An Alternative Strategy for Estimating the Parameters of Local Areas

Steven B. Cohen, University of North Carolina at Chapel Hill
William D. Kalsbeek, Research Triangle Institute, North Carolina

1. Introduction

The ever-growing need for good estimates of the social, political, economic, and health parameters has been rapidly gaining recognition. The allocation of federal aid to both states and municipalities is often dependent upon information pertaining to population, unemployment, and housing. Candidates vying for political office are particularly concerned with obtaining reliable estimates of voter preference and participation at the sub-national level. Similarly, rather precise small area estimates of retail trade are essential indicators for the commercial sector.

Some useful information has been obtained from sources which include the decennial census and vital registration systems. Generally, federal agencies have relied upon sample surveys to provide estimates of the data they require, though such estimates pertain to the entire United States or each of its four broad geographical regions. Estimates of data for small areas are unavailable primarily due to sample size requirements which are prohibitive with respect to cost and strata designs which often cross state and county limits. Consequently, several procedures have been developed which utilize available data from large areas, local data on population and accessible local data on ancillary (symptomatic) variables, in order to produce synthetically the desired estimates. Synthetic estimation is perhaps the most well known, defined by the United States Bureau of the Census as "the method of reference to a standard national distribution." Gonzalez (1974) has offered a more comprehensive explanation - "An unbiased estimate is obtained from a sample survey for a large area; when this estimate is used to derive estimates for subareas on the assumption that the small areas have the same characteristics as the larger area, we identify these estimates as synthetic estimates." Developed at the National Center for Health Statistics, the method was initially used to provide synthetic state estimates of disability from the results of the National Health Interview Survey (H.I.S.).

Procedurally, a number of demographic variables are selected (i.e., race, income, sex, age), and when possible, national sample surveys are used to determine estimates of a characteristic (criterion variable) of interest for each of the G mutually exclusive and exhaustive domains defined by the respective demographic cross-classifications. To produce the synthetic estimate of a criterion variable (Y) for local area ℓ, the NCHS model takes the form of a weighted average

$$Y^*_\ell = \sum_{j=1}^{G} P_{\ell j}\, Y_{\cdot j} \qquad (1.1)$$

where $P_{\ell j}$ is the proportion of local area ℓ's population represented by domain j so that $\sum_j^G P_{\ell j} = 1$, and $Y_{\cdot j}$ is the probability estimate of the criterion variable for domain j obtained from a national sample. The more detailed estimating equation includes a regional adjustment.

Considering the underlying model's structure, the synthetic estimates are biased. A popular measure used to assess their reliability is the average mean squared error (M.S.E.)

$$E[1/N \sum_{i=1}^{N} (Y^*_\ell - Y_\ell)^2]$$

calculated over all N local areas defined by the survey population. Gonzalez and Waksburg (1973) have derived an approximation for this expression, assuming that

i) the $P_{\ell j}$'s are fixed and measured without error, and
ii) the $\mathrm{Cov}(Y_{\cdot j}, Y_{\cdot k}) = 0$ for $j \neq k$.

Due to the nature of their derivation, the synthetic estimates will generally cluster near the mean for a specific geographic region. Consequently, the method is not particularly sensitive to many of the internal forces operating at the local level. By assuming the small areas share the same characteristics as a standard national distribution, they can only be distinguished by their respective demographic configurations. Recognizing this inherent limitation, Levy (1971) proposed a method which utilized available information at the local level on predictor (symptomatic) variables in conjunction with the NCHS estimator. The following model was considered:

$$Y^{**}_\ell = \alpha + \beta X_\ell + \varepsilon_\ell \qquad (1.2)$$

where X_ℓ is the value of the symptomatic variable for the ℓ^{th} subarea,

$$Y^{**}_\ell = (Y_\ell - Y^*_\ell)/Y^*_\ell \times 100$$

where ε_ℓ is a term representing random error, and α and β, regression coefficients to be estimated. Here, the percentage difference between the synthetic estimate and the true value is treated as a linear function of some related predictor variable X_ℓ. Were the estimates $\hat\alpha$ and $\hat\beta$ available and ε_ℓ omitted, an estimator $\hat Y_\ell$ of Y_ℓ could be derived from (1.2), taking the form:

$$\hat Y_\ell = Y^*_\ell[(\hat\alpha + \hat\beta X_\ell)/100 + 1] \qquad (1.3)$$

It is assumed that X_ℓ is available for every local area, but since Y^{**}_ℓ is a function of the true value Y_ℓ (which is unknown), a different strategy is used to estimate the linear coefficients. Briefly, α and β are estimated by least squares after combining local areas to form strata. The method can be extended to consider X_ℓ as a vector of symptomatic data, whereby $\hat Y_\ell$ is treated as a multiple regression estimator.

