SELECTION OF STRATIFIERS FOR THE NATIONAL CRIME SURVEY REDESIGN

Charles H. Alexander and Edward L. Kobilarcik, U.S. Bureau of the Census

I. INTRODUCTION AND SUMMARY

The National Crime Survey (NCS) measures criminal victimizations through interviews with individuals age 12 and over in a probability sample of living quarters throughout the U.S. Both crimes against persons (such as robbery or assault) and crimes against households (such as breaking-and-entering) are measured. Each year since 1973 occupants of approximately 60,000 housing units have been interviewed twice at intervals of 6 months. Crime rates as reported to interviewers are not regularly available from any other source. However, the FBI's Uniform Crime Reports (UCR) program, which measures criminal incidents as reported to police, has data for most localities. Unfortunately, the relationship between NCS rates and UCR rates is not well understood and little use was made of UCR data in designing the NCS.

Now that the NCS has been in operation for 8 years, it is possible to use information collected by the survey to evaluate and revise the survey design. This paper presents results to date of our investigation using NCS and UCR data to examine the stratification of primary sampling units (PSUs) in the current design. This is of particular interest because estimates of the components of variance (based on 1976 data) indicated that there may be a lack of homogeneity in the PSU stratification.

The investigation suggests that the inhomogeneity is concentrated in a fairly small number of strata and also that there may be some deficiencies in the estimates of the components of variance. It also identifies demographic variables which might be used as stratifiers to improve the stratification of PSUs. Portions of the original plans for the investigation remain to be completed; comments, suggestions, and references to similar studies for other surveys would be most welcome.

The NCS Sample Design

The NCS sample design was derived from the design of a portion of the Current Population Survey (CPS). Both surveys use a stratified multistage design in which the PSUs are counties or groups of contiguous counties. More details on this design may be found in references [1], [2], and [3].

The NCS sample PSUs were selected from a universe of 1,930 such PSUs. Prior to the selection, the 1,930 PSUs were grouped into 376 strata; one PSU was selected from each stratum. One hundred and fifty-six of the PSUs, generally the most populous, were made "self-representing" (SR). The remaining PSUs were formed into 220 ("nonself-representing" or NSR) strata by combining two or more PSUs with similar characteristics such as geographic region, population density, rate of growth in the 1960-1970 decade, proportion nonwhite, principal industry, etc. The NSR strata were formed so that their 1970 population sizes were rougly equal.

The variables used to form the NSR strata were chosen because of their relationship to the unemployment rate and other labor force characteristics, which are the main emphasis of the CPS. For the post-1980-census redesign of the NCS, it was decided to examine the current stratification and if necessary form new strata based on variables which were more closely related to crime rates.

A reason for this decision was the relatively large "between-PSU" component of variance for many NCS estimates, i.e., the component which arises because of the variation between PSU means within the NSR strata. The method of estimating NCS variances is described in reference [2]. The between-PSU variance is estimated using the method of "collapsed strata." This method groups or "collapses" pairs of similar strata. The total variance for each collapsed stratum is then estimated from the squared difference of the estimates from the two sample PSUs in the collapsed stratum, much as would be done with a two-PSU-per-stratum design. Weights (informally referred to as "p-weights") are used to adjust the two PSU estimates for the difference in stratum sizes for each pair of collapsed strata. The between-PSU component of variance will tend to be overestimated if the wrong measure of size is used in computing these "p-weights." A positive bias is also introduced if the two strata in each pair of collapsed strata have different stratum means. Thus, the large estimated between-PSU component may be due to inhomogeneity in the strata themselves, or to bias caused by the collapsing of dissimilar strata or use of incorrect p-weights in computing the estimate.

Outline of the Study

This paper presents results from a study which was conducted to (i) examine the causes of the large estimate of between-PSU variance in the NCS, and (ii) investigate methods of selecting stratifiers related to NCS crime rates.

The relationship of selected NCS crime rates to demographic variables was measured by fitting a linear regression model, using data for the NSR PSUs in the sample. Only sample PSUs can be used, because NCS crime rates must be available as the dependent variable. Data for 1976 were used, since these had been used in estimating the components of variance. One or, in some cases, two promising models were selected for each important crime category by applying a stepwise regression procedure to about 40 socio-demographic variables, using data from four states (Ohio, Pennsylvania, Michigan, and Texas), for which there were 40 NSR sample PSUs. The selected models will be tested on data from a different set of states (the southern census region, except Texas), for which there were 83 cases. Details of the model selection procedure are given in the third section.

