Early Survey Models and Their Impact on Survey Quality Work

Gösta Forsman

SCB

R&D Report
Statistics Sweden
Research - Methods - Development
U/STM - 37

INLEDNING

TILL

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Föregångare:

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

Efterföljare:

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report, U/STM 1987:37. Early survey models and their impact on survey quality work / Gösta Forsman.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-RnD-USTM-1987-37

Early Survey Models and Their Impact on Survey Quality Work

Gösta Forsman

R&D Report
Statistics Sweden
Research - Methods - Development
U/STM - 37

Från trycket	Oktober 1987
Producent	Statistiska centralbyrån, Enheten för statistiska metoder
Ansvarig utgivare	Staffan Wahlström
Förfrågningar	Gösta Forsman, tel. 08 7834180

ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1987

EARLY SURVEY MODELS AND THEIR IMPACT ON SURVEY QUALITY WORK

Gösta Forsman, Statistics Sweden

1987-09-17

INDEX

[1] Invited paper presented at the U.S. Bureau of the Census' third Annual Research Conference (ARC III) in Baltimore, MD, USA, March 30 to April 1, 1987.

EARLY SURVEY MODELS AND THEIR IMPACT ON SURVEY QUALITY WORK

Gösta Forsman, Statistics Sweden

ABSTRACT

There have been great advances in sampling models over the past 60 years. As these models have been developed, so has an awareness of the problem of non-sampling errors (or measurement errors) in surveys. Two lines have emerged in this work, namely (i) development of theory and methods for coping with specific sources of non-sampling errors, and (ii) development of a comprehensive theory of an integrated treatment of survey errors (or, using the terminology of the 1950s, of mixed error models). This paper deals with the early research in this field up to the early 1970s, and looks at its impact on survey quality.

KEYWORDS

Non-sampling errors, early survey models, survey quality.

1. INTRODUCTION

At a statistical agency the survey quality work includes a variety of procedures like evaluation studies, preventive control, and production control. One fundamental part of this work is the measurement of survey errors. Measurement studies provide information about quality useful for both the producer of the data as well as the user. The survey researcher needs data on survey quality to improve methods and to allocate resources more effectively. The user of the statistics needs quality data to determine whether the phenomena observed are real or the result of variability and bias. In this paper I discuss the measurement of survey errors from a historical point of view.

1.1 Some historical notes

Measurement errors in surveys (by survey we mean either a census or a sample survey) were recognized when censuses were conducted in the 18th and 19th centuries. Wargentin (1780) demonstrated that the then widely spread opinion that Sweden annually lost a considerable number of its inhabitants through emigration was based on erroneous data from the Swedish population statistics. He also showed that this "emigration" (that the government found very disturbing) was in reality almost negligible. Another example is the exceptional overestimation of "people over 100 years old" in a sequence of Bulgarian Censuses of Population beginning in 1887, which attracted international attention (see Strecker and Wiegert 1981). In these ancient times survey errors were, for the most part, neglected. A change took place around 1930 when the survey theory began to develop.

1.2 Specific error sources

The early development of survey theory focused on the measurement and control of specific error sources. An important example is the very successful research on sampling errors. Today there are a large number of techniques available for controlling these errors. Since the 1930s there has been an increasing awareness of the problems of non-sampling errors. India and the United States led the early development in this field.

In India, agricultural surveys were used to conduct a great deal of measurement error research. Two forerunners were the Indian Statistical Institute and the Indian Council of Agricultural Research. These agencies worked on problems concerning the enumerator bias and the overestimation of crop yields when small plot sizes were used (see, e.g., Mahalanobis 1946, and Sukhatme 1947). Interpenetrating subsamples is one of the most important tools developed from these studies.

In the United States most of the work was done at the Bureau of the Census. Work was done on both the data collection and the data processing. Early references are Rice (1929), Deming and Geoffrey (1942), Palmer (1943), and Hansen and Hurwitz (1946). In the 1940s the Bureau developed a number of evaluation procedures, see, e.g., Eckler and Pritzker (1951).

Contributions from the United Kingdom were also important. At the Rothamstead Experimental Station, Fisher, Yates, Cochran, and others did research on statistical experiments in the 1920s and 30s. This work had a strong influence on the development of survey theory. At the London School of Economics, extensive studies on interviewer effects were conducted in the 1950s. Among others, Kendall, Durbin, and Stuart were engaged in this work.

1.3 Survey models

Studies of specific error sources have continued to be an important part of survey quality work. In the 1940s a parallel development emerged that aimed at an integrated control of all sources of errors, and thus of the total error. In this research error models were developed which were characterized by the assumption that both sampling and various non-sampling errors could be present. These were called mixed error models; later, the term survey models has been widely used.

We can distinguish three fields of application for survey models:

1) As already indicated, a survey model allows an integrated treatment of various error sources. Thus, the total error under the model can be estimated.

2) Survey models can be used to estimate the relative impact of different error sources on the total error. For recurrent surveys, this allows a reallocation of resources to effectively control the error sources.

3) Survey models might also be applied on a specific source of error to study the magnitude of its components. For example, if applied to the response error, we can estimate the total error and the components of the total error, the response bias and the response variance. This also can lead to an improved allocation of the resources among survey operations.

The concept of total survey design (see, e.g., Dalenius 1974) is closely linked to survey models. This implies a balanced allocation of the resources available in a survey so that the total error of the estimate is minimized. The concept of total survey design takes into account factors such as the purposes of the survey, the resources available, data on sampling and non-sampling errors, and also the properties of alternative procedures.

1.4 The subsequent discussion

We will review the early development of survey models and discuss its impact on subsequent work on survey quality. The presentation is restricted to models that include estimation procedures for the error components. The reason for confining the presentation to early models is that it takes some time to adopt new theory in practical work. A discussion of the impact of the last 10 to 15 years of development might be a bit premature. We will, however, touch briefly on this research in the last section.

2 THE DEVELOPMENT OF SURVEY MODELS

2.1 Models for variable measurement errors

The work to develop survey models (led by Indian and American statisticians) concentrated on sampling variance and measurement variability. Two important sources of measurement variability were:

a) The error that depends on the tendency of the interviewers (or enumerators or observers, depending on the data collection mode) to affect the respondent's answers, and

b) The error that emerges from the fact that the answers to a question can be different if the respondent is asked more than once.

There has not yet emerged a common nomenclature for these sources of error. For simplicity, I refer to them as a and b.

In the United States, Rice (1929) showed that the interviewers' own attitudes affected the respondents and could lead to a response error. There was a great need to measure this type of error at the Census Bureau. The data collection in the decennial censuses of population and housing was conducted by thousands of temporarily employed interviewers whose skills could vary considerably. This type of error was also well-known in India, e.g., in the crop surveys, where the observers might classify the same field very differently. This was not only a result of mistakes. Mahalanobis (1946) pointed out that it could be deliberate dishonesty, depending on the conditions under which the crop enumerators had to work: "In summer the temperature would often go up to 110^0F or more; during the monsoon many of the roads are submerged; throughout the crop survey a large number of investigators suffer from malaria and other diseases every year. This makes it all the more necessary to have adequate controls at the point of collection of the material".

The modeling of the above type a) error was done somewhat differently in different agencies. Central to the Census Bureau's model was that each interviewer generated clusters of responses. Then, by allocating a random subsample to each interviewer, the error could be measured by a cluster sampling (variance) formula. The situation differs, however, from cluster sampling. In cluster sampling, the correlation seen in the data reflects the correlation that exists in the population. In the Census Bureau's model, the correlation is a result of the observation and data collection process. This error component was called the correlated response variance. In India and in some other agencies in the United States, this error was regarded as a bias due to the interviewer (or the observer). The interviewers were considered a simple random sample from a population of inteviewers. The variance among the biases connected to these interviewers in the population was often called the interviewer variance. This could then be estimated from the sample of interviewers. The interviewer variance is often regarded as identical to the correlated component of the response variance according to the Census Bureau model. This may be approximately true but there are different opinions among researchers whether it is correct or not, as we shall see later.

The modeling of the type b) error was also done differently at different agencies. At the Bureau of the Census the random nature of this error was discussed early (see Palmer 1943, and Deming 1944). In the model developed at the Census Bureau it was assumed that an answer to an interview question is generated by a random process. As a consequence - even a response from a given respondent to a given interviewer has a probability distribution. A similar situation is assumed in Sukhatme (1954) and Sukhatme and Seth (1952) and probably also in Mahalanobis (1946), although Mahalanobis does not explicity describe a survey model as we have defined it here. Another way of modeling the type b) error is to assume that only one answer is possible for each respondent-interviewer-question combination. On the other hand, a given respondent could provide different answers to the same question to different interviewers. The stochastic element in a survey model with this assumption is entirely due to the sampling processes and the allocation of respondents to interviewers. Both interviewers and respondents are normally regarded as sampled from large populations. One can interpret the survey model described by Stock and Hochstim (1951) as using this deterministic approach. The same goes for the later model by Murthy (1967) and the conceptual discussion in Zarkovich (1966).

The models for the total survey error, which considered the sampling error and the two variable measurement error types described above, were formulated according to two basic ideas. The Census Bureau used a mean square error decomposition approach founded in sampling theory, while other agencies used a linear model approach founded in the ANOVA technique.

A. <u>The mean square error decomposition approach</u>

The Census Bureau model assumes a set of general conditions under which the survey is conducted. The survey is regarded as one trial from among a large set of conceived repetitions of the survey under the same general conditions. This means that a measurement derived from the survey has a well-defined, but unknown, probability distribution. The model postulates the existence of a true value, x, for each sampling unit. We denote the measurement for the i:th element at the t:th trial by y_{it}. Now the conditional expected value of y_{it} over all possible samples that include the i:th element and all possible trials that have resulted in such a sample, is

$$E(y_{it}|i) = Y_i \tag{2.1}$$

The difference between the observation on the i:th unit in a particular survey and the expected value of that unit is

$$d_{it} = y_{it} - Y_i$$

This is the response deviation that is measured from the expected value.

