THE 1989 REDESIGN OF THE SAMPLE
FOR THE SURVEY OF LOCAL JAILS

by

Carma R. Hogue
Statistical Research Division
Bureau of the Census
Room 3000, F.O.B. #4
Washington, D.C. 20233 U.S.A.

Recommended by: Nash J. Monsour

Report completed: October 30, 1989

Report issued: November 8, 1989

The 1989 Redesign of the Sample for the
Survey of Local Jails

1. Introduction

In 1982 the first sample survey of local jails was conducted. The
Governmental Organization and Criminal Justice Statistics Branch of the Bureau
of the Census conducts this survey for the Bureau of Justice Statistics
(BJS). In 1978, 1983, and 1988, complete enumerations of all U.S. jails were
conducted, with full censuses scheduled for every 5 years hereafter. Between
censuses, annual sample surveys of jails provide estimates of national jail
inmate population characteristics. Surveys were conducted in 1984-87. This
documentation is for the redesign sample for 1989-92. For this survey, local
jails are facilities that are staffed by local officials and that hold inmates
beyond arraignment, usually more than 48 hours. Federally and State-
administered jails are excluded. Combined jail-prison systems in Connecticut,
Delaware, Hawaii, Rhode Island, and Vermont are also excluded.

The sample is a stratified simple random sample of jurisdictions. The
universe for the survey includes all county or city jurisdictions that
administer at least one local jail. When a jurisdiction is included in the
sample, all jails within the jurisdiction are surveyed. More details on the
sample design are given in section 2.

National estimates for the total inmate population on June 30 are
produced annually by sex, race, legal status, and conviction status. National
estimates by sex and legal status were also obtained for the average daily
inmate population (ADP) and for admissions and releases for the year ending
June 30. Questions were also asked on the jail's capacity. Jails in
jurisdictions with more than 100 inmates were given a long form to complete.
For these jails, questions on inmate deaths, overcrowding, and jails under

court order were also included. The estimation procedure is discussed in section 3.

The overall nonresponse rate was very low ($\leq$ .5%) in 1984. Over time, the nonresponse rate is rising as the local jails are being pressed for data from an increasing number of sources. The nonresponse rate for the 1987 survey was 4 percent. Item nonresponse is also present. The procedure for dealing with nonresponse error is given in section 4. Other nonsampling errors are discussed in section 5. Suggestions for further research are given in

section 6.

2. Sampling Procedures

As stated before, the estimates from the Sample Survey of Jails are made from a sample of city or county jurisdictions that administer at least one jail. The design is a stratified simple random sample of jurisdictions. Data from the 1988 Census of Local Jails were used to stratify. The 1988 average daily inmate population was used to stratify the sample with allocation to the strata based on 1983 ADP. Other variables for 1983 and 1988 were also considered in the final allocation: female average daily population, female inmates, expenditures, staff, and juvenile inmates. An optimum allocation to the strata was calculated for each of the variables listed above, and a compromise allocation, given in the table, was selected. More weight was given to the 1983 ADP allocation in selecting the compromise.

BJS decided that the initial certainties should include all jurisdictions with a 1988 ADP of 100 or more inmates. For the 1989 redesign there were 505 initial certainties, an increase of 142 over the 1984 design. (Some of this large increase in certainty jails was thought to have resulted from mergers of

smaller local jails into regional jails.) Three other jurisdictions were added with certainty because they contained either all-female jails or a substantially higher number of female inmates than the other jurisdictions of comparable size. These were 471022022, 491025025, and 181024024. Including these jurisdictions with certainty decreased the standard error and the sample size needed to meet the relative standard error requirements. No other jurisdictions of comparable ADP size demonstrated characteristics like these. Therefore, they received a weight of 1.0000, indicating that they represent only themselves. These jails should not receive the longer form as their 1988 ADP<100.

1989 Sample of Local Jail Jurisdictions

Stratum	'88 ADP	M_h	m_h	TE
1	Certainty	508	508	1.0000
2	60-99	241	64	3.7656
3	40-59	293	58	5.0517
4	25-39	414	59	7.0169
5	14-24	497	56	8.8750
6	0-13	1030	65	15.8462
		2983	810	

Prior to selecting the sample, the file of jurisdictions was ordered as follows: within stratum, jails containing at least one female, by region, and within region by ADP size; followed by jurisdictions with all-male jails by region, and within region by ADP size. This yields an implicit stratification (and insures adequate representation) of females and regions. The regions are the census-defined Northeast, Midwest, South, and West regions.

The general procedure that was followed in designing the sample was as follows:

1. Printouts of 1983 and 1988 census totals and sums of squares for ADP, female ADP, expenditures, staff, juvenile inmate count, and female inmate count for 21 specified preliminary strata were obtained. [These strata which were defined using 1988 ADP were collapsed to the strata given above.]