UCR crime rates at the PSU level were added to the best demographic model for each of the major NCS crime categories. This made surprisingly little improvement in the fit. Possible reasons for this are presented in section IV.

To determine whether collapsing or stratification was responsible for the large between-PSU variance, the most heterogeneous collapsed strata were identified by comparing the estimated crime rates in each collapsed stratum. The differences were standardized so that they would have a uniform distribution between 0 and 1, under the assumption of homogeneity. The deviations from

uniformity of the actual distribution of the standardized differences indicated that for crimes against persons a large portion of the between-PSU variance was due to a few extreme cases. For crimes against households, part of the problem appeared to be due to using "p-weights" based on the number of persons rather than the number of households. These conclusions are discussed in section II.

All PSUs in the most heterogeneous strata were examined with respect to the demographic variables in the best models from the stepwise regression. This part of the study is incomplete, but it so far points to problems with special situations both in the collapsing of strata and in the stratification itself.

II. IDENTIFYING THE HETEROGENEOUS COLLAPSED STRATA: LOOKING FOR CAUSES

The NCS's 220 strata of NSR PSUs are closely related to those defined for the CPS shortly after the 1970 census. The strata have been modified slightly since then, principally to reflect changes in SMSA definition following the 1960 and 1970 censuses. The CPS stratification is described in more detail in reference [1]. The stratifiers were chosen mainly because of their relationship to the unemployment rate and other labor force characteristics. The stratifying was basically by trial and error matching of similar PSUs until strata of suitable size were achieved. Geographic distance between PSUs was taken into account to some extent. Each stratum is wholly contained in a census region (Northeast, North Central, South, or West).

The pairing of strata to form the collapsed strata for variance estimation was based primarily on geographical proximity of the two strata and similarity of their stratum populations, with some consideration of the values of the original stratifiers.

If, as indicated by the large estimated between-PSU component of variance, some collapsed strata contain PSUs with very different crime rates, two explanations are available. The problem could be that variables related to crime rates were not used (or were given too little importance) in the original stratification. It could also be that dissimilar strata have been inadvertently paired in forming the collapsed strata.

As a first step to analyzing the NCS stratification, the most heterogeneous collapsed strata will be identified by comparing the estimated crime rates for the two sample PSUs from each collapsed stratum. In doing this, it will be necessary to take into account the sampling variability in the estimated rates.

Let the NSR PSUs in the sample be numbered i=1, 2,..., 220 and ordered such that for k=1,..., 110, the PSUs 2k-1 and 2 k are the ones from the k^{th} collapsed stratum. Let $\hat{V}_i$ be the estimated number of victimizations in the i^{th} stratum. Let N_i be the stratum population.

If collapsed stratum k is "homogeneous," then

$$(1) \qquad E\left(\frac{V_{2k-1}}{N_{2k-1}}\right) = E\left(\frac{\hat{V}_{2k}}{N_{2k}}\right).$$

Let $D_k = \left(\frac{N_{2k}}{N_{2k-1}}\right)^{\frac{1}{2}} \hat{V}_{2k-1} - \left(\frac{N_{2k-1}}{N_{2k}}\right)^{\frac{1}{2}} \hat{V}_{2k}$

Condition (1) implies that $E(D_k) = 0$.

The factors $\Pi_{2k-1} = \left(\frac{N_{2k}}{N_{2k-1}}\right)^{\frac{1}{2}}$ and $\Pi_{2k} = \left(\frac{N_{2k-1}}{N_{2k}}\right)^{\frac{1}{2}}$ are the "p-weights" used in the variance estimation procedure. D_k^2 gives an estimate of the variance of the estimated crime level in the collapsed stratum. This estimate will be approximately unbiased if the collapsed stratum is homogeneous. For the variance estimation, the p-weights were based on 1970 total population. Other measures of stratum size might be used. In this paper, we have also tried p-weights based on population age 12 and over and, for household crimes, on the number of households.