Assume now, that in a specific trial, t, the population mean, $\bar{X}$, is to be estimated by $\bar{y}_t$, the sample mean from a simple random sample of n units. Then the total error $\bar{y}_t - \bar{X}$ is measured by $MSE(\bar{y}_t)$, that can be decomposed as:

$$MSE(\bar{y}_t) = \sigma_S{}^2/n + \sigma_R{}^2[1+(n-1)\rho]/n + 2(n-1)\sigma_{RS}/n + B^2 \tag{2.2}$$

In (2.2), the first term is the sampling variance of $\bar{y}_t$, defined as the variance among the Y_i-values in the population, divided by n. The second term is the response variance, defined as the variance of $\bar{d}_t$, the average of the response deviations for the sample. It can be further decomposed into the simple response variance, $\sigma_R{}^2/n$, (which is the error component corresponding to the type b) error) and the correlated response variance, $\rho(n-1)\sigma\ ^2/n$. ρ_R is the intraclass correlation coefficient among the response deviations for a trial (survey), defined as

$$\rho = E(d_{it}, d_{i't})/\sigma_R{}^2, \quad i \neq i'$$

It is important to remember that the sampling variance measures variations caused by the sampling process, while the response variance measures variations assumed to characterize the measurement operation. The third term in (2.2) is the covariance of the response and sampling deviations, which is normally regarded as very small - it is zero for a complete census. The fourth term, finally, is the squared bias.

An important feature of the model is its broad applicability. It may be applied to any sequence of survey operations, i.e., either the full sequence or a subset of operations (for instance, interviewing and coding). Applied to the full sequence, the response variance reflects contributions from all operations such as interviewing, coding, editing and so forth. Applied to coding alone, the response variance reflects only coding and the response variance becomes a coding variance. Analogously to (2.2), coding gives a contribution to the MSE of the form

$$\sigma_c{}^2[1+(n-1)\rho_c]/n + B_c{}^2 \tag{2.3}$$

For surveys with interviewers, the correlated response variance may be especially large. It is then important to note that this component does not decrease when the number of sampling units within an interviewers' assignment increases. Thus a relatively low value of ρ can have a considerable effect on the total response variance and also on the total MSE. This is easily seen if the correlated response variance is assumed to depend entirely on the interviewers (i.e., on the type b error described in section 2.1). The response variance is then defined as

$$\sigma_R^2[1 + \rho(m-1)]/n \qquad (2.4)$$

where m is the (average) number of respondents assigned to an interviewer and σ_R^2 and ρ are slightly differently defined compared to (2.2).

The Census Bureau survey model was first presented in Hansen et.al. (1951). In their paper, Hansen et.al. assume that the correlated response variance depends entirely on the interviewers. They showed that the correlated component of the response variance could be estimated by means of interpenetrating subsamples. Hansen et.al. also showed that the "usual" textbook estimator of the sampling variance of $\bar{y}$ actually estimated the sum of the sampling and simple response variances. During the 1950s, the model was further elaborated on and eventually presented in the two widely recognized papers by Hansen, Hurwitz, Bershad (1961) and Hansen, Hurwitz, Pritzker (1964). In these, the correlated component of the response variance was defined as dependent not only on the interviewers, but on all field personnel. The correlation between answers to different interviewers was permitted to be non-zero, reflecting the possible correlations arising from supervisors, coders, editors, keyers, etc. This is the situation assumed in (2.2).

In Hansen et.al. (1964), an estimator of the simple response variance for 0,1-variables was presented. This estimator was developed under the assumption that independent repeated measurements of the sampling units were conducted. In this case the original survey and its replication (the reinterview) were considered two independent randomly selected trials. Now the gross difference rate,

$$g = \Sigma(y_{i1} - y_{i2})^2/n, \tag{2.5}$$

divided by two, is an unbiased estimate of $\sigma_R{}^2/n$. Hansen et.al. (1964), defined an index of inconsistency as the ratio of the simple response variance to the total variance of individual responses, $\sigma_y{}^2$, that is

$$I = \sigma_R{}^2/\sigma_y{}^2. \tag{2.6}$$

For a Bernoulli random variable, the total variance $\sigma_y{}^2$ is P(1-P), where P is the expectation of the sample mean. An estimator of the numerator is then g/2. The denominator may be estimated by $\bar{y}_t(1-\bar{y}_t)$, t=1, 2. $\bar{y}_t$ is then the proportion of sample units that belongs to the category of study in trial t (either data from trial 1 or from trial 2 or from both surveys may be used). Obviously, the index takes values between 0 and 1. I=0 when the entire variance is due to the sampling process and no response variance is present. I=1 when the measurement process is analogous to tossing a coin; the response variance is then equal to the total variance. Thus, low values of the index indicates that the measurement process is under control.

The two procedures for estimating error components mentioned above, interpenetration and repeated measurements, are the basic methods available in the estimation process. In Bailar and Dalenius (1969), these are applied (also in combination) in several basic study schemes aiming at estimating different variance components in the Census Bureau model. Repeated measurements can also be used for estimating the bias component. They must, however, be conducted with a preferred procedure that can be assumed to provide data close to the true values. Often, when the basic methods can not be applied in a survey, data from other surveys are used instead. For example, a census match with labor force survey data may be regarded as a re-enumeration, and data from two independently conducted surveys with similar questions and data collection procedures on the same population may be treated as if interpenetration had been applied (see Tepping and Boland 1972).

Fellegi (1964) elaborated on the Census Bureau Model. His sampling design involved both interpenetration and repeated observation. In his notation, the assignment of the j:th interviewer was $[S_{j(1)}, S_{j(2)}, j=1,\ldots,k]$, where $S_{j(1)}$ and $S_{j(2)}$ are the assignments for the j:th interviewer in the original and reinterview surveys, respectively. $S_{j(1)}$ and $S_{j(2)}$ are not the same for a given interviewer. The model differs from the Hansen, Hurwitz, Pritzker (1964) model in that the conditional expected values of a measured value y_{ijt} given a respondent, i, and an interviewer, j, over the trials need not be the same for the original survey and the reinterview. This model permits the definition of several types of correlation among the responses, e.g.,

i) The correlation of response deviations obtained by the same interviewer in the same survey.

ii) The correlation of response deviations obtained by different interviewers in the same survey.

iii) The correlation of response deviations obtained in the two surveys for the same units.

iv) The correlation of response deviations obtained by the same interviewer in different surveys.

v) The correlation between the sampling and response deviations for the same interviewer in the same survey.

Fellegi presents seven linearly independent estimators to be used for estimating the parameters outlined above and ten other parameters (including the simple response variances and the sampling variances in the two surveys). Unbiased estimators of all these parameters are not possible to obtain. The best available solution is to provide biased estimators for those parameters considered most important, and where the biases are provided in terms of the other parameters.

Murthy (1967) and Des Raj (1968) present variance decomposition models. They both arrive at expressions of the variance of the sample mean that are similar to the Census Bureau model decomposition although these components have different definitions. Murthy considers the survey to have two steps of randomization:

i) A sample of population elements, s, and

ii) A sample of survey personnel, r.

He defines y_{ij} as the value obtained by the j:th interviewer for the i:th element. Since Murthy assumes the deterministic response model described in 2.1, this value is not a random variable. He gives the following expression for the variance of the sample mean, $\bar{y}$:

$$V(\bar{y}) = \sigma_s^2/n + \sigma_d^2[1+(m-1)\rho]/n,$$

where the terms are called sampling variance, simple or uncorrelated response variance, and correlated response variance. The names are the same as the names of the components of the Census Bureau model, but they are not the same components. The response deviations are defined differently.

Des Raj gives an example of a survey model in which the sample design is not simple random sampling. In his design the interviewers are allocated randomly to primary sampling units that have been selected with probabilities proportional to size.

B. <u>The linear model approach</u>

Linear survey models may also be constructed in many different ways. This type of model emerged from the analysis of variance theory. In the late 1930s and early 1940s, the Indian Statistical Institute under the leadership of P.C. Mahalanobis developed sampling designs for their crop surveys with embedded experiments based on interpenetrating subsamples (Mahalanobis 1946). The purpose of these designs was to control the individual investigator bias that had been seen in surveys where different investigators had unknowingly been allotted the same fields. In their simplest form, the designs permitted the institute to test, using an F-statistic, if the mean values emerging from the investigators workloads differed significantly from each other. More complicated survey designs were also used, e.g., two-way ANOVA models made it possible to identify (by an F-test) measurement errors to parts of an investigator's workload.

One of the first examples in the literature of measuring the overall error by means of linear survey models was provided by Stock and Hochstim (1951). It seems as if the authors assume the deterministic response model as mentioned above, i.e., only one answer is possible for each respondent- interviewer-question combination. In this application, the deterministic model assumes a population of N respondents and a population of K interviewers. Then, assuming that each interviewer, j, interviews each respondent, i, in the population, the data generating process may be modeled in the following way:

$$y_{ij} = \bar{y}_{..} + I_j + e_{ij}, \quad i=1,\dots,N, \quad j=1,\dots,K,$$

where $\bar{y}_{..}$ is the mean of all NxK observations.

I_j is the individual interviewer bias for interviewer j, i.e., the difference between the overall mean μ and the mean of the N observations of interviewer j.

e_{ij} is the deviation between the observed value and $\bar{y}_{..} + I_j$ when interviewer j interviews respondent i.

It is assumed that no correlation is present between I and e. The survey design is such that an interpenetrated subsample, drawn by simple random sampling, is randomly allotted to each of k interviewers. The interviewers are sampled by simple random sampling from a population of K interviewers.

Now, if we further assume that the sampling fractions are small, the variance of the estimated mean, $\bar{y}$, is approximately:

$$\mathrm{Var}(\bar{y}) = \sigma_I^2/k + \sigma_e^2/n,$$

where n is the total sample size. σ_I^2 is the variation between the I_j:s. This term is the interviewer variance mentioned in section 2.1, and is present in most ANOVA-type survey models. σ_e^2 is the average variation between respondents within interviewers. This is the sampling variance and would be the total variance under the model if there were no interviewer effects. Note that the model does not take into account the existence of true values. If we denote the mean squares between and within interviewers by S_I^2 and S_e^2, respectively, an estimator of σ_I^2 is $[S_I^2 - S_e^2]/(n/k)$ and an estimator of σ_e^2 is S_e^2.