Ericksen (1974) developed another technique for computing local area estimates which, unlike the NCHS estimator, solely combines symptomatic information and sample data into a multiple regression format (assuming an underlying linear model). Referred to as the regression-sample data of local area estimation, the procedure can be outlined as follows:

1. Initially, a sample of n local areas, referred to as primary sampling units (PSU's), is selected from the N local areas in the population. Estimates of the criterion variable are then computed for the respective PSU's in the sample.
2. Collect symptomatic information for both sample and non-sample PSU's. Typical predictor variables are the number of births, deaths, and school enrollment.
3. Compute the linear least squares regression estimate using data for the sample PSU's only. Estimates for all subareas are then determined by substituting values of the symptomatic indicators, whether included in the respective sample or not.

The model assumes the availability of criterion variable estimates for each of n sample PSU's and the values of p symptomatic indicators for the universe of N local areas. It takes the matrix representation:

$$Y = XB + u \tag{1.4}$$

where Y, an n×1 vector, is the criterion variable consisting of a set of actual unobserved values; X, an n×(p+1) matrix denoting the set of predictor variables;

B, the (p+1)×1 vector of regression coefficients; and

u, an n×1 vector, a stochastic error term.

Under the assumption of linearity, B could be estimated by ordinary least squares regression were the Y values observed. Because the individual observations of Y are affected by sampling variability, the model may be revised to explain the within-PSU sampling error in the following manner:

$$Y_0 = XB + u + v \tag{1.5}$$

where v is an n×1 vector of sampling error deviations and Y_0 the observed values.

The regression equation is then computed, substituting the observed values of Y_0 for Y. Hence, the regression coefficients are unbiased in the absence of correlations between v and Y. The mean square error of the regression estimates is expressed as:

$$E(Y-\hat{Y})'(Y-\hat{Y})/n = [(n-p-1)\sigma_u^2/n] + [(p+1)\sigma_v^2/n] \tag{1.6}$$

where σ_u^2 is the between-PSU variance unexplained by the predictor variables, and σ_v^2 is the within PSU variance.

This method was tested for counties and states using 1970 census data on population growth. The resulting estimates were found to be more accurate than estimates computed by standard demographic procedures for the same period.

2. An Alternative Strategy

2.1 Methodology

The method advanced by Ericksen is most feasible when the linearity assumption is satisfied and the observed multiple correlation is high. But what decision is reached when the multiple correlation level is moderate (.5-.8) and a non-linear model is more suitable? The inclusion of all possible symptomatic variables into the regression would increase the R^2 but most probably at the expense of an "over-fit" model which increases the mean square error of the final estimate. More generally, in those situations where assumptions are too strict or unrealistic, the need for a more flexible approach is most obvious. Kalsbeek (1973) has developed one such procedure in which the most limiting assumption is the availability of good symptomatic information.

It has usually been common practice to treat the local area units as the smallest level for which the estimates are made. Contrarily, Kalsbeek suggests breaking up the local unit into constituent geographical sectors called "base units," such as townships, enumeration districts, or other geographical subunits of a county. The local area for which a variable of interest is to be estimated is referred to as the "target area" and further subdivided into "target area base units." Unlike other methods which use symptomatic information directly for the purposes of estimation, this procedure uses the information to group base units (sample base units) from the total population. The symptomatic information is also used to classify "target area base units" into the appropriate group.

Initially, a random sample of n base units is selected from the total population of $\dot{N}$ base units. The sample base units (possibly including some "target area base units") are required to possess both symptomatic and criterion information. These units are divided into K groups (strata) using either or both types of the information available. The object is to form groups which are most homogeneous within while dissimilar between themselves. Grouping can be handled be any one of several iterative procedures in cluster analysis (i.e., Automatic Interaction Detection (A.I.D.), Multivariate Iterative K-Means Cluster Analysis (MIKCA)). It is noteworthy that the respective groups may be defined by either rectilinear or non-rectilinear boundaries.

