 1984*. Any attempt to move to a more detailed level would be limited to subnational statistics, as are found in the "Data Briefs" section of the 1984 report.

Content Option 3 assumes that any detailed national estimates are unsupportable. A *Juvenile Court Statistics* report would assume in large part the character of the FBI's *Crime in the United States* series. Except for a single presentation of a national estimate of the overall volume of delinquency and status offense activity and an offense profile of referred cases, all presentations would be sample specific and trending over years would be based only on changes in commonly reporting courts.

Content Option 4 assumes that national estimates are not at all supportable. The *Juvenile Court Statistics* report would contain no national estimates. The report would be similar to the recent BJS report entitled *Sentencing Outcomes in 28 Felony Courts 1985* (although the *Juvenile Court Statistics 1984* might be subtitled *Delinquency and Status Offense Activity in 1,040 Juvenile Courts*). The report would describe the activities in these courts only. There would be no mention of trends over years and even summaries across the reporting courts would be limited.

The Concerns About Nonpetitioned Data

In addition to these options, one other factor must be considered. Following the recommendations of the advisors, *Juvenile Court Statistics 1984* discontinued the reporting of national estimates of dependency cases because so few juvenile courts were collecting and reporting this information. A major reason for this decline is that in many jurisdictions the primary responsibility for handling these matters has shifted from the courts to social service agencies. The courts are still involved in an oversight role and, when called upon, the power of the court is used to enforce an

intervention plan. In more and more jurisdictions the detailed record keeping of dependency cases is based in the social service agencies, resulting in less and less dependency data found in the juvenile court data files tapped by this project.

This process is happening with officially recognized and informally processed, or nonpetitioned, delinquency and status offense cases. The traditional juvenile court had an intake unit which screened all referrals. Those referrals which court intake determined could be handled informally were serviced either by the court's intake or diversion unit or referred out of the court for service. Those that needed formal attention were recommended to the prosecutor for petitioning. Often the prosecutor's role was *pro forma*, with the real decision making responsibility centered in the court's intake unit. However, in recent years more of the court's intake decision making has been diverted to child service agencies and prosecutors' offices. The traditional intake screening function has moved to other parts of the system. Matters that would have been handled and counted as informal juvenile court cases 20 years ago now enter the juvenile justice system through executive branch agencies, with the court loosing its initial screening function. In fact in many areas the 'juvenile court' consists of little more than courtroom activities with executive branch social service agencies handling intake screening and probation.

Any attempt to develop a conceptual definition of officially recognized, informally handled delinquency and status offense cases that is common across jurisdictions is a frustrating task. Mounting a data collection effort attempting to describe these cases is a major undertaking. In any single jurisdiction as many as three or four agencies would have to be surveyed to capture the entire informal delinquency and status offense caseload. Unless all of these agencies were tapped, any change in the volume or nature of informal cases would more likely reflect a change in system policy than a change in the nature of juvenile behavior. Certainly

the *Juvenile Court Statistics* reports were never designed to reflect the incidence of juvenile law-violating behavior. The series was designed to describe the activities of juvenile courts. So, it is not important to the mission of the series that the juvenile court is becoming only one of many possible options for handling officially recognized law-violating behavior. Unlike the dependency cases, nonpetitioned delinquency and status offense cases which are not serviced by court intake have no other court involvement. Therefore, the lack of information on these matters does not distort our understanding of the activities of juvenile courts. Certainly it is a loss to those who want to use court data to measure delinquent behavior in a community or to study the development of delinquent careers, but in terms of describing court workloads and activities the loss is unimportant.

Over the years the *Juvenile Court Statistics* series has attempted to define courts generically. Regardless of the organizational structure in a specific jurisdiction, the data collection goal was to collect information on cases that were handled by agencies that did the work of the traditional juvenile court. Consequently, many of the data suppliers to the project are social service or child welfare departments. As the responsibilities are further diffused to other agencies, this data collection goal becomes more and more unobtainable on a national level and even within specific jurisdictions.

Should the Juvenile Court Statistics Reports Include Nonpetitioned Cases?