Stock and Hochstim applied the model on a quota sampling design, which made the analysis of the results more difficult than if the model, as described above, had been applied strictly. The interviewer's selection bias could not be distinguished from the response bias due to the interviewers.

Sukhatme (1954) presents a linear survey model rather different from the Stock and Hochstim model. He lets

$$y_{ij} = x_i + I_j + \varepsilon_{ij},$$

where y_{ij} and I_j are defined as above, x_i is the true value and ε_{ij} is "the random deviation of $x_i + I_j$ from the reported value." It is assumed that the ε_{ij} are independently distributed with mean 0 and variance σ_ε^2. Assuming a similar sampling and measurement design as Stock and Hochstim did, Sukhatme derives the following expression for $\mathrm{Var}(\bar{y})$:

$$\mathrm{Var}(\bar{y}) = \sigma_x^2(1/n - 1/N) + \sigma_I^2(1/k - 1/K) + \sigma_\varepsilon^2/n$$

or, if N and K are large:

$$\text{Var}(\bar{y}) = (\sigma_x^2 + \sigma_\varepsilon^2)/n + \sigma_I^2/k$$

σ_x^2 is the variance among the true values in the population. It can not be estimated separately from σ_ε^2 with Sukatme's design. Sukhatme shows that the mean square between observations within enumerators is an estimator of the sum of σ_x^2 and σ_ε^2. This result is similar to the finding that the "usual" sampling variance estimator estimates the sum of the sampling and simple response variances in the Census Bureau model. However, the components are differently defined in the two models.

Sukhatme also derives the correlation, ρ', between responses obtained by the same interviewer. If the finite population correction is small, the variance for a single obsevation is approximately $\sigma_y^2 \simeq \sigma_x^2 + \sigma_I^2 + \sigma_\varepsilon^2$. An approximate expression for ρ' is

$$\rho' \simeq \sigma_I^2/\sigma_y^2 \simeq \sigma_I^2/[\sigma_x^2+\sigma_I^2+\sigma_\varepsilon^2]$$

which gives $\text{Var}(\bar{y}) = \sigma_y^2[1+\rho'((n/k)-1)]/n$

Kish (1962) presents a similar model. He, however, pools the x and ε components. The reason for this is that (as is mentioned above) the variance components σ_x^2 and σ_ε^2 cannot be estimated separately with Kish's (or Sukhatme's) design since it does not include repeated measurements. Kish's model can be written

$$y_{ij} = y'_{ij} + I_j,$$

where $y'_{ij} = x_i + \varepsilon_{ij}$. The definitions of the components seem to be the same as in the Sukhatme model. The intraclass correlation, ρ^*, between responses obtained by the same interviewer is defined as:

$$\rho^* = s_a^2/(s_b^2 + s_a^2), \qquad (2.7)$$

where s_a^2 and s_b^2 are the sample estimates of the variance between interviewers and within interviewers, respectively.

The Sukhatme (1954) model is a simplified version of the model presented in Sukhatme and Seth (1952). The latter model allows several observations on each unit, and an interaction between the j:th interviewer and the i:th respondent, denoted δ_{ij}. The model is expressed as:

$$y_{ijk} = x_i + I_j + \delta_{ij} + \varepsilon_{ijk},$$

where y_{ijk} is the response obtained by the j-th interviewer from the i:th sample respondent on the k:th occasion. x, I, and δ are defined as above and ε_{ijk} is the random deviation associated with y_{ijk} that is not accounted for by interviewer and interaction effects. Sukhatme and Seth presented estimators for the variance components, or linear combinations of them, for different types of measurement designs.

1) Each unit is observed once
2) Each unit is observed p times by the same enumerator
3) Each unit is observed once by each of p enumerators
4) Some of the units are observed once and some are observed twice.

Case 4 is the only situation where the σ_ε^2 can be estimated since this estimation requires repeated observations.

All of the above linear models are based on the assumption that the expected values of the measurements do not depend on the sample. This assumption is not present in the variance decomposition models.

C. Other models for measurement variability

There are other approaches to developing survey models for surveys that estimate proportions in the presence of classification errors. In these models, the key concept is what is called misclassification probability. Assume, for example, that the true value, x_i, for unit i is 1 if i belongs to some category, say C, and 0 otherwise. Errors that give rise to the misclassification of a unit are considered. The survey procedure is assumed to generate a stochastic variable y_i, such that

$$P0 = Pr(y_i = 1 \mid x_i = 0)$$

$$P1 = Pr(y_i = 0 \mid x_i = 1) \qquad i = 1,\ldots,N$$

P0 and P1 are the misclassification probabilities. P1 denotes the probability that a unit is misclassified from C into another category and P0 denotes the probability that a unit is misclassified from another category into C.

Casady (1966) derived two such survey models for the analysis of reinterview data in the Health Interview Survey conducted by the U.S. National Center for Health Statistics. Casady defined the "within element response variability" and an index of inconsistency and presented estimators of these parameters under the two models. The models differed in that the misclassification probabilities in the first model were assumed to be constant over different trials while the misclassification probabilities were permitted to vary between trias in the second model.

Swensson (1969) showed that the first Casady model could be regarded as a special case of the Census Bureau model (only the sampling variance and the simple response variance are regarded - the correlated response variance can not be studied under the given definition of misclassification probabilities).

The misclassification approach was discussed in Cochran (1968), who defined the misclassification probabilities as dependent on the unit. Bailar and Biemer (1984) showed that the misclassification probabilities can be formulated as dependent on both the unit and the operator (i.e., an interviewer, coder, supervisor, etc.). This allows a correlated measurement error component to be estimated with the misclassification probability approach.

2.2 Survey models for systematic errors

Systematic errors are normally studied by comparing the survey data to preferred data. Higher quality data are usually obtained from reinterviews or record checks. This can be done without applying survey models. The method also allows an estimation of the bias term as it appears in, e.g., the Census Bureau model (see formula 2.2). This term has been extensively studied within the U.S. census evaluation programs.

In addition to the Census Bureau work, bias models were developed for specific survey situations. Kish and Lansing (1954) developed a model for the situation when not only the observed values (y_i) but also preferred values (a_i), obtained from a preferred data collection procedure, are available. These preferred values were, however, not regarded as good as the true values (x_i). This model was to estimate the error in a study of the market value of houses, a study that was a part of the 1950 Survey of Consumer Finances in the United States. Let the y_i:s be estimates made by the owners, and the a_i:s estimates made by experts. The model expresses, in terms of the mean square error, the difference in accuracy between estimates obtained from the original method and the preferred method. The model can be summarized as follows: Denote the population means and variances by

$E(x) = \bar{X}$, $E(y) = \bar{Y}$, $E(a) = \bar{A}$ and

$V(x) = E(x-\bar{X})$, $V(y) = E(y-\bar{Y})$, $V(a) = E(a-\bar{A})$

The individual response bias is for the ordinary method $y_i - x_i$, and $a_i - x_i$ for the preferred method. The difference between the two response errors is $d_i = (y_i - x_i) - (a_i - x_i) = y_i - a_i$.

Kish and Lansing also define

$\bar{Y} - \bar{X}$ = the response bias when the ordinary method is used

$\bar{A} - \bar{X}$ = the response bias when the preferred method is used

$\bar{D} = \bar{Y} - \bar{A}$ = the difference between the two biases

Then a term called the mean square difference of the measurements is defined as: $M.S.(d) = E(d) = E(y-a)$.

The covariance between the difference d and a is

$$Cov(d,a) = E(d-\bar{D})(a-\bar{A}) = Cov(y,a) - V(a).$$

With the above definitions, it is possible to derive an equation that shows how much larger the total mean square error is if the ordinary method is used than if the preferred method is used:

$$V(y) + \bar{D}^2 - V(a) = M.S.(d) + 2Cov(d,a). \qquad (2.8)$$

If the results obtained with the preferred method are regarded as the true values (i.e., if a = x), the right-hand side is the increase in total MSE that depends on measurement errors. If the true values, x, had been available, the equation would have been

$$V(y) + (\bar{Y} - \bar{X}) - V(x) = E(y - x) + 2Cov[\,(y-x),\ x].$$

Kish and Lansing derive unbiased estimators for the five terms in the equation (2.8) assuming simple random sampling. The estimator of $\bar{D}$ can assume negative values. Thus, a positively biased estimator may be used or the estimator can be truncated at 0.

The well-known randomized response model presented by Warner (1965) can be regarded as a survey model since it takes into account both response (and nonresponse) bias and sampling variance. In the Warner's model, error probabilities are intentionally introduced to eliminate nonresponse and erroneous answers to sensitive questions. This technique makes it possible to construct unbiased maximum likelihood estimators of population means and totals.

Warner assumed that every person in the population belongs to either of two groups, group A (x_i=1) or group B (x_i=0), and that the purpose of the interview survey is to estimate the proportion, $\bar{X}$, belonging to group A. Warner's original randomizing device was set up in the following way. "Before the interviews, each interviewer is furnished with an identical spinner with a face marked so that the spinner points to the letter A with probability P and to the letter B with probability (1-P). Then, in each interview, the interviewee is asked to spin the spinner unobserved by the interviewer and report whether or not the spinner points to the letter representing the group to which the interviewee belongs." The value of P is chosen in advance by the statistician, i.e., P is a known parameter and is not estimated. It is assumed that the respondents answer correctly. So, if P=1 or P=0, true values from all respondents are obtained. Of course, P is never 1 or 0 in practice. If P was equal to 1 or 0, then the situation would be identical to the traditional. Randomized response does not work very well if P-values close to 1 or 0 are used. The point is that the respondent shall be convinced that both A and B can occur with a fairly large probability. The closer the P-value is to 1/2, the less need the respondent reveal about which group he/she belongs to.