All "target area base units" belonging to the local area in question are then assigned (classified) to one of the K groups with respect to symptomatic information. Consequently, each "target area base unit" is associated with a group of base units both similar to itself and internally homogeneous. An estimate for each of the "target area base units" with respect to the criterion variable is obtained from the sample base units in the group to which it has been assigned. These estimates are then pooled to arrive at a final estimate for the respective target area.

2.2 Notation

Consider a population consisting of L local areas, indexed by $\ell=1, 2, \ldots, L$, which have further been subdivided into constituent geographical sectors called "base units." There are $_{\ell}N_{.}$ base units in the ℓ^{th} local area, and

$$\sum_{\ell=1}^{L} {}_{\ell}N_{.} = {}_{.}N_{.}$$

in the population, individually indexed by $i=1, 2, \ldots, {}_{\ell}N_{.}$, to denote the i^{th} base unit from the ℓ^{th} local area. When the local area reference is dropped, each base unit is indexed by $i=1, 2, \ldots, {}_{.}N_{.}$. Furthermore, each base unit i consists of a cluster of M_i smaller units referred to as elements. Hence, there are

$${}_{\ell}M_{.} = \sum_{i=1}^{{}_{\ell}N_{.}} M_i$$

elements in the ℓ^{th} local area and

$${}_{.}M_{.} = \sum_{\ell=1}^{L} {}_{\ell}M_{.} = \sum_{i=1}^{{}_{.}N_{.}} M_i$$

elements in the population. Let y_{ij} represent the observed value of the criterion variable for the j^{th} element within the i^{th} base unit, where

$$Y_i = \sum_{j=1}^{M_i} y_{ij}$$

is the i^{th} unit total.

In practice, a multi-stage sampling design is most appropriate. To facilitate the presentation, we assume a two-stage sampling design whereby a simple random sample of n base units (first stage units) is initially drawn from the ${}_{.}N_{.}$ base units in the population. A subsample of m_i out of the M_i elements is then selected with equal probabilities of selection from each of the chosen sample base units. Here, the subunits are chosen independently in different units. The units are then divided into K groups (strata), indexed by $g=1, 2, \ldots, K$, by one of the aforementioned procedures (Section 2.1). Consequently, estimates of the group means are obtained by a method which most closely resembles post-stratification. To determine the criterion variable estimator for the ℓ^{th} local area, each "target base unit" is assigned to the group most similar with respect to symptomatic information. Thus, we have a two-way classification of all base units in the population by respective strata and local areas, where ${}_{\ell}N_g$ is the total number of base units in the g^{th} strata from the ℓ^{th} local area.

2.3 Representation of the Model

The local area estimator of the criterion variable may be expressed in terms of an average, a proportion, or a total. Initially, we direct attention to the mean per element representation.

Assuming a two-stage sampling design with sub-units of unequal sizes, we define

$$\bar{y}_i = \sum_{j=1}^{m_i} \frac{y_{ij}}{m_i}$$

as the sample mean per element in the i^{th} base unit and

$$\bar{Y}_i = \sum_{j=1}^{M_i} \frac{y_{ij}}{M_i}$$

as the overall mean per element in the i^{th} base unit. To obtain an estimate of the g^{th} stratum mean per element, we also define the indicator variables I_{gi} (once more dropping the local area reference), such that

I_{gi} = 1 if the (first stage) base unit falls in the g^{th} stratum;

= 0, otherwise

for $g=1, 2, \ldots, K$ and $i=1, 2, \ldots, {}_{.}N_{.}$. Here, $\sum_{i=1}^{n} I_{gi} = n_g$, the number of sample base units belonging to the g^{th} stratum, and

$$\sum_{i=1}^{{}_{.}N_{.}} I_{gi} = {}_{.}N_g \quad .$$

Consequently, let

$$\hat{\bar{\bar{y}}}_g = \frac{\sum_{i=1}^{n} I_{gi} M_i \bar{y}_i}{\sum_{i=1}^{n} I_{gi} M_i} = \frac{\sum^{n_g} M_i \bar{y}_i}{\sum^{n_g} M_i}$$