The estimates $\hat{V}_i$ may denote either the NCS's usual "final" estimates or alternatively the "unbiased" estimates, which are computed omitting noninterview adjustments and post-stratification factors. It turns out that the results of the analysis are similar for the two different kinds of estimate.

Let $Z_k = D_k/S.E.(D_k)$. Assuming (1), Z_k will be approximately standard normal (details are discussed in Appendix A), so Z_k^2 will have approximately a chisquare distribution with df = 1. Finally, let $Y_k = 1-F(Z_k^2)$, where F represents the cumulative distribution function for a chisquare distribution with df = 1. Assuming (1), the standardized differences $Y_1,\ldots, Y_{110}$ should be uniformly distributed between zero and one. Since they represent estimates from distinct parts of the country, they are essentially independent. If Y_k is close to zero, this indicates a large difference between the estimated crime rates for the PSUs from collapsed stratum k; i.e., the collapsed stratum is heterogeneous.

A histogram of the 110 standardized differences for the crime category "total personal victimizations" is given in Appendix B. Except for the interval (0.1, 0.0) the histogram is not inconsistent with a uniform distribution, but there is a strong suggestion of too many values Y_k in the interval (0.1, 0.). In particular, there are 11 values in the interval (0.01, 0.), compared to the expected number 1.1 under the assumption of homogeneity.

To look at this another way, under the assumption of homogeneity the probability is .95 that none of the 110 collapsed strata would have $Y_k < .000466 = 1-.95^{1/110}$. In fact three of the values are smaller than this. Comparing personal and household crimes, the most heterogeneous with respect to household crimes are not generally the worst for personal crimes.

The histograms for the other crime categories are similar to those for total personal victimizations, although the distribution for [1.0,0.1] is not as uniform. There is apparently less extreme heterogeneity for household crimes. This heterogeneity can further be reduced by using p-

weights based on the number of households rather than the number of persons.

Conclusions

A few collapsed strata contribute a large portion of the NCS variance. For the three crime categories considered, the heterogeneity in the four worst collapsed strata contributed about 30 percent of the total estimated variance from the NSR sample. However, we cannot give a simple answer as to whether the problem is collapsing or stratification.

Various socio-economic characteristics of the two sample PSUs from some of the most heterogenous collapsed strata were compared. There does not appear to be a single reason for heterogeneity. A few variables, such as percent urban, percent nonwhite, percent in agriculture, seem to be associated with higher crime in some cases. However, the association is not consistent. It is apparent that in a few cases implausible strata or collapsed strata were formed for want of a better alternative. In some cases, there are PSUs which cannot be combined with any others to form a heterogeneous stratum. In other cases, a reasonably heterogeneous stratum exists, but is not similar to any other stratum for collapsing purposes.

In general, to determine to what extent poor collapsing or poor stratification is responsible for the heterogeneity, it is necessary to consider information about the non-sample PSUs in each stratum. A limited investigation of this type listed values of important potential stratifiers (as selected in the next section) for all PSUs in the strata whose crime rates were "most heterogeneous." The pattern in the stratifiers was not consistent. In some cases, strata were evidently heterogeneous, in others, the strata seemed homogeneous but the collapsing did not seem appropriate, and in others there was no evident explanation for the large differences in crime rates. This investigation has not been completed and the details are not presented in this paper.

III. SELECTION OF NEW STRATIFIERS

Current plans are to restratify the NCS PSUs based on information collected about each PSU in the 1980 census. Unlike 1970, when the NCS had to use the CPS stratification in order to take advantage of the existing interviewer staff, current plans are for the NCS to select its strata separately. To prepare for this restratification, the relationship of 1976 NCS crime rates and 1970 census variables has been examined for a portion of the NCS sample PSUs.

Potential stratifiers are many. Data from the regular NCS survey show that rates for many types of crime vary according to the race, age, income, sex, marital status, employment status, and household size of the potential victims. Rates are highest in urban areas and lowest in rural areas. Also of potential interest are measures of neighborhood age, degree of crowding, and neighborhood transiency.

It should be kept in mind, however, that variables associated with crime at the household or individual level are not necessarily useful stratifiers at the PSU level. For example, rates of violent crime are much higher among males than among females, but the proportion of males varies so little from PSU to PSU that sex ratio is not a useful stratifier.