To derive an estimate of $\bar{X}$, we use y_i to denote the result of an observation on a unit drawn by simple random sampling from the population. Then

$$\Pr(y_i=1) = \bar{X}P + (1-\bar{X})(1-P) \text{ and}$$

$$\Pr(y_i=0) = (1-\bar{X})P + \bar{X}(1-P).$$

If we assume that the sample has resulted in n_1 1:s and $(n-n_1)$ 0:s an unbiased maximum likelihood estimator of $\bar{X}$ is

$$\hat{\bar{X}} = (P-1)/(2P-1) + n_1/n(2P-1)$$

with the variance

$$\sigma^2(\hat{\bar{X}}) = \bar{X}(1-\bar{X})/n + \left[1/16(P-1/2)^2 - 1/4\right]/n.$$

The first variance term is the sampling variance and the second is a "response variance," due to the randomization device. The "response variance" is the price one has to pay for the presumed reduction in bias obtained with this method.

The connection between survey models and randomized response models is even more evident in Abul-Ela, Greenberg, and Horvitz (1967). They extended the Warner model to a trichotomous randomized response model in which the respondent was assumed to tell the truth with a probability that was allowed to be less than 1.

3 THE IMPACT OF SURVEY MODELS ON SURVEY QUALITY WORK

Several aspects of survey quality work have been affected by the use of early survey models. Among these are:

i) The early survey models provided a framework that has completely permeated survey practice. The simple and correlated components of the response variance, the interviewer variance, and the index of inconsistency have become well-known and useful concepts by the use of survey models and are often theoretically discussed in technical reports even in situations where models have not been explicitly applied. These discussions serve as reminders to the users of the survey results that the figures may suffer from errors other than those that have been measured. The knowledge of such error components is also important when designing the data collection and the data processing procedures.

ii) Models have been used in regular surveys to measure error components and the total error. Of course, the survey must be designed according to the model postulates, e.g., using interpenetrating subsamples. The error components under study are usually the sampling variance and the enumerator variance (or the correlated response variance).

iii) In evaluation studies, like post enumeration surveys, survey models have been used to assess the relative impact of different sources of error on the total error. This kind of evaluation study can be reenumerations that are then compared to the original survey to assess uncorrelated error components or biasses. Other evaluation studies have been designed to measure correlated components. This type of work has been done within the content evaluation programs of the population censuses in the United States and Canada during the last decades.

iv) Survey models have also been used to develop new methods of data collection and processing. These models are then used to measure differences and likenesses when comparing different methods. The Institute of Social Research, University of Michigan has used a model-based approach in its research on computer assisted telephone interviews.

3.1 Use of survey models in different countries

In this section we will give examples of quality work mentioned in ii)-iv) above. It should be emphasized that the list is by no means a comphrehensive review of the quality work guided and inspired by the early models. The error component estimates presented from different surveys should be compared with caution, since the original models are in some cases modified and the definitions of the components might differ slightly. Any comparison should also take into account survey design features that might affect interviewer and response errors, such as the question wording, interviewer training, etc. This is information that is not normally reported.

THE UNITED STATES

The Bureau of the Census

There is no doubt that the most extensive quality work based on survey models has been conducted at the U.S. Bureau of the Census. The bulk of the work has been done within two of the Bureau's major projects: the Decennial Census of Population and Housing and the Current Population Survey.

A continuing program of research, evaluation, and experimental studies has been conducted as a part of the censuses and during the inter-censal periods. The results of the 1950 Census experiments led to important changes in procedures adopted for the 1960 Census. In one of these experiments, a set of interviewer-assignment areas was designated. In these areas, the interviewers' assignments were randomly allocated according to the design postulated in the Hansen et. al. (1951) model. This experiment dealt with the variance between and within interviewers. The intraclass correlation of response errors within interviewers was also estimated. In Table 1 the intraclass correlations for some items are shown.

For items that are typically difficult to measure (i.e., occupation, education, and income) the correlation was often around .03 (see also Hansen and Tepping, 1969, p.11). This seems small, but since the average size of an interviewers assignment was about 700, the factor $[1+\rho(m-1)]$ in (2.4) becomes larger than 20, and leads to a substantial contribution to the total variance even for moderate σ_R^2. Findings like these showed that the variability of the complete census results was as large as if only a 25% sample had been taken (in the absence of interviewer effects). This was true even for areas with populations smaller than 5000 people. These findings, complemented by studies of the bias and experimental studies of self-enumeration etc., led to the following procedural changes for the most difficult items to measure in the 1960 Census:

i) The data collection was based on a 25% sample.

ii) A self-enumeration procedure was introduced for the sample questions.

The interviewers were, however, still involved in the data collection for the 1960 Census. Interviewers delivered the questionnaires to the households and completed them for those households that did not mail in a completed form or whose questionnaires were inconsistent. This led to an interviewer influence on the variance in the 1960 Census too. It was much smaller than in the 1950 Census, but it was still important for a number of items.

Table 1 Values of ρ for selected characteristics, U.S. Population: 1950.

Characteristic	ρ
Race:	
Negro	.0165
Other	.0043
Age:	
Under 1 year	.0002
1 year or older	.0009
Under 14 years	.0030
25 years or older	.0040
55 years or older	.0026
Educational attainment:	
Grade 12 or over	.0125
Grade 9 or over	.0064
Grade 8 or over	.0064
Grade 5 or over	.0027
Not reported	.0543
Residence:	
Farm	.0609
Nonfarm:	
Male	.0330
Female	.0278
Income--earned:	
Less than 2000	.0059
2000 to 4999	.0060
5000 or more	.0087
Not reported	.0160
Income--unearned:	
None	.0313
Less than 2500	.0246
2500 or more	.0009
Not reported	.0599

Source: U.S. Bureau of the Census (1985)

In the 1970 and 1980 Censuses changes were made in the census-taking procedures in that the questionnaires were delivered by mail to most of the population (95% in the 1980 Census). The enumerators still had an important role in the follow-up procedures, and enumerator variance studies were also made in the evaluation programs of these censuses.

Within all content evaluation programs of the censuses from 1950 to 1980, large-scale reenumeration studies were conducted to obtain estimates of response variance and bias. The reenumerations were conducted as reinterviews or as a record match to the Current Population Survey (CPS).

The Census Bureau model was also applied to the coding process, as described above in formula (2.3). Jabine and Tepping (1973) presented estimates on the simple and correlated coding variance components (presented as relvariances) for 1960 Population Census data. These were related to sampling and total response relvariances as well as to response and coder bias (the latter was based on 1970 Census data).

In the Current Population Survey a continuing reinterview program has been conducted since the beginning of the 1950s. These studies are primarily designed to control the field procedures rather than measuring the simple response variance according to the Census Bureau model. Nevertheless, the reinterview data is continually used to derive the index of inconsistency for various items. According to U.S. Bureau of the Census (1978), this measure has an important role in the CPS quality work: "The index is used primarily to monitor the measurement procedures over time. Substantial changes in the indexes that persist for several months result in review of field procedures to determine and remedy the cause."

Experiments aiming at measuring the correlated components of the response variance are not conducted in the CPS. Tepping and Boland (1972) report, however, from a study where data from the Monthly Labor Survey (MLS), carried out during six months in 1966 concurrently with the CPS, provided estimates independent of the regular CPS estimates for several items. The two estimates could then be used for estimating the correlated response variance component. In this paper, Tepping and Boland present estimates of the ratio of the correlated response variance to the sum of the sampling variance and the simple response variance, i.e., in terms of section 2.1:

$$(m-1)\rho\sigma_R^2/(\sigma_R^2 + \sigma_S^2)$$

Table 2 shows some estimated ratios as given in Tepping and Boland (1972).

Table 2 Ratio values of correlated component of response variance to sampling variance for selected characteristics in the CPS/MLS study.

Item	Estimate ratio
Married, spouse present	.77
Single	.57
Male head	.94
Wife of head	.74
White	.66
At work--full time	.80
At work--part time	.60
House	.64
Other employment status	.80
Employment in nonagriculture	.90
Private wage worker	.64
Self-employed worker	.53
1 to 34 hours, usually full time, other	.66
35 to 40 hours	.67
41 hours or more	.72
Nonworkers	.67
Dwelling unit	1.0
Total population	1.0

Source: Tepping and Boland (1972)

The estimation of interviewer variance in the CPS is currently under consideration, see Biemer et.al. (1985).

The Census Bureau model was also applied, in a modified form, to study the interviewer variance in the National Crime Survey, conducted in eight cities. In Bailey, Moore, and Bailar (1978), the ratio of the correlated response variance to the sampling variance was presented for the victimization rates for major crimes against persons. The correlated response variance was defined here as the sum of the within interviewer correlated response variance and twice the covariance of response and sampling deviations of different units. The ratios presented ranged from 0.00 (negative values of the correlated response variance were replaced by 0.00) to 1.40. The latter value was exceptionally large and indicated that the sampling variance in this case must be multiplied by 2.4 to reflect the total variance of the statistic. As is seen in Table 3, several values between .50 and 1.00 indicate the strong influence of interviewers on the precision of the estimates.

Institute for Social Research (ISR)

At the Survey Research Center at the ISR, University of Michigan, different models have been developed. The Kish (1962) model has been frequently applied in measuring interviewer effects, first by Kish in studies of factory workers' job attitudes. In recent years, Groves and others have applied the Kish model in various telephone surveys. The parameter of study in these applications is the intra-class (or intra-interviewer) correlation, ρ^*, defined by (2.7). Groves and Magilavy (1986) reviewed nine ISR telephone surveys and the estimates of ρ^* for 297 items. Other interview surveys were also reviewed in which similar models for interviewer effects were applied. The results of the nine ISR surveys are summarized in Table 4.

The average values of ρ^* in the surveys were in eight cases under .01, but varied considerably between different statistics. The lowest average ρ^*, .0018, was found in the survey with the largest interviewer workload, which, together with other observations, led Groves and Magilavy to the interesting conclusion that the interviewer variability might be larger in the initial cases completed by the interviewers. The lowest row in Table 4 shows the design effect $1+\rho^*(m-1)$.