2. The sample frame was a listing of all jurisdictions that administered at least one jail. A listing of jurisdictions on the 1988 Census universe file with jurisdictions ordered from largest to smallest ADP was acquired. Data from 1983 and 1988 for ADP, female ADP, expenditure, and staff were requested for each jurisdiction. Several variables were needed in order to give an idea of the variability from one characteristic to the next. If we used only ADP for the allocation, ADP is the only variable that we could safely assume would have a 1 percent CV. Others could have a much higher CV. ADP remained as the most important variable but sample sizes were adjusted to give reasonable CV's on other variables. Expenditure and staff are not Jail Sample Survey variables, but they are the only other related variables from the census that could give an idea of variability from one characteristic to the next.

3. The sample size needed to obtain a 1 percent coefficient of variation for ADP, female ADP, expenditures, staff, and female inmates was determined and a sample size of 810 was found to satisfy the CV requirement for the important variables (ADP, inmate population, female ADP, and female inmate population).

4. The strata were collapsed, and allocations were made to the strata using each of the variables. A compromise allocation was selected. The allocation using ADP was slightly altered in order to give better estimates for the other characteristics. The compromise did not harm ADP estimates. The total sample size was increased by less than 10 jurisdictions in order to give the better estimates.

5. Using the compromise allocation, CV's were calculated for 1983 ADP and inmates and 1983 and 1988 female ADP and female inmates. They were checked to see if they met the CV requirements.

6. A second universe list was obtained. Jurisdictions were ordered by stratum and within each stratum by jurisdictions with jails that confined females by region and within region by 1988 ADP size, and jurisdictions with all male jails by region and within region by ADP size.

7. Take-every rates (the inverse of the selection probability) were obtained for each stratum. A random start was obtained for each stratum,and a systematic sample was drawn using the new ordered universe listing.

3. Estimation

3.1 Estimate of Total

Each year BJS publishes certain totals on jail capacity, total average daily population, total inmates on June 30, and total admissions and discharges categorized by race, sex, and/or adult/juvenile status.

The estimate of total that is currently used is the sum of the weighted stratum totals for each characteristic of interest. This equation is given in (3.1).

$$x' = \sum_{h=1}^{6} \frac{M_h}{m_h} \sum_{i=1}^{m_h} X_{hi} \qquad (3.1)$$

where

x' is the estimate of the characteristic of interest for the current year.

M_h is the total number of jurisdictions in stratum h.

m_h is the number of jurisdictions in stratum h that are in the sample.

X_{hi} is the total of the characteristic of interest for the i-th jurisdiction in stratum h. (Totals from all jails within the jurisdiction are to be added to get this jurisdiction total.)

3.2 Estimation of the Difference

Each year several comparisons are made to the previous year's data. This year-to-year change is estimated by simply dividing the current year's total by the previous year's total and subtracting one from the ratio. (This difference is then tested to see if it is

significantly different from 1.) Equation (3.2) gives the formula for calculating change:

$$z' = x'/y' - 1 \qquad (3.2)$$

where z' is the estimate of change;
                x' is the current year's total of interest; and
                y' is the previous year's estimate of total.

3.3 Estimate of the Variance of the Estimated Total

The estimated sampling variance of x', $s^2_{x'}$, is calculated using equation (3.3). <u>The sum is only over the noncertainty strata.</u>

$$s^2_{x'} = \sum_{h=2}^{6} M_h(M_h - m_h) \, s^2_{hx}/m_h \qquad (3.3)$$

where

$$s^2_{hx} = \sum_{i=1}^{m_h} (x_{hi} - \bar{x}_h)^2/(m_h - 1) \text{ and}$$

$$\bar{x}_h = \sum_{i=1}^{m_h} x_{hi}/m_h.$$

All other terms were defined in section 3.1. The algorithm given in the attachment is used to calculate s^2_{hx}. A(I) is the vector of x values for the stratum and N is the number of sample jurisdictions in the stratum.

3.4 Estimate of the Variance of Estimated Change

The variance of the year-to-year change is estimated by using equation (3.4). If the previous year is a census year, $s_{z'}^2 = s_{x'}^2$.

$$s_{z'}^2 = (x'/y')^2 \left[\frac{s_{x'}^2}{x'^2} + \frac{s_{y'}^2}{y'^2} - 2 \frac{s_{x'y'}}{x'y'} \right] \tag{3.4}$$

where

$s_{y'}^2$ is estimated by using equation (3.3) for the previous year's data.

$$s_{x'y'} = \sum_{h=2}^{6} M_h(M_h - M_h) \, s_{hxy}/m_h$$

where

$$s_{hxy} = \sum_{i=1}^{m_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)/(m_h - 1)$$

As stated in section 3.3, s_{hy} and s_{hxy} are calculated using the algorithm given in the attachment. B(I) is the vector of y values for the stratum. These y-values are in the same order as the x-values. In the output of the subroutine AVAR is s_{hx}^2; BVAR is s_{hy}^2; and COV is s_{hxy}.