There are other reasons the usual NCS national data on characteristics of victims or offenders may not be a good guide to selection of PSU stratification variables. For one thing, only NSR areas are of interest in determining stratifiers. Factors influencing crime in the largest cities may not affect the NSR areas. More important, interactions may be present at the PSU level which are not apparent in the national data. For example, crimes of violence are less frequent against high-income individuals than low-income individuals. This does not mean that high-income PSUs will have low rates of crime of violence; it could be that crime rates are exceptionally high among those low-income individuals who live in the kinds of area which tend to have high per capita income.

The complexity of the crime phenomenon means that extremely good prediction of NCS crime rates based on a few demographic stratifiers should not be expected. Additionally, a limit on the explainable portion of the variance of NCS PSU-level crime rates is imposed by the fact that the NCS rates are sample estimates. The proportion of the total variation among observed PSU rates which is due to the selection of the sample within the PSU is estimated at values from 23 percent for total personal victimization to 35 percent for crimes of violence. (See reference [4]. The remaining 77 percent or 65 percent of the variation is due to differences in the rates which would be measured by a complete census of the PSU. This variation could in principle be predicted from characteristics of the PSU. This "maximum explainable" proportion of the variance is referred to as the "maximum R^2" in Figure 1.

The Model Selection Process

Models were selected for the two broad NCS crime categories, total personal victimizations and total crimes against households, and the two subcategories of greatest interest, personal crimes of violence and the household crime of burglary.

The models were originally fit using data from the 40 NSR sample PSUs in the states of Pennsylvania, Ohio, Michigan, and Texas. These states were selected because they had a large number of NSR sample PSUs and because both SMSAs and rural PSUs were well represented.

The best model for each crime variable was selected by stepwise regression from a list of about 40 potential stratifiers. The BMDP stepwise regression program was used. Attempts were made to improve the best-fitting model by experimenting with transformations of the selected variables.

Models were tested by randomly splitting the data set into two sets of 20 PSUs each, fitting the model to each half and comparing the coefficients. Models in which these differences were statistically significant or seemed extreme were rejected in favor of more stable models with slightly worse fit. Similarly, models were tested for stability by the removal of obvious outliers and by "trimming" of extreme observations.

In some cases, even when the best-fitting model was stable, alternatives were sought if the best-fitting model included variables which did not make sense as predictors of crime or if they differed drastically from the models for other crimes.

The final stage of the model selection procedure consists of fitting models using the independent variables from the best four-state model to data for 83 NSR PSUs from the southern census region (excluding Texas). This will give a better indication of the true fit of the models, since we will not be selecting the best of many models for these PSUs. This final stage has not been completed.

Figure 1 shows the selected model for total crimes of violence.

FIGURE 1

$$CVIOL = .035 - .908\,OWNOC35\text{-}50 - .149\,NHS16\text{-}21 + .085\,U50\text{-}54 + .075\,ROC5+$$

$$(\sigma = .288) \quad (\sigma = .047) \quad (\sigma = .025) \quad (\sigma = .041)$$

$$R^2 = .44 \quad (MAX\ R^2 = .65)$$

CVIOL = rate per person of crimes of violence
OWNOC35-50 = proportion of occupied housing units which are owner-occupied and valued between $35,000-$49,999
NHS16-21 = proportion of persons age 16-21 who are not in school and are not high school graduates
U50-54 = proportion of occupied housing units built between 1950-54
ROC5+ = proportion of occupied housing units which are in renter-occupied structures with 5 or more units

Conclusions

A question which arose at the beginning of the study was whether a study like this one should be used merely to give general guidance in the selection of a more or less traditional set of stratifiers consisting of 1980 census characteristics, or whether the predicted crime rates from the models could be used as stratifiers.

Using the latter approach, two PSUs might be stratified together if they have different characteristics which are combined by the regression equation to give nearly equal predicted crime rates. The question is really whether the fit and validity of the model are sufficiently trustworthy to apply it in different areas and for different time periods. Because of the mediocre fit and our lack of understanding of the meaning of some variables in the models, it seems wise not to place excessive reliance on the model. Our conclusion is that more traditional stratifiers should be used, including the kinds of characteristics which are important in the models.