Table 3. Ratio of correlated response variance to sampling variance for eight cities in the 1975 National Crime Survey.

Kinds of victimization crimes	Atlanta	Baltimore	Cleveland	Dallas	Denver	Newark	Portland	St. Louis	Average estimated standard error
	Personal victimizations								
Total	.73	.61	.18	.51	.34	1.40	.59	.96	.33
Assaultive violence	.72	.63	.31	.45	.09	.85	.68	1.16	.27
Assaultive violence with theft	.00	.17	.33	.00	.00	.10	.24	.00	.14
Assaultive violence without theft	.80	.54	.37	.50	.12	1.21	.70	1.16	.27
Personal theft without assault	.28	.30	.00	.22	.47	.79	.27	.06	.24

Source: Bailey, Moore and Bailar (1978)

Table 4. Descriptions of nine studies of ρ* and summary of findings.

Study Description and Summary Statistics	Study of Telephone Methodology	Health and Television Viewing	Health in America	1980 Post Election study	Monthly Survey of Consumer Attitudes				
					Nov'81	Dec'81	Jan'82	Feb'82	Mar'82
Field period	Apr-May 1976	Mar-Apr 1979	Oct-Dec 1979	Nov-Dec 1980	Nov'81	Dec'81	Jan'82	Feb'82	Mar'82
Response rate	59%	67%	80%	87.3%	73.5%	72.8%	73.3%	76.9%	73.8%
No. inerviewers	37	30	33	22	31	28	21	25	26
No. inteviews	1529	954	1918	697	370	350	386	379	366
Average workload	41.3	31.8	58.1	31.7	11.9	12.5	18.4	15.2	14.1
Major topics	Political economic & social issues	Health and TV viewing	Health	Political issues	Economic issues	Economic issues	Economic issues	Economic issues	Economic issues
No. variables analyzed	25	55	25	42	30	30	30	30	30
Range of ρ*	-.0080 .0560	-.0150 .1650	-.0070 .0097	-.0154 .1710	-.0217 .0895	-.0373 .0546	-.0221 .0916	-.0419 .0657	-.0356 .0729
Mean values									
ρ*	.0089	.0074	.0018	.0086	.0184	.0057	.0163	.0090	.0067
$deff_{int}$	1.36	1.23	1.10	1.26	1.20	1.07	1.28	1.13	1.09

Source: Groves and Magilavy (1986)

Groves and Magilavy also discuss two issues concerning the study of interviewer effects that have been largely overlooked in the literature. These are the stability of the estimates of ρ^* and the causes of interviewer effects.

CANADA

In Canada, the Fellegi model was applied in an experimental pilot study preceeding the 1961 Canadian Census of Population. The results were similar to those found in the U.S. Census in that the correlated response variance, derived as the mean of the correlated response variances in the two surveys, was "several times as large as the simple response variance for all except the basic population counts, such as the number of males, sons, married persons, persons of certain age, etc." (Fellegi, 1964).

Fellegi concluded that, for most characteristics, "considerable gains in the total response variance may be made by reducing the size of the enumerators' assignments." Fellegi argued that the Canadian Census should use a self-enumeration procedure. To find out if such a procedure would increase the simple response variance, he compared the index of inconsistency for a self-enumeration survey with the index of inconsistency for a an interview survey. He used items from the 1960 U.S. Census (self-enumeration) that corresponded to items in his pilot study for the 1961 Canadian Census (interviews). Fellegi found that the values of the simple response variance were rather similar despite the different procedures. As a result of these findings, the 1971 Population Census of Canada was substantially modified. Self-enumeration was introduced along with a sample based collection of most census questions. Later, Krotki and Hill (1978) compared the Fellegi estimates of the correlated response variance with the corresponding estimates from the 1971 and the 1976 Canadian Censuses (see Table 5). They found that for almost all characteristics examined, the magnitude of the estimates were considerably reduced. Note that in Table 5, all estimates are based on census data only. In Fellegi (1964) the 1961 estimates were based on both the census and the reinterview.

Table 5 A comparison of 1961, 1971 and 1976 correlated response variance estimates for characteristics published in Fellegi (1964).

	1961 estimate x 10^4 comparable to 1971	1971 estimate x 10^4	1976 estimate x 10^4 (4)
Sex: Male (1)	13.0	0.2	- 3.4
Age: 5 (1)	- 0.9	0.4	- 0.6
Ethnic Group: French (1)	1688.0	151.1	-
Highest grade of school, Attended: High school, grade 5 or univ. (2)	23.3	9.8	-
Persons looking for work last week	40.8	5.6	-
Persons who usually work 40 hours a week (3)	83.1	48.9	-
Industry: Manufacturing	6.7	-27.0	-
Industry: Trade	18.5	- 3.4	-

Notes: (1) Age, sex and ethnic group were 100 % variables in all three censuses. The other variables were 100 % in 1961 but sample variables in 1971 and 1976.

(2) 1961 wording. In 1971 and 1976 the estimate is for grade 12, 13 or university.

(3) 1961 wording. In 1971 the estimate is for persons who usually work 40-44 hours a week.

(4) Entries marked with - are unavailable for 1976.

Source: Krotki and Hill (1978)

EUROPE

In Europe, surprisingly few applications of survey models seem to have been conducted in government agencies. We do not know of any example from eastern Europe, from the Nordic countries outside Sweden, or from most other countries in Western Europe. Even in otherwise methodologically active central bureaus like those in Sweden and the Netherlands, the applications of survey models are rare.

In Spain, the General Population Survey is a continuing interview survey that gives bi-monthly and quarterly estimates on several items concerning households and individuals, like labor force items, family budgets, health, purchase plans, etc. Since the early 1970, an evaluation program of the General Population Survey has been conducted, based on 3000 reinterviews each quarter. The purposes of the program are to control the work of the interviewers and to evaluate the general quality of the results. According to Sanchez-Crespo (1973, 1981), the quality evaluation is based on the U.S. Census Bureau model. In the 1981 paper, estimates of the total response variance, the simple, and correlated variance components are presented for the variable "unemployed" (the study design used for estimating the correlated component is, however, not described). The intraclass correlation coefficient was estimated to .02 and .029 for two time periods and the correlated component was found to give the largest contribution to the total response variance.

In the Federal Republic of Germany, the Federal Statistical Office conducted a study following the 1970 Population Census, based on a survey model that differed somewhat from the U.S. Census Bureau model. The German model was used for measuring error components in metrically scaled data, especially the response variability. According to Strecker et.al. (1984), this is the only application of survey models in the Federal Republic of Germany. Theoretical work has, however, been conducted in the 1970s and 1980s by Strecker and Wiegert at the University of Tübingen.

In Belgium, a variance decomposition model for surveys with reenumerations, developed by Strecker and Wiegert, was applied in the 1979 Census of Agriculture. The application was limited to one kind of data, viz., the number of pigs. The model, related to the U.S. Census Bureau model, was originally developed for a study design with k enumerations, which permitted estimation of several correlation components. Since k=2 in the application, only the simple response variance component could be estimated by means of reenumeration data. The data was collected by mail and therefore the correlated response variance component was assumed to be small and not considered in the model. The "reenumeration" was conducted before the census and thus it could not reflect the same time point as the census did. Because of that, adjustments were made for change over time by means of a control study (after the census) to make the data comparable. This study also provided "true" values, on which bias estimates could be based.

The results from the Belgian study are analyzed in detail in, e.g., Strecker, Wiegert, and Kafka (1984). The impact of the simple response variance component was considerable, as the following example shows.

The mean square error was defined as the sum of the simple response variance, the sampling variance, and the squared bias. The relative MSE for the estimate of the mean number of pigs per holding was estimated to 4.61%. If the relative MSE for this variable had been defined as the sum of the sampling variance and the squared bias only it would have been 1.92%. Thus the simple response variance more than doubled the relative MSE.

At Statistics Sweden, the quinquennial Censuses of Population have been evaluated during the last decades. An evaluation based on a survey model is conducted only for labor force items, though. This is based on two sets of data:

i) The labor force items are matched, by means of personal identity numbers, with Labor Force Survey (LFS) data collected during the same time as the census is taken. Some questions are added to the standard LFS questionnaire to meet the census definitions. The LFS data are then regarded as (independent) reinterview data.

ii) A reconciliation of the census-LFS match provide better values than the census data.

By applying the U.S. Census Bureau model to the labor force items, we can derive the bias and the simple response variance of the census estimates. Since the census data are collected by mail, the correlated response variance is assumed to be small. The model assumptions are, however, not fully met for two reasons. First, the LFS data are collected by telephone, thus the LFS is not a replication of the census under the same general conditions. Second, the matching and reconciliation procedures can not be assumed to provide "true" data. If the same unit is misclassified in both surveys, the case is not reconciled. By reason of that, the estimates of the bias (the net difference rate) and of the simple response variance (g/2) are not unbiased.

Table 6 shows the estimated error components for some items for the 1980 Census data. The simple response variance is estimated for typical population sizes of communes (28200) and parishes (3100).

In general, the simple response variance is small compared to the squared bias. This (and the assumption that the correlated response variance is small) has led to the conclusion that the bias is the major problem in the Swedish population census. However, we have neither studied the impact of the editing personnel on the estimates, nor other items except those in Table 6.

No other application of survey models has been reported at Statistics Sweden.

Table 6 Estimates of bias and simple response variance for selected labor force items in the Swedish 1980 Census.

		Per cent in census	Squared bias x 10^4	Simple response variance x 10^4	
				Area of population size	
				28200	3100
Hours/week at work					
	35 - ∞	55.4	1.80	0.02	0.16
	20 - 34	14.4	0.12	0.02	0.15
	1 - 19	4.3	0.01	0.01	0.09
Outside the labor force					
	Student	7.3	0.41	0.00	0.03
	Others	18.5	0.22	0.01	0.06

Source: I Lyberg (1986)

Table 7 Cumulative distribution of ρ^*-values in three experiments conducted in United Kingdom.

