OPTIMUM SAMPLE ALLOCATION TO DIFFERENT MODES OF ENUMERATION

Kirk M. Wolter, U.S. Bureau of the Census, David A. Pyne, The Johns Hopkins University

1. Introduction

One of the principal objectives of survey design is the construction of a sampling and data collection system which is as efficient as possible, within budgetary and other constraints, for estimating the main parameters of the survey. Many trade-offs are typically involved in attempting to achieve this objective. For example, in planning a household survey, the statistician must decide whether it is more efficient to invest a portion of the survey budget on rigorous interviewer training (thus attempting to reduce measurement errors) or to use the entire budget on an increased sample size (thus reducing the sampling error of the survey). The concept of total survey design (TSD), as defined in Horvitz and Wolter (1975), recognizes the various sampling and nonsampling components of error, and implies an allocation of the total survey resources to the various error components in a manner which minimizes the total error of estimate.

There has been at least one recent effort to formalize the ideas of TSD. Lessler (1975) developed a double sample sampling scheme in which the first stage sample is enumerated using an inexpensive but imperfect measurement method, while the smaller second stage sample is also enumerated by a method that is both costly and accurate. Thus, this scheme invests a portion of the survey budget to reduce or eliminate the estimator bias due to measurement errors.

In this paper, we formalize the ideas of TSD in another direction. The general problem we consider is the following:

> A given survey may be conducted under either of two (or more) sets of general survey conditions, each giving rise to different measurement error distributions, and each being associated with different per unit costs. What is the optimum allocation of the sample to the two conditions? That is, should the entire sample be enumerated under one set of conditions or the other, or should a portion of the sample be enumerated under each of the alternative conditions.

An important example of this problem is where a sample may alternatively be enumerated by telephone or personal interview, with the remaining general conditions the same in both cases.

Section 2 describes the specific allocation problem to be considered. Our technique for modeling measurement errors is also introduced in this section, and then developed in section 3. The technique is a straightforward extension of the Census Bureau model, discussed by Hansen, Hurwitz, and Bershad (1961); Koch, (1973); and others. The optimum allocation results are worked out in section 4. Section 5 discusses the allocation results vis-à-vis a model which incorporates the effects of interviewers, and section 6 closes the paper with some extensions and a general discussion.

2. The Allocation Problem

We assume the finite population is composed of a fixed number N of identifiable units. Attached to each unit j is the true value, say Y_j^o, of a characteristic y. The specific allocation problem we are interested in presumes that it is desired to estimate the population mean

$$\bar{Y}^o = \sum_{j=1}^{N} Y_j^o/N .$$

We assume, however, that the Y_j^o are unobservable, and instead associate an observable bivariate random variable

$$\underset{\sim}{Y}_{tj} = (Y_{tj1}, Y_{tj2})'$$

to each unit $j=1,\ldots,N$. The elements of $\underset{\sim}{Y}_{tj}$ represent erroneous measurements of Y_j^o under two sets of general survey conditions, G_1 and G_2, at time (or trial) t .

We shall assume that a simple random sample of size n is selected without replacement. In turn, this sample is randomly split into two groups, n_1 and n_2, the first being enumerated under conditions G_1 and the second under conditions G_2. Let s_1 and s_2 denote the two samples, and let $s=s_1 \cup s_2$ denote the combined sample. Then,

$$\bar{y}_{t1} = \sum_{j\varepsilon s_1} Y_{tj1}/n_1$$

and

$$\bar{y}_{t2} = \sum_{j\varepsilon s_2} Y_{tj2}/n_2$$

denote the sample means of the first and second samples, respectively.

We shall consider two members of the class of estimators which are defined by a weighted combination of $\bar{y}_{t1}$ and $\bar{y}_{t2}$. The first estimator

$$\bar{y}_t(1) = (n_1/n)\bar{y}_{t1} + (n_2/n)\bar{y}_{t2} \qquad (2.1)$$

weights according to the sample size, while the second estimator

$$\bar{y}_t(2) = \alpha_0\bar{y}_{t1} + (1 - \alpha_0)\bar{y}_{t2} \qquad (2.2)$$

uses the minimum variance weights. If we let v_1^2, v_2^2, and v_{12} denote the variances of and covariance between $\bar{y}_{t1}$ and $\bar{y}_{t2}$, then the minimum variance coefficient is

$$\alpha_0 = (v_2^2 - v_{12})/(v_1^2 + v_2^2 - 2v_{12}).$$

Finally, we shall adopt the simple cost function

$$C = c_1n_1 + c_2n_2 \quad . \qquad (2.3)$$

In this notation, C denotes the total survey budget, c_1 denotes the per unit cost of enumerating the sample under conditions G_1, and c_2 denotes the per unit cost under conditions G_2.