The results of the study indicate the kinds of characteristics which make the best stratifiers for NSR PSUs. All models except Model 1 for crimes of violence involve at least one variable which may be interpreted as a measure of urbanization or population density. This includes U1/2P, ROC50, CROWD1, PRUR, ROC5-49, ROC5+, CROWD2, U6+P. Most involve a measure of income or value of housing: PCINC, OWNOCHI, or OWNOC35-50. It seems, however, that income and value of housing are differently related to crime rates. The model for burglary involves two variables which relate to age and transiency of the neighborhood: PRSH and U1950. Only NHS16-21 and PR16-21 describe directly the socio-economic or demographic characteristics of the individuals living in the housing units, and the sign of the coefficient is difficult to interpret in these cases.

Variables relating to education level, racial composition, poverty status, etc., showed no signs of being useful stratifiers at the PSU level in NSR areas. This conflicts both with conventional wisdom and with NCS national estimates. The introduction to this section mentions several possible explanations why PSU-level data for NSR areas may not show associations which are present at the household level for all areas. Which, if any, of these explanations are correct needs to be determined by further study.

IV. UCR CRIME RATES AS STRATIFIERS

This is treated as a separate topic because UCR rates are not available in a form compatible with NCS's PSU definitions. Since UCR rates will be much less convenient than the more compatible census variables, they would only be used as stratifiers if they substantially improved the prediction of NCS rates. It is known that criminal incidents (one incident may have several victims) as measured by NCS are usually much higher than comparable UCR rates, even after the latter has been inflated by dividing by NCS estimates of the proportion of crimes which have been reported to police. In spite of this discrepancy, UCR rates might still have proved to be good stratifiers if the ratio between NCS and UCR rates were relatively constant from PSU to PSU. For this study, county level UCR rates were taken from references [5] and [6], and used to find UCR rates for NCS PSUs.

For all four NCS crime categories considered, NCS rates were positively correlated with the closest corresponding UCR crime rates. (UCR crime categories differ from the NCS categories. In particular, the UCR does not make the distinction between personal and household crimes.) The comparisons were:

(a) NCS total personal victimization vs. UCR total crimes,
(b) NCS crimes of violence vs. UCR assault plus robbery,
(c) NCS total household crimes vs. UCR total crimes, and
(d) NCS burglary vs. UCR burglary.

Only in cases (a) and (d) were the correlations significant at the .05 level.

Both 1976 and 1972 UCR rates were compared to 1976 NCS rates. Surprisingly the 2 years did about equally well; in some cases the 1972 rates were more highly correlated.

In only two of the four regression models (total personal victimizations and burglary) the UCR rates made a significant improvement over the best demographic model for the four-state data, and even then the improvement was slight. There are several possible explanations for this disappointing result:

(1) The differences between the NCS and UCR crime definitions may be partially at fault. This is supported by the result that the greatest relationship was for burglary, where the correspondence between the two definitions is closest.

(2) For reasons of cost, in this study no effort was made to check the correspondence of geographic units as described in reference [5,6] to the NCS PSUs. In some cases, several police agencies reported crimes for the same PSU. We may not have handled this correctly in all cases when our data file was prepared. If this is the problem, use of UCR rates as stratifiers will require close consultation with the FBI staff and detailed individual consideration of all PSUs in creating a data file.

(3) There may be biases in either or both crime series which vary from PSU to PSU. Differences between police jurisdictions in rates of reporting crime to police or in police record-keeping procedures may conceivably affect UCR crime rates. Differences between interviewers in different PSUs, whose behavior may reduce or inflate reports by victims, may affect NCS rates.

Because of the limitations of this study, no explanations of the discrepancies between the series can be tested. To more fully investigate the discrepancies a future study would have to:

(i) verify the PSU-level UCR rates more carefully
(ii) include SR areas
(iii) include all NSR sample PSUs
(iv) investigate the most different PSUs to try to reconcile and explore the discrepancies

Regardless of the cause of the discrepancy, UCR incident rates cannot readily be used as NCS stratifiers without further study.