(summed only over the n_g sample base units from the g^{th} stratum) be our (post-stratified) estimator of the g^{th} stratum mean per element. To facilitate the presentation, we assume the values of M_i in the sample are known. Since $\hat{\bar{\bar{y}}}_g$ is a ratio estimator of

$$\bar{\bar{Y}}_g = \frac{\sum_{i=1}^{{}_{.}N_{.}} I_{gi} M_i \bar{Y}_i}{\sum_{i=1}^{{}_{.}N_{.}} I_{gi} M_i} = \frac{\sum^{{}_{.}N_g} M_i \bar{Y}_i}{{}_{.}M_g}$$

(where the sum is over the ${}_{.}N_g$ base units assigned to the g^{th} stratum), it is biased to the order of 1/n. Yet, when n is large (i.e., $n \geq 100$), the bias is negligible and the expectation of $\hat{\bar{\bar{y}}}_g$ is approximately equivalent to $\bar{\bar{Y}}_g$,

$$E(\hat{\bar{\bar{y}}}_g) \doteq \bar{\bar{Y}}_g \ , \ g = 1, 2, \ldots, K \quad .$$

Returning to the ℓ^{th} local area, we focus attention on the "target base unit" alignment in order to weight appropriately the stratum estimators $(\hat{\bar{\bar{y}}}_g)$ by the proportion of base units so classified. Therefore, the estimator of the criterion variable for the ℓ^{th} local area takes the following form:

$${}_{\ell}\hat{\bar{\bar{y}}}^{*} = \sum_{g=1}^{K} \frac{{}_{\ell}M_g}{{}_{\ell}M_{.}} \hat{\bar{\bar{y}}}_g \qquad (2.3.1)$$

such that

$$E({}_\ell\hat{\bar{\bar{y}}}^*) = \sum_{g=1}^{K} \frac{{}_\ell M_g}{{}_\ell M_.} E(\hat{\bar{\bar{y}}}_g) \doteq \sum_{g=1}^{K} \frac{{}_\ell M_g}{{}_\ell M_.} \bar{\bar{Y}}_g$$

when n is large. Often the sizes of ${}_\ell M_g$ and ${}_\ell M_.$ are only known approximately. When this occurs, the respective estimators of the strata means are weighted by the ratio of available estimates ${}_\ell M'_g$ and ${}_\ell M'_.$, or by the cruder ratio ${}_\ell N_g / {}_\ell N_.$.

Due to the nature of its derivation, the local area estimator ${}_\ell\hat{\bar{\bar{y}}}^*$ of ${}_\ell\bar{\bar{Y}}$ is biased. The observed value of the criterion variable mean per element is

$${}_\ell\bar{\bar{Y}} = \frac{\sum^{{}_\ell N_.} M_i \bar{Y}_i}{\sum^{{}_\ell N_.} M_i} = \frac{\sum^{{}_\ell N_.} M_i \bar{Y}_i}{{}_\ell M_.}$$

summed across only those base units in the ℓ^{th} local area. The bias,

$$B = [E({}_\ell\hat{\bar{\bar{Y}}}^*) - {}_\ell\bar{\bar{Y}}]$$

can be approximated by

$$B' = [\sum_{g=1}^{K} \frac{{}_\ell M_g}{{}_\ell M_.} \bar{\bar{Y}}_g - \frac{\sum^{{}_\ell N_.} M_i \bar{Y}_i}{{}_\ell M_.}]$$

Similarly, to express the local area estimator in terms of a proportion, y_{ij} is redefined, so that

y_{ij} = 1 when the j^{th} element in the i^{th} base unit has the characteristic of interest;

= 0 otherwise,

so that

$$\sum_{j=1}^{M_i} y_{ij} = Y_i$$

is the total number of elements in the i^{th} base unit with the characteristic of interest. Model (2.3.1) can then be used.

2.4 An Expression for the Mean Squared Error of the Local Area Estimator

It has already been observed that the local area estimator ${}_\ell\hat{\bar{\bar{y}}}^*$ is biased. Consequently, the mean squared error term takes the form:

$$E[({}_\ell\hat{\bar{\bar{y}}}^* - {}_\ell\bar{\bar{Y}})^2] = E({}_\ell\hat{\bar{\bar{y}}}^* - E({}_\ell\hat{\bar{\bar{y}}}^*))^2 + (E({}_\ell\hat{\bar{\bar{y}}}^*) - {}_\ell\bar{\bar{Y}})^2$$
$$= Var({}_\ell\hat{\bar{\bar{y}}}^*) + Bias^2 \quad . \qquad (2.4.1)$$