In section 4, we shall work out the optimum values of n_1 and n_2. In this development, the term optimum shall mean those values of n_1 and n_2 which minimize the variance of $\bar{y}_t(1)$ or $\bar{y}_t(2)$ subject to fixed cost C.

3. Measurement Error Model

In preparation for solving the allocation problem, we develop explicit expressions for the variances of $\bar{y}_t(1)$ and $\bar{y}_t(2)$. We assume that the random variables $\underset{\sim}{Y}_{tj}$ have bounded second moments, and let

$$\mu_{j1} = E\{Y_{tj1}|j1\} \quad ,$$

$$\mu_{j2} = E\{Y_{tj2}|j2\} \quad ,$$

$$d_{tj1} = Y_{tj1} - \mu_{j1} \quad ,$$

$$d_{tj2} = Y_{tj2} - \mu_{j2} \quad ,$$

$$\sigma_{j1}^2 = E\{d_{tj1}^2|j1\} \quad ,$$

$$\sigma_{j2}^2 = E\{d_{tj2}^2|j2\} \quad ,$$

$$\sigma_{j1i1} = E\{d_{tj1}d_{ti1}|j1, i1\} \quad ,$$

$$\sigma_{j2i2} = E\{d_{tj2}d_{ti2}|j2, i2\} \quad ,$$

and $\sigma_{j1i2} = E\{(d_{tj1})(d_{ti2})|j1, i2\}$ for $i \neq j = 1,\ldots,N$, where the expectations are taken with respect to the distribution of measurement errors. The means μ_{j1} and μ_{j2} are not necessarily assumed to be equal, nor are they assumed to equal the true value Y_j^o. The covariances σ_{j1i1} and σ_{j2i2} may be nonzero if the two units, i and j, are enumerated by the same interviewer, processed by the same clerk, or otherwise "measured" by a common organization. We would assume that the σ_{j1i2} are zero if different enumeration and Processing teams are employed in the two samples, s_1 and s_2; otherwise, the σ_{j1i2} may also be nonzero.

It can easily be shown that the expectations and variances of $\bar{y}_{t1}$ and $\bar{y}_{t2}$ are

$$E\{\bar{y}_{t1}\} = \bar{M}_1 \quad , \qquad (3.1)$$

$$E\{\bar{y}_{t2}\} = \bar{M}_2 \quad , \qquad (3.2)$$

$$v_1^2 = (1-n_1/N)S_{\mu1}^2/n_1+(\sigma_{d1}^2/n_1)[1+(n_1-1)\rho_{11}], \quad (3.3)$$

$$v_2^2 = (1-n_2/N)S_{\mu2}^2/n_2+(\sigma_{d2}^2/n_2)[1+(n_2-1)\rho_{22}], \quad (3.4)$$

where

$$\bar{M}_1 = \sum_j^N \mu_{j1}/N \quad ,$$

$$\bar{M}_2 = \sum_j^N \mu_{j2}/N \quad ,$$

$$S_{\mu1}^2 = (N-1)^{-1} \sum_j^N (\mu_{j1}-\bar{M}_1)^2 \quad ,$$

$$S_{\mu2}^2 = (N-1)^{-1} \sum_j^N (\mu_{j2}-\bar{M}_2)^2 \quad ,$$

$$\sigma_{d1}^2 = \sum_j^N \sigma_{j1}^2/N \quad ,$$

$$\sigma_{d2}^2 = \sum_j^N \sigma_{j2}^2/N \quad ,$$

$$\sigma_{d1}^2\rho_{11} = N^{-1}(N-1)^{-1} \sum_{i\neq j}^N\sum \sigma_{i1j1} \quad .$$

and

$$\sigma_{d2}^2\rho_{22} = N^{-1}(N-1)^{-1} \sum_{i\neq j}^N\sum \sigma_{i2j2} \quad .$$

For a complete development of these expressions see, e.g., Hansen, Hurwitz, and Bershad (1961). The first term on the right side of (3.3) represents the sampling variance of $\bar{y}_{t1}$, while the second term is the measurement error variance. Similar remarks apply to the two terms on the right side of (3.4). We do not have a covariance (or interaction) between sampling error and measurement error as do some other authors, because we assume the expectations μ_{j1} and μ_{j2} depend only on the unit j and not on the other units in the sample.