* * * * * * * *

Definition of Variables

ROC5+ = proportion of occupied housing units which are in renter-occupied structures with 5 or more units

PRUR = proportion of persons residing in urban areas

OWNOCHI = proportion of occupied housing units which are owner-occupied and valued $50,000 or more

PRSH = proportion of persons age 5 or more residing in the same house in 1965 and 1970

OWNOC35-50 = proportion of occupied housing units which are owner-occupied and valued between $35,000-$49,999

PR16-21 = proportion of persons age 16-21

ROC5-49 = proportion of occupied housing units which are in renter-occupied structures with 5-49 units

CROWD2 = number of persons residing in urban areas divided by the number of urban occupied and vacant housing units

PCINC = per capita income, measured in thousands

REFERENCES

[1] U.S. Department of Commerce, Bureau of the Census. "The Current Population Survey: Design and Methodology," by Robert Hanson. Technical Paper 40. Superintendent of Documents No.: C2.2.2:40; January 1978.

[2] Bateman, David and Bettin, Paul. "Standard Error Estimation for the National Crime Survey," Presented at the Annual Meeting of the American Statistical Association, August 1975.

[3] "National Academy of Sciences, Panel for the Evaluation of Crime Surveys, Surveying Crime," Bettye K. Eidson Penick, Editor, 1976.

[4] Alexander, Charles H. "NCS Redesign: Upper Bound on R^2 in the Stratification Variables Study." Bureau of the Census internal memorandum for Rajendra P. Singh from Charles H. Alexander, May 11, 1981.

[5] U.S. Department of Justice, Federal Bureau of Investigation. "Crime in the United States 1976, Uniform Crime Reports," issued by Clarence M. Kelley, Director, FBI.

[6] U.S. Department of Justice, Federal Bureau of Investigation. "Crime in the United States 1972, Uniform Crime Reports," issued by Clarence M. Kelley, Director, FBI.

U 1950 = proportion of occupied housing units built before 1950

U6+P = proportion of occupied housing units containing 6 or more persons

Appendix A

Variances of D_k

$\text{Var}(\hat{V}_i) = \delta_i N_i \text{ SI } p_i (1-p_i)$, where δ_i is the design effect for PSU i, N_i is the population of the stratum containing PSU i, and SI is the overall sampling interval. The design effects are assumed to be the same for all PSUs.

$$\text{Var}(D_k) = \delta\left(\pi_{k2}^2 N_{2k-1} \text{ SI } p_{2k-1} (1-p_{2k-1}) + \pi_{k1}^2 N_{2k} \text{ SI } p_{2k} (1-p_{2k})\right)$$

$$= \delta\, p_{2k} (1-p_{2k}) \text{ SI}\left(\pi_{k2}^2 N_{2k-1} + \pi_{21}^2 N_{2k}\right) \text{ assuming } p_{2k-1} = p_{2k}.$$

p_{2k} is estimated by $\hat{p}_{2k} = \left(\hat{V}_{2k-1} + \hat{V}_{2k}\right) / \left(\hat{N}_{2k-1} + \hat{N}_{2k}\right)$. SI=1104 for the 1976 victimizations.

The design effect will be estimated as follows. Let $\hat{V}_i^{(1)}$ and $V_i^{(2)}$ be the two Keyfitz half-samples for PSU i.

$$E\left(\hat{V}_i^{(1)} - \hat{V}_i^{(2)}\right) = \text{Var}(\hat{V}_i) = \delta_i N_i \text{ SI } p_i (1-p_i)$$

$$E\left(\sum_{i=1}^{220}\left(\hat{V}_i^{(1)} - \hat{V}_i^{(2)}\right)\right) = \delta \text{ SI} \sum_{i=1}^{220} N_i p_i (1-p_i)$$

$$\delta = E\left(\sum_{i=1}^{220}\left(\hat{V}_i^{(1)} - \hat{V}_i^{(2)}\right)\right) \Big/ \left(\text{SI} \sum_{i=1}^{220} N_i p_i (1-p_i)\right)$$

p_i will be estimated by $\hat{V}_i / \hat{N}_i$. N_i will be estimated by $\hat{N}_i$. $\sum_{i=1}^{220}\left(\hat{V}_i^{(1)} - \hat{V}_i^{(2)}\right)$ is the NSR contribution to 1976 within-PSU variance, and is available from 1976 variance estimates.

Appendix B

Frequency Histogram for
Standardized Differences for
Total Personal Victimizations