ρ^*	Southampton	North Yorkshire	Milton Keynes
	(163 items)	(175 items)	(61 items)
Estimated interviewer variance as percentage of total:	%	%	%
Under 1 %	36	45	48
Under 3 %	63	74	80
Under 5 %	80	90	92
All items	100	100	100

Source: Collins (1980)

In the United Kingdom, three experiments on interviewer variability were reported by the Social and Community Planning Research (SCPR) as a part of a methodological research programme. The experiments took place in Southampton, North Yorkshire, and Milton Keynes. The questionnaires dealt with the problems faced by the disabled, environmental preferences, and different aspects on living and working, respectively. In the North Yorkshire experiment each interviewer was assigned a sample from the entire experimental area, while in Southampton and in Milton Keynes the survey area was subdivided into smaller areas. The interviewers were assigned samples within these subdivided areas. The Kish (1962) model was applied and the intraclass correlation coefficients, ρ^*, were derived for 399 items in the three experiments. Table 7 from Collins (1980) shows the cumulative distributions of the ρ^*-values.

The estimated interviewer effects were generally larger in the Southampton study than in the other studies. It was assumed that this difference was due to the nature of the study of the disabled, but it could also reflect greater problems with the questionnaire in this than in the other studies.

INDIA

The early research on non-sampling errors and survey models at the Indian Statistical Institute and the Indian Council of Agricultural Research has already been mentioned.

A recent example of a survey model application is the evaluation of the 1981 Indian Censuses. Two weeks after the census a reinterview was conducted with about 91500 households. The reinterview enumerators were better trained and supervised compared to the regular enumerators and thus the reinterview data were regarded as closer to the true values. Table 8 shows the type of analysis that was conducted. The index of inconsistency and the net difference rate were presented for the variables literacy and age by sex and rural/urban area.

Table 8 Index of inconsistency and net difference rate by sex, literacy, and age in the 1981 Indian Census.

Characteristic	Index of inconsistency		Net difference rate	
	Male	Female	Male	Female
LITERACY				
Urban	7.71	7.71	.11	-.57
Rural	9.19	9.39	.47	-.26
AGE				
0 to 4 years:				
Urban	13.75	17.35	.52	-.01
Rural	7.97	5.24	.20	.17
5 to 9 years:				
Urban	14.73	16.04	.21	-.22
Rural	9.44	8.65	.14	-.95
20 to 24 years:				
Urban	15.99	18.27	-.34	-.79
Rural	21.95	28.40	-.41	-.25
25 to 29 years:				
Urban	19.94	22.61	.19	-.17
Rural	23.80	26.11	.14	.08
45 to 49 years:				
Urban	24.92	29.76	.66	.23
Rural	31.06	34.06	-.31	-.39
50 to 54 years:				
Urban	31.51	23.00	-.15	.08
Rural	31.80	28.62	-.13	.38
65 to 69 years:				
Urban	38.80	34.77	-.37	.33
Rural	38.80	41.71	-.13	-.17
70 years or older:				
Urban	16.81	19.66	.00	.03
Rural	14.50	16.72	.21	.00

Source: Indian Administrative Service (1982)

The net difference rate was regarded as an approximation to the level of under- or overreporting. Similar analyses were presented for labor force items.

An interesting application of survey models has been reported by the Maharashtra Association for the Cultivation of Science in Pune, where P.V. Sukhatme is currently working. In recent years, he has applied his model to biological problems, see, e.g., Sukhatme (1985). Two quotations from Sukhatme (1986) may illustrate his view.

"The main purpose of my model is to test whether enumerator biases in sample surveys are under control. I have used it not only for testing whether the biases cancel each other on average in estimating the population mean but also for estimating the total error and components thereof. In most of my recent work, this has taken the form of testing the hierarchical non-random structure of variation such as one finds in the study of behavioural traits in man and society. Testing for the stability of intra-individual variance and its implications for testing for homeostasis and for higher order feed-back under a sustained perturbation of environment is the most important development resulting from the application of my model to biological problems. I have shown that the stabilized portion of the intra-individual variation has the property of heritability, i.e., of cultural inheritance as distinct from biological inheritance."

and

"I have also used the model for continually monitoring quality of data on morbidity in children in villages in India. When I introduced the model in 1952, I did it primarily for testing of quality of data in crop surveys and censuses. Extending the application to monitor morbidity was even more fascinating because morbidity is conceptually more difficult to assess than agricultural products."

OTHER APPLICATIONS: The World Fertility Survey

The World Fertility Survey (WFS) was conducted in several developing countries, mainly during the 1970s. The variables under study were various fertility items, some of which rather sensitive and thus subject to response errors. Typically, the data were collected in personal interviews. As part of the WFS analysis programme, evaluations of the quality of the data were conducted in many countries. These evaluations were based on interpenetrated survey designs using reinterview studies. These were also interpenetrated as in the experimental design used by Fellegi (1964). A survey model was developed that resembles the U.S. Census Bureau model as well as the simple ANOVA models described by Kish and Sukhatme, see, e.g., O'Muircheartaigh and Marckwardt (1980).

The analysis includes the estimation of the index of inconsistency (I) as well as of the measure ρI, the product of the intra-interviewer correlation, ρ, and I, ρI is basically the same parameter as Kish's intraclass correlation coefficient, ρ^*. In that paper, estimates of ρI from the evaluations in Peru and Leshoto were presented for fertility variables. Questions on these variables can be difficult or sensitive for the interviewers to ask. The estimates ranged from 0.00 to 0.19 (mean = 0.7). Indices of inconsistency were derived for several variables in Fiji and Peru, and for some variables in Indonesia, as presented in Table 9.

The pattern is similar in the different countries. For eight of the variables, I is less than 10%, which is usually considered a low value. The remaining seven variables show, however, much higher values. The most serious signs of inconsistency are displayed for "first birth interval," "desired number of children," and "ever use of contraception," for which I in one country was around 50%.

Table 9. Index of inconsistency for selected WFS items for Fiji, Peru and Indonesia.

Variable	Fiji	Peru	Indonesia
First birth interval	.35	.60	/
Last closed interval	.18	.29	/
Year of first birth	.01	.02	.04
Children ever born	.02	.02	.02
Month of last birth	.02	.02	/
Year of last birth	.02	.02	/
Desired no. of children	.25	.59	.32
Current age	.02	.02	.06
Age at marriage	.20	.20	.29
Year of marriage	.02	.04	/
Age group	.03	.03	/
Whether worked	.31	.30	/
Ever-use of contraception	.48	.35	/
Births in past 5 years	.06	.09	/

Source: O'Muircheartaigh and Marckwardt (1980).

4 DISCUSSION

4.1 The Early Models

The usefulness of survey models for quality assessment and as a basis for allocating survey resources between different components of the error is well-known among survey practitioners all over the world. We have in this paper reviewed several important applications of early survey models. Some of them, like the U.S. Census Bureau model and the Kish model, have proved a broad applicability. They have been used not only within the agency for which they primarily were developed, but also in other contexts, where the survey environment might be quite different.

Despite this, we have reasons to ask why survey models have not been even more broadly applied in survey quality work. After all, outside the United States, Canada, and perhaps some other countries, applications of survey models are rare. Applications might appear in certain experiments and surveys, but often as a result of a single researchers' interest in the field. These scattered applications often concern a small number of variables only. There are different reasons for this state of affairs.

i) The models do not cover all possible error sources. The survey models we reviewed in Section 3 mainly concern content errors and sampling errors. They do not account for frame errors, coverage errors, and non-response errors, and thus their usefulness are limited, at least in the context of total survey design.

ii) The models are based on assumptions that are not met in most survey situations. For instance, when estimating the simple response variance, a common assumption is that reinterviews are independent of the original interviews and equally distributed. Bailar and Dalenius (1969) showed that the simple response variance component could be estimated even if the reinterviews and the original interviews were permitted to be dependent. This, however, requires a second reinterview survey, which for practical reasons may be difficult and certainly expensive to implement. Another example is that the early models almost always assume a simple sampling design. While, in practise, survey designs are usually much more complex.

iii) The costs. The experimental designs necessary for estimating the components in the early models are expensive to implement. When personal interviews are used, interpenetration of the interviewers workloads might be enormously expensive if the study area is large. This problem can be diminished if the population under study and the population of interviewers are stratified and the model is applied in each stratum, as suggested by Sukhatme, or if the populations are grouped as assumed in the Hansen et.al. (1951) model. However, even these designs would be expensive for organizations like Statistics Sweden, where the interviewers are spread out over the country and work alone in large areas. Another practical problem associated with interpenetration of interviewer assignments occurs in countries where the sampling units are individuals (and not housing units). Tracking respondents then becomes an important part of the interviewers work and, as this requires good knowledge of the local environment, interpenetrating would lead to increased nonresponse problems. In telephone interviews, the cost problems with interpenetration can almost be ignored, but the non-response problem can not.

The reinterview costs are considerable since large reinterview samples are needed for estimating the simple response variance component with an acceptable precision.

4.2 Recent contributions

The subsequent development of survey models after 1970 has to some degree coped with the drawbacks of earlier models. It is beyond the scope of this paper to review this research, but we may briefly mention a few examples. At the Research Triangle Institute (RTI) in North Carolina, Koch (1973) extended the U.S. Bureau of the Census' model to the multivariate case for continuous and qualitative variables. His model is not confined to simple random sampling, but may be applied to multistage clustered unequal probability sampling designs. Other contributions from the RTI include Koop (1974) and Lessler (1974). Efforts have also been made to include the nonresponse error in survey models, see Platek, Singh, and Tremblay (1977), Lessler (1983), and Platek and Gray (1983). Others have developed methods for estimating error components without adopting special designs like interpenetration or repetition. Hartley and Rao (1978) presented a mixed ANOVA model and recommended a synthetic model based method of variance component estimation. Biemer (1986a) generalized the U.S. Census Bureau model and introduced a spatial autoregressive model to estimate the true values. With this model the error components can be estimated from a more simple design.