To give an expression for the variances of $\bar{y}_t(1)$ and $\bar{y}_t(2)$, it only remains to find the covariance between $\bar{y}_{t1}$ and $\bar{y}_{t2}$. Towards this end, let

$$\bar{\mu}_1 = \sum_{j\varepsilon s_1} \mu_{j1}/n_1 \quad ,$$

$$\bar{\mu}_2 = \sum_{j\varepsilon s_2} \mu_{j2}/n_2 \quad ,$$

$$\bar{d}_{t1} = \bar{y}_{t1} - \bar{\mu}_2 \quad ,$$

and $\bar{d}_{t2} = \bar{y}_{t2} - \bar{\mu}_2$. Then it can be shown that

$$v_{12} = \mathrm{Cov}\{\bar{y}_{t1},\bar{y}_{t2}\} = \mathrm{Cov}\{\bar{\mu}_1,\bar{\mu}_2\}+\mathrm{Cov}\{\bar{d}_{t1},\bar{d}_{t2}\}$$

$$= -S_{\mu1\mu2}/N+\rho_{12}\sigma_{d1}\sigma_{d2}, \quad (3.5)$$

where

$$S_{\mu1\mu2} = (N-1)^{-1} \sum_{j}^{N}(\mu_{j1}-\bar{M}_1)(\mu_{j2}-\bar{M}_2)$$

and

$$\rho_{12}\sigma_{d1}\sigma_{d2} = N^{-1}(N-1)^{-1} \sum_{i\neq j}^{N}\sum \sigma_{j1i2} \quad .$$

From (3.3), (3.4), and (3.5) we obtain the following expressions for the variances of $\bar{y}_t(1)$ and $\bar{y}_t(2)$:

$$V\{\bar{y}_t(1)\} = (n_1/n)^2v_1^2+(n_2/n)^2v_2^2+2(n_1/n)(n_2/n)v_{12} \quad (3.6)$$

$$V\{\bar{y}_t(2)\} = \alpha_0^2v_1^2+(1-\alpha_0)^2v_2^2+2\alpha_0(1-\alpha_0)v_{12} \quad . \quad (3.7)$$

Our preparation is now complete, and in the next section, we minimize the variances subject to fixed cost.

4. Optimum Allocation Results

Throughout this section we assume the sampling fractions n_1/N and n_2/N are negligible. Thus,

$$v_1^2 = S_{\mu1}^2/n_1+(\sigma_{d1}^2/n_1)[1+(n_1-1)\rho_{11}] \quad ,$$

$$v_2^2 = S_{\mu2}^2/n_2+(\sigma_{d2}^2/n_2)[1+(n_2-1)\rho_{22}] \quad ,$$

and

$$v_{12} = \rho_{12}\sigma_{d1}\sigma_{d2} \quad .$$

Our strategy is to first work out the optimum allocation in a general framework and then specialize to a specific case.

In the case of (3.6), weights according to sample size, if the substitution

$$n_1 = \Theta n, \quad n_2 = (1-\Theta)n \quad (4.1)$$

is made, the allocation problem reduces to

$$\min_{n,\Theta}(1/n)[\Theta k_1+(1-\Theta)k_2]+\Theta^2k_3+(1-\Theta)^2k_4+2\Theta(1-\Theta)k_5$$

subject to $n\Theta c_1 + n(1 - \Theta)c_2 \leq C$, and $0 \leq \Theta \leq 1$, where

$$k_1 = S_{\mu1}^2 + \sigma_{d1}^2(1 - \rho_{11}) \quad ,$$

$$k_2 = S_{\mu2}^2 + \sigma_{d2}^2(1 - \rho_{22}) \quad ,$$

$$k_3 = \sigma_{d1}^2\rho_{11} \quad ,$$

$$k_4 = \sigma_{d2}^2\rho_{22} \quad ,$$

$$k_5 = v_{12} \quad . \quad (4.2)$$

It can be shown that for fixed θ the objective function is a decreasing function of n. This implies that the minimum will occur at largest n possible, i.e.