The effect of this general dispersion of traditional juvenile court responsibilities is that it is difficult to define the universe of 'nonpetitioned' cases. Petitioned cases, since they are by definition those matters in which the court is asked to make an adjudication or a waiver decision, still form a defined universe. This situation presents three options for the *Juvenile Court Statistics* series and the data collection activities of NJCDA.

Nonpetitioned Option 1 is that the series continues to define the juvenile court generically and continues to collect data on nonpetitioned cases from all juvenile courts and also from executive branch agencies that serve in the role of the traditional juvenile court. States or communities would only be incorporated into a national nonpetitioned estimation sample if we received data from all agencies handling nonpetitioned cases.

Nonpetitioned Option 2 is that the series begin to define nonpetitioned cases more narrowly, as only those cases informally handled by the juvenile court in the community. The generic definition of juvenile court would be abandoned. In many states and communities, therefore, the number of nonpetitioned cases would be defined as zero since the court handles only petitioned cases. National estimates could be developed for these cases by incorporating an understanding of the juvenile court's responsibilities in each state. For example, if a court which did not report data to NJCDA had intake or diversion responsibilities its youth population at risk would be used within the estimation base, while courts without these responsibilities would be assumed to have no such cases.

Nonpetitioned Option 3 assumes that (1) the concept of the nonpetitioned case is so vague and (2) fluctuations in the estimates of the volume and type of nonpetitioned cases handled from year to year by juvenile courts would be so hard to interpret, that all estimation of nonpetitioned cases should be stopped.

Within each of these three options the subnational reporting of nonpetitioned cases (however they are defined) could still be a part of the *Juvenile Court Statistics* reports. The "Data Briefs" section could still display the nature of these cases. Even if national estimates of nonpetitioned cases are abandoned, NJCDA could continue to collect the information just as it now does with the dependency data, but the priorities of the data collection activities of NJCDA would be on the court's formally handled, petitioned, cases.

The Basis for Reommendations

The final recommendations are based on the following assumptions:

Assumption 1: Even if the *Juvenile Court Statistics* reports does not present national estimates, individuals who need national estimates of court activity will extrapolate them from the presentations. This is certainly the experience of the *Crime in the United States* series. Almost everyone reads the percentage increase in arrests for index violent crimes as a national estimate, even though the careful reader knows that all these tables are sample based. Users of their report would prefer the FBI to incorporate all that it knows about the biases in its sample into a (provisional) national estimate, instead of encouraging a straight-line extrapolation from their sample statistics.

Assumption 2: The split-half and rolling sample analyses demonstrate the stability of the estimates and give some confidence in their accuracy.

Assumption 3: The new survey of large courts should protect the estimates against the error caused by estimating for a large court that is actually an outlier. The collection of more detailed aggregate statistics from these large courts will improve the face validity of the estimates.

Assumption 4: The estimates and the estimation procedures are not cast in stone. It is common practice for the FBI and the Bureau of the Census to adjust previously reported estimates if they find there is some error in the published figures. The *Juvenile Court Statistics* series can do the same. In addition, if more current county-based sociological, economic or crime data become available, investigations could be conducted to determine if these data should be incorporated into the estimation procedure.

Assumption 5: The conceptual definition of a petitioned case is commonly understood and measurable across courts. With very few exceptions, there is only

one court in a community that has the original authority to adjudicate or waive a youth for a delinquent or status offense. The boundaries are less clear for the definition of a nonpetitioned case, unless it includes only those informal cases processed by an actual juvenile court and not a social service or child welfare agency.

The Final Recommendations

The *Juvenile Court Statistics* series should adopt Nonpetitioned Option 4. The report should present subnational descriptions of nonpetitioned cases in a "Data Briefs" format, as it should dependency case information. The generic definition of court is too difficult to support in any reasonable data collection effort. The more limited definition, informal cases handled by a actual juvenile court, may support some national estimates, but any variations in estimates over time would be difficult to interpret. Certainly some of the same interpretation problems also apply to petitioned case estimates, but this is clearly a more organizationally stable process.

The *Juvenile Court Statistics* series should adopt Content Option 2. The report should present national estimates of the general nature and processing of petitioned cases. More detailed analyses should be included in the "Data Briefs" section or in special reports prepared by the project.