The above indicates that developments in both variance decomposition models and ANOVA models have been aimed at eliminating some of the drawbacks listed in Section 4.1 above. But this has led to more complex models and the complexity itself might be an obstacle to their application. Another school of survey practitioners continues to use the early models. The appeal of these models is their simplicity - or as Kish (1986) puts it: "My central philosophy appeared in print here and there: Go for a lot of data, because the parameters vary so much in several dimensions. Then we cannot afford the expenses of a few "critical experiments," because there won't be any. That means: be robust, rough, approximate. For example, on interviewer variance, the simple model of my 1962 paper and the tables in 13.2 of Survey Sampling are good enough. This model has been well used by R. Groves in several papers. The finer models that include the respondents' variation separate from the interviewers' cannot be measured in practise." Other survey model researchers stress the need for developing better models, as indicated in the following quotations:

"I do not know that we have appropriate models in mind for errors of response, models that take into account the correlations and so on that exist, and when this is done on a within survey basis we will collect very little data that will help us. When it is done by comparing with outside records, the biases are often enormous and no one knows how to use the information. I think the Bureau has done more in the way of nonsampling error research than probably anyone else and certainly much more than those engaged in sampling, but I still don't think we know what are the appropriate models and I doubt very much that they will be general. I think they will depend very much on the survey technique, and I think therefore, that this problem is difficult. I am delighted with the development of models but maybe we will have better approaches than we've had so far". (Madow, 1981).

"I believe there are two major reasons for the relatively slow development of non-sampling error theory and control:

(1) There exists no comprehensive model which has direct practical utility. Models which do exist are either so general that they provide no guidance in the design of surveys, or are quite application specific with limited generalizability. As a result we are not able to enlist effectively the help of statistical theory, academic statisticians are not adequately interested, and most importantly, we miss the flashes of insight which analytically exploited good error models can provide. See the impact of the Hansen-Hurwitz-Bershad paper on census taking for a positive illustration of the point, or even the impact of my papers on record linkage and imputation.

(2) Bias measurement is obviously difficult, expensive, and often outright impossible. So not only do we miss the existence of good models, we also have a poor empirical base - for assessment, motivation, or to estimate model parameters." (Fellegi, 1986).

4.3 The future

We have seen that survey models are founded in sampling theory and that ANOVA models are applicable to the estimation of survey errors. It is indicated in the above quotations by Madow and Fellegi that other approaches are needed. Biemer (1986b) requested new methods for analyzing survey data: "I believe evaluation study data, in general, have been underanalyzed owing to a lack of resources and a dearth of analytical techniques that exploit the data to their fullest." When looking for other approaches to survey error models, sampling statisticians should perhaps become more familiar with methods of dealing with measurement error that are outside of sampling and even outside of statistics. Biemer continues: "Today, sampling statisticians are discovering the models of the cognitive psychologists and psychometricians and adopting these to survey modeling; for example, latent trait theory (see Clyde Tucker's paper in the 1985 Proceedings of the ASA)."

Survey model researchers certainly have ideas on how their own models could be developed. Sukhatme (1986) and Koch (1986) both mention longitudinal studies as an area of extension. Koch also mentions applications to extended classes of estimators. In a more general perspective, Biemer (1986b) suggests the following areas for future research on survey models: (i) model validation and (ii) models for exploring the causes of non-sampling errors. He gives the following examples of questions. For model validation:

- How well are the assumptions of independence and identically distributed observations satisfied for reinterview studies that measure the simple response variance component?

- Is a linear additive model appropriate for modeling interviewer effects?

- Does reconciliation in a reinterview survey produce responses which are closer to the truth?"

For exploring the causes of non-sampling error:

"- What is the role of log-linear modeling in survey error modeling?

- How can the estimation of response error be routinized in sample surveys?

- What models are appropriate for linking readily available non-sampling error "indicators" - such as non-response rates, edit failure rates, interviewing learning curves, interviewer turnover rates, etc. - to the total mean squared error."

The ideas put forward by these statisticians are interesting and form a good basis for future research. To the list we might add the issue of precision of estimates of the error components. This is indeed a neglected area, but a very important one. It is unfortunate if we, after having conducted expensive and complex evaluations, end up with error estimates that are extremely imprecise. Then we might as well rely on intuition and devote more effort to preventive measures, i.e., dealing with errors at their sources and hope for the best when it comes to the total error.

ACKNOWLEDGEMENTS

I would like to thank the following persons who kindly supplied information. B.A. Bailar, P.P. Biemer, T. Dalenius, J.C. Deville, L. Fabbris, I. Fellegi, G.B. Gray, M. Hansen, W. Keller, L. Kish, G. Koch, J. Koop, J.T. Lessler, L. Lyberg, I. Lyberg, R. Platek, J.N.K. Rao, M. Ribe, H. Strecker, P.V. Sukhatme, G. Theodore, A. Sunter, V. Verma, and S.S. Zarkovich. I am, however, responsible for any defects that might be found in the paper.

I am also indepted to Avdelningsdirektör Patricia A. Dean, Statistics Sweden, for her expert editing and to Ms. Sivan Rosén and Ms. Kerstin Averbäck, Statistics Sweden, for their very skillful typing.

References

ABUL-ELA, A.A., GREENBERG, B.G., and HORVITZ, D.G (1967). A Multiproportions Randomized Response Model. Journal of the American Statistical Association, 62, 990-1008.

BAILAR, B.A., and BIEMER, P.P. (1984). Some Methods for Evaluating Nonsampling Error in Household Censuses and Surveys. In Rao, P.S.R.S, and Sedransk, J. (ed.:s): W.G. Cocran's Impact on Statistics. 253-274. New York: John Wiley & Sons.

BAILAR, B.A., and DALENIUS, T. (1969). Estimating the Response Variance Components of the U.S. Bureau of the Census' Survey Model. Sankhya: Series B, 341-360.

BAILEY, L., MOORE, T.F., and BAILAR, B.A. (1978). An Interviewer Variance Study for the Eight Impact Cities of the National Crime Survey Cities Sample. Journal of the American Statistical Association, 73, 16-23.

BIEMER, P.P., BUSHERY, J., KOSTANICH, D., and NASH, F. (1985). The Feasibility of Interviewer Variance Estimation in the CPS: Final Report. Memo. U.S. Bureau of the Census.

BIEMER, P.P. (1986a). A Spatial Modeling Approach to the Evaluation of Census Nonsampling Error. Proceedings of the Second Annual Research Conference, 7-20. U.S. Bureau of the Census.

BIEMER, P.P. (1986b). Personal communication.

CASADY, R.J. (1966). A Model for Analysis of Reinterview Data in the NCHS Health Interview Survey. Memo. U.S. Department of Health, Education, and Welfare.

COCHRAN, W.G. (1968). Errors of Measurement in Statistics. Technometrics, 10, 637-666.

COLLINS, M. (1980). Interviewer Variability: A Review of the Problem. Methodological Working Paper No. 19. Social and Community Planning Research. London.

DALENIUS, T. (1974). Ends and Means of Total Survey Design. The Errors in Surveys research project, Institute of Statistics, University of Stockholm.

DEMING, W.E. (1944). On Errors in Surveys. American Sociological Review, 9, 359-369.

DEMING, W.E., and GEOFFREY, L. (1941). On Sample Inspection in the Processing of Census Returns. Journal of the American Statistical Association, 36, 351-360.

DES RAJ (1968). Sampling Theory. New York: McGraw-Hill Book Co.

ECKLER, A.R., and PRITZKER, L. (1951). Measuring the Accuracy of Enumerative Surveys. Bulletin of the International Statistical Institute, 33:4, 7-24.

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

FELLEGI, I. (1986). Personal communication.

GROVES, R.M., and MAGILAVY, L.J. (1986). Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys. Public Opinion Quarterly, 50, 251-266.

HANSEN, M.H., and HURWITZ, W.N. (1946). The Problem of Non-response in Sample Surveys. Journal of the American Statistical Association, 41, 517-529.

HANSEN, M.H., and TEPPING, B.J. (1969). Progress and Problems in Survey Methods and Theory Illustrated by the Work of the United States Bureau of the Census. In Johnson, N.L, and Smith, H. (ed.:s.): New Developments in Survey Sampling. New York: Wiley-Interscience. 1-26.

HANSEN, M.H., HURWITZ W.N. and BERSHAD, M.A. (1961). Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38:2, 359-374.

HANSEN, M.H., HURWITZ, W.N., and PRITZKER, L. (1964). The Estimation and Interpenetration of Gross Differences and the Simple Response Variance. In C.R. Rao (ed.): Contributions to Statistics. Calcutta: Statistical Publishing Society.

HANSEN, M.H., HURWITZ, W.N., MARKS, E.S., and MAULDIN, W.P. (1951). Response Errors in Surveys. Journal of the American Statistical Association, 46, 147-190.

HARTLEY, H.O., and RAO, J.N.K. (1978). Estimation of Nonsampling Variance Components in Sample Surveys. In N.K. Namboodiri(ed.): Survey Sampling and Measurement. New York: Academic Press.

INDIAN ADMINISTRATIVE SERVICE (1982). Census of India 1981: Report on the Post-Enumeration Check. Series 1, No. 4, New Delhi.

JABINE, T.B., and TEPPING, B.J. (1973). Controlling the Quality of Occupation and Industry Data. Bulletin of the International Statistical Institute, 45, 360-389.

KISH, L. (1962). Studies of Interviewer Variance for Attitudinal Variables. Journal of the American Statistical Association, 57, 92-115.

KISH, L. (1986). Personal communication.

KISH, L., and LANSING, J.B. (1954). Response Errors in Estimating the Value of Homes. Journal of the American Statistical Association, 49, 520-538.

KOCH, G.G. (1973). 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, 906-913.

KOCH, G.G. (1986). Personal communication.

KOOP, J.C. (1974). Notes for a Unified Theory of Estimation for Sample Surveys Taking Into Account Response Errors. Metrika, 21, 19-39.