3.5 Other Estimates of Error

The standard error of the estimate, x' or z', is the square root of the variance of x' or z'. The relative standard error is the standard error of the estimate divided by the total of interest, as given below:

$$s_{x'}^2/x' \text{ or } s_{z'}^2/z'.$$

The relative standard error or coefficient of variation is used in the publication.

4. Nonresponse

The total response rate for the survey has been very good, ranging from 100% in 1982 to about 96% in 1986. These high response rates were achieved through second and third mailouts and followup telephone calls. The response rate has gradually decreased since 1982 when the survey was started. The use of BJS's "SAC" groups in the nonresponse followup effort is expected to bring the response rate back up. Although total response rate is very good, there is a problem with item nonresponse, particularly with breakdowns of the data by race. Most of the other items have a very high response rate. Every 5 years in the Census of Local Jails, all jails respond to all items.

In the past, if a record was missing, the most recently available values were substituted for the value. Prior to 1986, if race was missing, the previous year's race breakdowns were applied to the current year's jail population. Since 1986, the "DK - don't know" category has been used for race. For total nonresponse in the current sample, a growth factor will be applied to the substituted data for each jail. If the jail has responded in the past, and if a factor can be calculated from past reports from the jail, that factor will be used. If there is an inadequate reporting history, a

growth factor can either be calculated by using the jail's 1983 and 1988 census reports, or by using the average growth rate of the other jails in the stratum, or by using a hot deck procedure to obtain a growth rate from another jail with similar characteristics.

5. Nonsampling Error

As in any survey, errors other than sampling errors are present. Section 4 includes a discussion of nonresponse errors and what is being done to limit them. Other errors result from questionnaire design, improper reporting, and clerical errors in examination and tabulation procedures.

In order to restrict these types of errors, the data were edited, and items that were questionable were verified by telephone with the respondent. Data had to be internally consistent and consistent with previously reported data. In some cases, independent sources of the data are available, and they were used in the verification process.

6. Future Research

There are a few considerations for future research. First, since some small local jails are now consolidating to form larger regional jails and since jail sizes are growing, some consideration should be given to raising the certainty cutoff, which is currently jurisdictions with an ADP of 100 inmates. In 1982, there were only 287 certainty jurisdictions. For the current design there are 505 certainty jurisdictions above the 100+ certainty cutoff. If the total sample size is to remain around 800-900 jurisdictions and if the targeted relative standard error is to remain about the same, this cutoff should be raised. Certainty units represent only themselves. The 505 jurisdictions represent themselves, but a sample must still be taken of the

remaining noncertainty jurisdictions. If the sample size remains at 900,

fewer jurisdictions can be used to represent the noncertainty portion. As the

noncertainty sample size decreases, the variance increases. Therefore, if the

certainty cutoff remains at 100, we will either have to increase the sample

size or accept more variability in the data. More error in the data will

decrease the analyst's ability to draw any valid conclusions from the data.

A second area for research is in estimation. Since there is a census

every 5 years, it may be beneficial to use a ratio or regression estimator

instead of a simple unbiased estimator. The ratio or regression estimator is

used to acquire greater precision by taking advantage of the correlation

between the census and sample data. Data from the past 5 years could be used

to study the different estimators.

In the area of nonresponse, different estimators of the growth factors

could be studied to see which method of calculating the factor seems to

produce the best imputation.

Finally, in the area of evaluating nonsampling errors, several evaluation

studies could be planned to attempt to measure the amount of reporting error

by the jail, to check the edit system currently used, or to evaluate the

questionnaire design.

Attachment

```
C    THIS IS A SUBROUTINE FOR CALCULATING SUMS, CROSS PRODUCTS AND SUMS OF
     SQUARED DEVIATIONS FROM THE MEAN.

     SUBRØUTINE SUMSQD(A(I),B(I),N,ASUM,BSUM,ASQD,BSQD,CRØSS)
     DØUBLE PRECISIØN ASUM,BSUM,CRØSS,ASQD,BSQD
     ASUM=A(1)
     BSUM=B(1)
     CRØSS=0.
     ASQD=0.
     BSQD=0.
     DØ 1 I=2,N
     ASUM=ASUM+A(I)
     BSUM=BSUM+B(I)
     CRØSS=CRØSS+(I*A(I)-ASUM)*(I*B(I)-BSUM)/(I*(I-1))
     ASQD=ASQD+((I*A(I)-ASUM)**2)/(I*(I-1))
     BSQD=BSQD+((I*B(I)-BSUM)**2)/(I*(I-1))
1    CØNTINUE
     COV=CROSS/(N-1)
     AVAR=ASQD/(N-1)
     BVAR=BSQD/(N-1)
     RETURN
     END
```