$$n = C/[\Theta c_1+(1 - \Theta)c_2] \quad . \quad (4.3)$$

With the substitution of (4.3) into the objective function, the problem becomes

$$\min_{0\leq\Theta\leq1} \Theta^2d_1 + \Theta d_2 + d_3 \quad ,$$

where

$$d_1 = (c_1-c_2)(k_1-k_2)/C+k_3+k_4-2k_5$$

$$d_2 = c_2(k_1-k_2)/C+k_2(c_1-c_2)/C-2k_4+2k_5$$

$$d_3 = k_4+c_2k_2/C \quad .$$

It is easily seen that the optimal solution is

$$\theta_0 = \begin{cases} -d_2/2d_1 & \text{if } d_1 > 0 \text{ and } 0 < -d_2/2d_1 < 1, \\ 0 & \text{if } d_1 > 0 \text{ and } -d_2/2d_1 \le 0 \\ & \text{or } d_1 < 0 \text{ and } c_1k_1+Ck_3 \ge c_2k_2 \\ & \qquad + Ck_4 \\ & \text{or } d_1 = 0 \text{ and } d_2 > 0 \ , \\ 1 & \text{if } d_1 > 0 \text{ and } -d_2/2d_1 \ge 1 \ , \\ & \text{if } d_1 < 0 \text{ and } c_1k_1+Ck_3 \le c_2k_2 \\ & \qquad + Ck_4 \ , \\ & \text{if } d_1 = 0 \text{ and } d_2 < 0 \ , \\ \text{any value in } [0,1] & \text{if } d_1 = 0 \text{ and } d_2 = 0 \ . \end{cases} \tag{4.4}$$

The special case of uncorrelated measurement errors (i.e. $\rho_{12} = \rho_{11} = \rho_{22} = 0$) is of considerable interest. Given these conditions, $k_3 = k_4 = k_5 = 0$, and it follows that the minimum occurs at

$$\theta_0 = \begin{cases} 1 & \text{if } d_1c_1 < d_2c_2 \ , \\ 0 & \text{if } d_1c_1 > d_2c_2 \ , \\ \underline{\text{any}} \text{ value in } [0,1] & \text{if } d_1c_1 = d_2c_2 \ . \end{cases} \tag{4.5}$$

Thus, the optimal solution in this case will usually be to carry out the entire survey under one set of general conditions or the other.

In the case of (3.7), optimal weights, the problem becomes

$$\min_{n,\theta} \frac{[k_1/n\theta+k_3][k_2/n(1-\theta)+k_4]-k_5^2}{k_1/n\theta+k_3+k_2/n(1-\theta)+k_4-2k_5} \ . \tag{4.6}$$

subject to $c_1n\theta + c_2n(1-\theta) \le C$, and $0 \le \theta \le 1$. We can show this is a decreasing function of n by taking a derivative and showing it be less than 0, and therefore we may use the substitution (4.3). In this case the problem reduces to

$$\min_{0\le\theta\le 1} \frac{[b_1+b_2/\theta][b_3+b_4/(1-\theta)]-k_5^2}{b_1+b_2/\theta+b_3+b_4/(1-\theta)-2k_5} \ ,$$

where

$$b_1 = k_1(c_1-c_2)/C+k_3 \ ,$$

$$b_2 = k_1c_1/C \ ,$$

$$b_3 = k_2(c_2-c_1)/C + k_4 \ ,$$

$$b_4 = k_2c_1/C \ .$$

Setting

$$f_1(\theta) = b_1 + b_2/\theta \ ,$$

$$f_2(\theta) = b_3 + b_4/(1-\theta) \ ,$$

the optimum occurs at the same value of θ as does the optimum of

$$\max_{0\le\theta\le 1} \frac{1}{f_1(\theta) - k_5} + \frac{1}{f_2(\theta) - k_5} \ . \tag{4.7}$$

Taking a derivative of the objective function and setting it equal 0, we find the roots of the following equation are the stationary values:

$$\theta^2 e_1 + \theta e_2 + e_3 = 0 \ , \tag{4.8}$$

where

$$e_1 = 2(b_3-k_5)^2-b_4(b_1-k_5)^2 \ ,$$

$$e_2 = -2[b_2(b_3-k_5)^2+b_2b_4(b_3-k_5)+b_2b_4(b_1-k_5)] \ ,$$

$$e_3 = b_2(b_3-k_5)^2+2b_2b_4(b_3-k_5)+b_2b_4^2-b_4b_2^2 \ .$$

The maximum in (4.7) must occur at one of the roots of (4.8) or at 0 or at 1. These four values can be calculated and the maximum found. In general this characterization seems to be the easiest to apply of those we have investigated.