KROTKI, K.P., and HILL, C.J. (1978). A Comparison of Correlated Response Variance Estimates Obtained in the 1961, 1971 and 1976 Censuses. Survey Methodology, 4, 87-99.

LESSLER, J.T. (1974). A Double Sampling Scheme Model for Eliminating Measurement Process Bias and Estimating Measurement Errors in Surveys. Mimeo Series No. 949. Chapel Hill, N.C.: University of North Carolina

LESSLER, J.T. (1983). An Expanded Survey Error Model. In: Incomplete Data in Sample Surveys, 3, 259-270. New York: Academic Press.

LYBERG, I. (1986). Om evalveringsmetoder for Folk- och bostadsräkningen (FoB). Memo. Statistics Sweden. (In Swedish)

MADOW, W. (1981). General discussion of the meeting on non-sampling errors. In Krewski, Platek, Rao (ed.:s): Current Topics in Survey Sampling. New York: Academic Press.

MAHALANOBIS, P.C. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, 109, 327-378.

MURTHY, M. (1967). Sampling Theory and Methods. Calcutta: Statistical Publishing Society.

O'MUIRCHEARTAIGH, C.A., and MARCKWARDT, A.M. (1980). An Assessment of the Reliability of WFS Data. World Fertility Survey Conference. Methodology Session No. 6.

PALMER, G.L. (1943). Factors in the Variability of Response in Enumerative Studies. Journal of the American Statistical Association, 38, 143-152.

PLATEK, R., and GRAY, G.B. (1983). Imputation Metohodology: Total Survey Error. In: Incomplete Data in Sample Surveys, 2, 249-333. New York: Academic Press.

PLATEK, R., SINGH, M.P., and TREMBLAY, V. (1977). Adjustment for Non-Response in Surveys. Survey Methodology, 3, 1-24.

RICE, S.A. (1929). Contagious Bias in the Interview. American Journal of Sociology, 35, 420-423.

SANCHEZ-CRESPO, J.L. (1973). Balance Between Sampling and Non-sampling Errors in Spanish Official Statistics. Bulletin of the International Statistical Institute, 45, 329-367.

SANCHEZ-CRESPO, J.L. (1981). Spanish Experience on the Estimation of Some Components of the Total Error. Statistical Commission and Economic Comission for Europe. Conference of European Statisticians, June 1-4, 1981.

STOCK, J.S., and HOCHSTIM, J.R. (1951). A Method of Measuring Interviewer Variability. Public Opinion Quarterly, 15, 322-334.

STRECKER, H., and WIEGERT, R. (1981). Fehler in Statistischen Erhebungen, Darstellung anhand von Beispielen. In von Mückl, W.J., and Ott, A.E. (ed.:s): Wirdschaftstheorie und Wirdschaftspolitik. Gedenkschrift für Erich Preiser, Passau. 439-458. (In German).

STRECKER, H., WIEGERT, R., and KAFKA, K. (1984). Practical Determination of a Response Variance on the Basis on Survey Models with Reenumerations. Jahrbücher für Nationalökonomie und Statistik, 199, 1-31.

SUKHATME, P.V. (1947). The Problem of Plot Size in Large-Scale Yield Surveys. Journal of the American Statistical Association, 42, 297-310.

SUKHATME, P.V. (1954). Sampling Theory of Surveys With Applications. Rome: U.N. Food and Agriculture Organization. Published by: The Iowa State College Press and The Indian Society of Agricultural Research.

SUKHATME, P.V. (1985). The Nature of Energy Requirement and Its Implications for Measurement of Undernutrition. Bulletin of the International Statistical Institute, 51:4, pp. 30.1: 1-16.

SUKHATME, P.V. (1986). Personal communication.

SUKHATME, P.V., and SETH, G.R. (1952). Non-sampling Errors in Surveys. Journal of the Indian Society of Agricultural Statistics, 4, 5-41.

SWENSSON, B. (1969). Alternativa surveymodeller för dikotoma variabler. Report no. 19 of the Errors in Surveys research project. Institute of Statistics. University of Stockholm. (In Swdish).

TEPPING, B.J., and BOLAND, K.L. (1972). Response Variance in the Current Population Survey. U.S. Working Paper No. 36, Bureau of the Census. U.S. Government Printing Office, Washington, D.C.

U.S. BUREAU OF THE CENSUS (1978). The Current Population Survey. Design and Methodology. Technical Paper 40. U.S. Government Printing Office, Washington, D.C.

U.S. BUREAU OF THE CENSUS (1985). Evaulating Censuses of Population and Housing, Statistical Training Document, ISP-TR-5, Washington, D.C.

WARGENTIN, P. (1780). Undersökning om Folk-utflyttningen, såväl utur hela riket, som utur hvart Höfdingedöme särskilt, i anledning av Tabellverket för åren 1750 till 1773. Kongl. Vetenskaps Academiens handlingar, Stockholm. (In Swedish).

WARNER, S.L. (1965). A Randomized Response Technique for Eliminating Evasive Answer Bias. Journal of the American Stastical Association, 60, 63-69.

ZARKOVIC, S.S. (1966). Quality of Statistical Data. Rome: Food and Agriculture Organization of the United Nations.

Tidigare nummer av Promemorior från U/STM:
NR

1 Bayesianska idéer vid planeringen av sample surveys. Lars Lyberg (1978-11-01)

2 Litteraturförteckning över artiklar om kontingenstabeller. Anders Andersson (1978-11-07)

3 En presentation av Box-Jenkins metod för analys och prognos av tidsserier. Åke Holmén (1979-12-20)

4 Handledning i AID-analys. Anders Norberg (1980-10-22)

5 Utredning angående statistisk analysverksamhet vid SCB: Slutrapport. P/STM, Analysprojektet (1980-10-31)

6 Metoder för evalvering av noggrannheten i SCBs statistik. En översikt. Jörgen Dalén (1981-03-02)

7 Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Gösta Forsman och Tomas Garås (1982-11-01)

8 How large must the sample size be? Nominal confidence levels versus actual coverage probabilities in simple random sampling. Jörgen Dalén (1983-02-14)

9 Regression analysis and ratio analysis for domains. A randomization theory approach. Eva Elvers, Carl Erik Särndal, Jan Wretman och Göran Örnberg (1983-06-20)

10 Current survey research at Statistics Sweden. Lars Lyberg, Bengt Swensson och Jan Håkan Wretman (1983-09-01)

11 Utjämningsmetoder vid nivåkorrigering av tidsserier med tillämpning på nationalräkenskapsdata. Lars-Otto Sjöberg (1984-01-11)

12 Regressionsanalys för f d statistikstuderande. Harry Lütjohann (1984-02-01)

13 Estimating Gini and Entropy inequality parameters. Fredrik Nygård och Arne Sandström (1985-01-09)

14 Income inequality measures based on sample surveys. Fredrik Nygård och Arne Sandström (1985-05-20)

15 Granskning och evalvering av surveymodeller, tiden före 1960. Gösta Forsman (1985-05-30)

16 Variance estimators of the Gini coefficient - simple random sampling. Arne Sandström, Jan Wretman och Bertil Waldén (Memo, Februari 1985)

17 Variance estimators of the Gini coefficient - probability sampling. Arne Sandström, Jan Wretman och Bertil Waldén (1985-07-05)

18 Reconciling tables and margins using least-squares. Harry Lütjohann (1985-08-01)

19 *Ersättningens och uppgiftslämnarbördans betydelse för kvaliteten i undersökningarna om hushållens utgifter.* Håkan L. Lindström (1985-11-29)

20 A general view of estimation for two phases of selection. Carl-Erik Särndal och Bengt Swensson (1985-12-05)

21 On the use of automated coding at Statistics Sweden. Lars Lyberg (1986-01-16)

22 Quality Control of Coding Operations at Statistics Sweden. Lars Lyberg (1986-03-20)

23 A General View of Nonresponse Bias in Some Sample Surveys of the Swedish Population. Håkan L Lindström (1986-05-16)

24 Nonresponse rates in 1970 - 1985 in surveys of Individuals and Households. Håkan L. Lindström och Pat Dean (1986-06-06)

25 Two Evaluation Studies of Small Area Estimation Methods: The Case of Estimating Population Characteristics in Swedish Municipalities for the Intercensal Period. Sixten Lundström (1986-10-14)

26 A Survey Practitioner's Notion of Nonresponse. Richard Platek (1986-10-20)

27 Factors to be Considered in Developing a Reinterview Program and Interviewer Debriefings at SCB. Dawn D. Nelson (1986-10-24)

28 Utredning kring bortfallet i samband med den s k Metropolitdebatten. Sixten Lundström (1986-11-10)

29 Om HINK-undersökningens design- och allokeringsproblematik. Bengt Rosén (1986-10-06)

30 Om HINK-undersökningens estimationsförfarande. Bengt Rosén (1987-01-28)

31 Program för minskat bortfall i SCBs individ- och hushållsundersökningar - förslag från aktionsgruppen för uppgiftslämnarfrågor. Lars Lundgren (1987-02-20)

32 Practical estimators of a population total which reduce the impact of large observations. Jörgen Dalén (87-03-03)

33 A Multivariate Approach to Social Inequality. M Ribe (1987-05-05)

34 BORTFALLSBAROMETERN NR 2. Bortfall vid I-avdelningen, AKU och F-avdelningen t o m 1986. Håkan L. Lindström, Peter Lundquist och Tomas Garås (1987-08-01)

35 Use of Laboratory Methods and Cognitive Science for the Design and Testing of Questionnaires. Handbook of Methods. Judith T. Lessler (June 1987)

36 Some Optimal Problems When Estimating Household Data on the Basis of a "Primary", Stratified Sample of Individuals.
Bengt Rosén (1987-08-10)

Kvarvarande exemplar av ovanstående promemorior kan rekvireras från Elseliv Lindfors, U/STM, SCB, 115 81 Stockholm, eller per telefon 08 7834178

GARNISONSTRYCKERIET ☆ STOCKHOLM ☆ 1987 ☆