By assuming

$$E({}_\ell\hat{\bar{\bar{y}}}^*) \doteq \sum_{g=1}^{K} \frac{{}_\ell M_g}{{}_\ell M_.} \bar{\bar{Y}}_g \quad ,$$

where ${}_\ell\hat{\bar{\bar{y}}}^*$ is a linear combination of the ratio estimators $\hat{\bar{\bar{y}}}_g$, g=1, 2, ..., K with neglible bias, the variance of ${}_\ell\hat{\bar{\bar{y}}}^*$ can be approximated by $Var({}_\ell\hat{\bar{\bar{y}}}^*)$ =

$$\sum_{g=1}^{K} (\frac{{}_\ell M_g}{{}_\ell M_.})^2 Var(\hat{\bar{\bar{y}}}_g) + \sum_{g \neq g'}\sum (\frac{{}_\ell M_g}{{}_\ell M_.})(\frac{{}_\ell M_{g'}}{{}_\ell M_.}) Cov(\hat{\bar{\bar{y}}}_g, \hat{\bar{\bar{y}}}_{g'})$$

If we also assume

$$\frac{\sum_{i=1}^{n} I_{gi} M_i \bar{y}_i}{\sum_{i=1}^{n} I_{gi} M_i} - \bar{\bar{Y}}_g \doteq \frac{\sum_{i=1}^{n} I_{gi} M_i (\bar{y}_i - \bar{\bar{Y}}_g)}{n(\frac{M_{.g}}{N_{..}})} \quad ,$$

then

$$Var(\hat{\bar{\bar{y}}}_g) \doteq \frac{(N_{..} - n)}{n N_{..}} (\frac{N_{..}^2}{M_{.g}^2}) \frac{\sum_{i=1}^{N_{..}} I_{gi}^2 M_i^2 (\bar{Y}_i - \bar{\bar{Y}}_g)^2}{(N_{..} - 1)}$$
$$+ \frac{N_{..}}{n M_{.g}^2} \frac{\sum_{i=1}^{N_{..}} I_{gi}^2 M_i^2}{m_i} (1 - \frac{m_i}{M_i}) \frac{\sum_{j=1}^{M_i} (y_{ij} - \bar{Y}_i)^2}{(M_i - 1)}$$

This is the standard form of the approximate variance of ratio estimator for a two-stage sampling design where the base units have equal probabilities of selection. Here, the first term represents the between base unit component of the variance, whereas the second denotes the within-base unit contribution. A nearly unbiased sample estimate of $Var(\hat{\bar{\bar{y}}}_g)$ takes the form:

$$var(\hat{\bar{\bar{y}}}_g) = (\frac{N_{..} - n}{N_{..}\, n})(\frac{N_{..}^2}{M_{.g}^2}) \frac{\sum_{i=1}^{n} I_{gi}^2 M_i^2 (y_i - \hat{\bar{\bar{y}}}_g)^2}{(n-1)}$$
$$+ \frac{N_{..}}{n M_{.g}^2} \frac{\sum_{i=1}^{n} I_{gi}^2 M_i^2}{m_i} (1 - \frac{m_i}{M_i}) \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{(m_i - 1)}$$

Since our sampling design requires the independent selection of subsamples from different sample base units, and the respective strata estimators are defined in terms of the indicator variable I_{gi}, it can also be shown that $Cov(\hat{\bar{\bar{y}}}_g, \hat{\bar{\bar{y}}}_{g'}) \doteq 0$. Hence, the mean squared error of our small area estimator can be expressed as:

$$MSE({}_\ell\hat{\bar{\bar{y}}}^*) \doteq \sum_{g=1}^{k} (\frac{{}_\ell M_g}{M_{.g}})^2 Var(\hat{\bar{\bar{y}}}_g) + (Bias)^2$$

3. An Illustrative Example

The availability of Census data on population and per capita income for 1970 allowed for an examination of the method's accuracy. State estimates of population growth (from 1960-1970) and per capita income were generated by Kalsbeek, using the Current Population Survey (a national multi-stage probability sample of the U.S. conducted monthly) as the source of sample information. Here, the sample base units correspond to the first stage primary sampling units (PSU's) in the C.P.S., which are counties or groups of contiguous counties. The symptomatic variables considered when estimating population growth include total school enrollment, live births, and deaths, all expressed in ratio form (1970 total/1960 total). Those considered in the per capita income example include the percent natural increase in population between 1960 and 1965, the 1960 per capita aggregate income, and the 1964 percent of the population on public assistance (all obtained from the 1967 County-City Data Book).