It is not particularly easy to obtain the solution for the special case $\rho_{12} = \rho_{11} = \rho_{22} = 0$ from the above characterization. However, the original problem can be written

$$\min_{n_1,n_2} \frac{(k_1/n_1)(k_2/n_2)}{k_1/n_1 + k_2/n_2} \ ,$$

subject to $c_1 n_1 + c_2n_2 \le C$, and $n_1, n_2 \ge 0$. The solution of this problem can be obtained as the solution of

$$\max_{n_1,n_2} \frac{n_1}{k_1} + \frac{n_2}{k_2} \ ,$$

subject to $c_1 n_1 + c_2n_2 \le C$ and $n_1, n_2 \ge 0$. This linear programming problem is solved by

i) $n_1 = 0$, $n_2 = C/c_2$ if $c_2k_2 < c_1k_1$,

ii) $n_1 = C/c_1$, $n_2 = 0$ if $c_2k_2 > c_1k_1$,

iii) <u>any</u> feasible n_1, n_2 if $c_2k_2 = c_1k_1$.

Clearly this is equivalent to (4.5) and again we would expect to use one source or the other for the whole survey.

While our work here has been done for continuous n_1 and n_2, it should be realized that these variables are discrete and therefore may not exactly achieve the optimal values. However, the variances (objective functions) are not particularly ill-behaved for values near the optimal values and as a result we feel it is quite safe to use integer values for n_1 and n_2 which are close to optimal.

5. Optimum Allocation for a Model with Interviewers

In this section we briefly consider a simple situation where the presence of correlated measurement errors is solely due to the effects of interviewers. We assume that a large population K of interviewers is available for the survey. A simple random sample of k interviewers is selected without replacement, and of these k_1 are assigned to carry out the interviewing under conditions G_1 and $k_2 = k - k_1$ under conditions G_2 . We select k <u>independent</u> simple random samples without replacement from the target population, each of size m. These samples are assigned to the selected interviewers. Thus, each interviewer's work-load is equal, and the total sample size is n = km units.

The question we shall be addressing here is the following: What is the optimum allocation of the k interviewers (and hence the sample) to the two sources, G_1 and G_2 , such that the estimator variance is minimized subject to fixed cost $c_1k_1m + c_2k_2m \le C$?

Assuming that the measurement errors of units enumerated by different interviewers are uncorrelated; that sampling error is uncorrelated with measurement error; and that the sampling fraction k/K is negligible, it can be shown that the covariance v_{12} is approximately zero and that the variances of $\bar{y}_{t1}$ and $\bar{y}_{t2}$ are of the form $v_1^2 = a_1/k_1$ and $v_2^2 = a_2/k_2$, where a_1 and a_2 are population parameters independent of the sample sizes k_1 and k_2.

It follows that the present allocation problem is mathematically equivalent to the special case of uncorrelated measurement errors studied in section 4. By the solution obtained there, we conclude that the optimum allocation for either estimator will usually be to conduct the entire survey under one set of conditions or the other. Specifically, the solution is

i) $k_1 = 0$, $k_2 = C/c_2m$ if $c_2a_2 < c_1a_1$

ii) $k_1 = C/c_1m$, $k_2 = 0$ if $c_2a_2 > c_1a_1$

iii) <u>any</u> k_1, k_2 with $c_1mk_1 + c_2mk_2 = C$ if $c_2a_2 = c_1a_1$.

Once again, this solution is somewhat unsatisfactory because it assumes k_1 and k_2 are continuous, whereas in practice they are integer valued. This problem is more important here than it was in section 4, because now each unit change in k_1 or k_2 transfers a block of m sampling units from one source to the other. Nevertheless, we feel that by rounding the above solution to the nearest integer values of k_1 and k_2 we achieve a useful allocation which is close to optimum.

6. Discussion

There are a number of extensions of the work presented in sections 4 and 5. First, if there is a net nonzero bias, then we would want to allocate based on a minimum mean square error (MSE) criterion, rather than on minimum variance. Let $B_1 = \bar{M}_1 - \bar{Y}^o$ and $B_2 = \bar{M}_2 - \bar{Y}^o$ denote the biases of $\bar{y}_{t1}$ and $\bar{y}_{t2}$, respectively. Then, the minimum MSE allocation is the allocation given in sections 4 and 5, provided we redefine $k_3 = \sigma_{d1}^2\rho_{11} + B_1^2$, $k_4 = \sigma_{d2}^2\rho_{22} + B_2^2$, and $k_5 = v_{12} + B_1B_2$.

Second, although our work was for the case of the sample mean with simple random sampling, the allocation results easily generalize to a wider class of problems. Specifically, let $\hat{\Theta}_1$ denote an arbitrary estimator (of an arbitrary parameter Θ) computed from the sample enumerated under conditions G_1 , and $\hat{\Theta}_2$ denote the estimator under conditions G_2 . Assume that the variances and covariance are of the form $V\{\hat{\Theta}_1\} = k_1/n_1 + k_3$, $V\{\hat{\Theta}_2\} = k_2/n_2 + k_4$, and $Cov\{\hat{\Theta}_1, \hat{\Theta}_2\} = k_5$, where k_1, k_2, k_3, k_4, and k_5 are population parameters not dependent on the sample sizes. Aside from these restrictions, we leave the nature of the sampling design unspecified. Then, the optimum allocation can be obtained from (4.4) or (4.8), depending on how we weight the estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$. If $k_3 = k_4 = k_5 = 0$, i.e. the estimators are uncorrelated and the variances are inversely proportional to the sample sizes, then the optimum allocation is to carry out the entire survey under one set of conditions or the other.

Finally, we look at the allocation problem when the survey may be conducted under any of $p \ge 2$ sets of general survey conditions. Let $\hat{\Theta}_\alpha$ ($\alpha = 1, \ldots, p$) denote the estimators derived under the various general conditions, and consider the estimator $\hat{\Theta} = \sum_{\alpha=1}^{p} w_\alpha\hat{\Theta}_\alpha$, where $\{w_\alpha\}$ is an arbitrary weighting scheme (with $w_\alpha \ge 0$ and $\sum_{\alpha=1}^{p} w_\alpha = 1$) which puts weight 1 on source α if the total allocation is to that source. Assume that the variance of $\hat{\Theta}$ is of the form

$$V\{\hat{\theta}\} = \sum_{\alpha=1}^{p} w_\alpha^2 \sigma_\alpha^2/n_\alpha ,$$

where σ_α^2 is a population parameter independent of the sample sizes. Let the cost function be $C = \sum_{\alpha=1}^{p} c_\alpha n_\alpha$. Then, it can be shown that the variance is minimized subject to fixed cost by taking the single source for which $\sigma_\alpha^2 c_\alpha/C$ is a minimum.

The allocation results obtained here are useful in that they provide a rational means of choosing between (or combining) survey alternatives. To apply the results, we only require knowledge of the per unit costs of the alternative procedures and the moment properties of the associated measurement errors.

An important application of the allocation results is to the question of telephone versus personal interviewing. The trend in survey research seems to be towards more telephone interviewing, in many cases computer-assisted telephone interviewing. This trend is precipitated, at least partially, by the desire of survey sponsors to use the significant cost advantages of telephone interviewing. We feel, however, that the decision to use the telephone or any other survey procedure should not be on the basis of cost alone, but on both cost and accuracy considerations. Such improved decision making can only occur in an environment where more is known about the measurement errors associated with the alternative survey procedures.

References

Fellegi, I.P. "Response Variance and its Estimation." Journal of the American Statistical Association 59(1964): 1016-1041.

Hansen, M.H.; Hurwitz, W.N.; and Bershad, M.A. "Measurement Errors in Censuses and Surveys." Bulletin of the International Statistical Institute 38(1961): 359-374.

Horvitz, D.G. and Wolter, K.M. "Total Survey Design." in Advances in Health Survey Research Methods, ed. L. G. Reeder, National Center for Health Services Research and Health Resources Administration, DHEW Publication No. (HRS) 77-3145, 1975.

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 63(1973): 906-913.

Lessler, J.T. "A Double Sampling Scheme Model for Eliminating Measurement Process Bias and Estimating Measurement Errors in Surveys." Institute of Statistics Mimeo Series No. 949, University of North Carolina at Chapel Hill, 1974.

Raj, D. Sampling Theory. New York: McGraw-Hill, 1968.