The grouping of the sample base units (PSU's) was done using the Automatic Interaction Detector, version II (AID II), which is essentially a clustering algorithm that uses both symptomatic and criterion variable information. Since the respective groups (strata) formed have rectilinear boundaries, the "target base units" (here counties) are assigned to the group whose boundaries include the observation's symptomatic values. Hence, each target base unit takes on the group estimate of the criterion variable to which it is assigned. For the ℓ^{th} state, one considers the respective target base unit alignment, and weights the group (strata) estimators ($\hat{\hat{y}}_g$) by the proportion of the state's population in the target base units so classified. Here, the 1960 county populations were used.

The method was compared with Ericksen's procedure since both are applicable under essentially the same circumstances. The criterion for measuring the accuracy of the estimates was the relative absolute deviation from the true value: $\left|\frac{\text{Estimated-True}}{\text{True}}\right|$. Ericksen's procedure would be expected to give better results for the population growth example due to the inclusion of three symptomatic variables with a high level of multiple correlation and an underlying linear relationship. Still, the proposed method yielded more accurate estimates for more than 25 percent of the states considered (11 out of 42). It was observed that the results tended to improve with increases in population size for both methods. The proposed method did much better in generating state estimates of per capita income, yielding more accurate results in 29 of the 47 states considered. In general, the proposed method produced better results with moderate per capita income states, while Ericksen's approach was more successful at the extremes.

4. Summary

Reliable estimates at the local level are generally difficult, if not impossible, to obtain from sample surveys, primarily due to the constraints of sample size and design. Yet, the very nature of the problem has served as the motivating force in the development of several alternative procedures. The strategy suggested here offers a quick though not consistently clean method of generating the desired estimates. Here, a trade-off exists between the considerations of cost and accuracy. Generally, one is willing to sacrifice a degree of exactness when confronted with the harsh realities of limited resources. This strategy is particularly attractive in that no particular functional model between the criterion and symptomatic variables must be specified. Estimates for the base units of the "target areas" are available as a by-product of the technique. Finally, the method performs reasonably well even for a linear setting, though here it would be better to choose Ericksen's approach.

ACKNOWLEDGEMENTS

The research was partially supported by the National Institute of Child Health and Human Development (Grant HD-00371) and the U.S. Bureau of the Census (JSA-76-93). The authors wish to thank Gary G. Koch for his discussions of the paper and Jean McKinney for her conscientious typing of this manuscript.

REFERENCES

Cochran, W. G. Sampling Techniques. New York: John Wiley and Sons, 1963.

Ericksen, E. P. "A Regression Method for Estimating Population Changes of Local Areas." Journal of the American Statistical Association, 69 (1974), 867-875.

Gonzalez, M. E. and Waksberg, J. "Estimation of the Error of Synthetic Estimates." Presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria, 1973.

Gonzalez, M. E. "Use and Evaluation of Synthetic Estimates." American Statistical Association, Proceedings of the Social Statistics Section (1975).

Gonzalez, M. E. and Hoza, C. "Small Area Estimation of Unemployment." American Statistical Association, Proceedings of the Social Statistics Section (1975).

Kalsbeek, W. D. "A Method for Obtaining Local Postcensal Estimates for Several Types of Variables." Unpublished doctoral dissertation, University of Michigan (1973).

Koch, G. G. "An Alternative Approach to Multivariate Response Error Models for Sample Survey Data with Applications to Estimators Involving Subclass Means." Journal of the American Statistical Association, 68 (1973), 328-331.

Levy, P. S. "The Use of Mortality Data in Evaluating Synthetic Estimates." American Statistical Association, Proceedings of the Social Statistics Section (1971), 328-331.

U. S. Bureau of the Census. "Federal State Cooperative Program for Local Population Estimates: Test Results - April 1, 1970." Current Population Reports, P-26, No. 21, 1973.

U. S. Bureau of the Census. "Coverage of the

Population in the 1970 Census and Some Implications for Public Programs." Current Population Reports, P-23, No. 56, 1975.

U. S. National Center for Health Statistics. Synthetic Estimates of Disability, PHS publication, No. 1759, 1968.