Journal of Official Statistics
Vol. 9, No. 3, 1993, pp. 611–621

Total Survey Design – Application to a Collection of the Construction Industry

Susan J. Linacre[1] *and Dennis J. Trewin*[2]

Abstract: Statistical agencies are placing increasing emphasis on non-sampling error in surveys as well as efficient collection design. This paper argues that by considering both sampling and non-sampling errors, and optimising resource allocations so that resources are deployed where error reduction is most effective, superior survey designs can be derived. The paper describes the application of the Total Survey Design approach to one collection – the Construction Industry Survey (CIS). The study is a simplification in many respects, but highlights the potential gains in efficiency from such an approach.

Key words: Data quality; optimal resource allocation; non-sampling error.

1. Introduction

Groves (1989) points out that although sampling statisticians have devoted their efforts to a thorough understanding of the sampling error properties of estimates, it has been mostly independent of work undertaken to reduce and measure the various non-sampling errors which can affect survey data. He points out that a more logical approach is to minimise the mean squared error of estimates, considering all sources of error, rather than just the sampling variance. It is the objective of this paper to illustrate an application of this technique.

A growing appreciation of the relative importance of non-sampling error in survey estimates has led to an increase in evaluation studies conducted in conjunction with statistical collections. Evaluation studies have several purposes. First, they provide indicators, and sometimes measures of the reliability of the estimates, making more informed use of the statistics possible. Second, by better understanding the sources of error, it is possible to develop methods to reduce them, e.g., through better forms design, interviewer training or frame enhancement. In the Australian Bureau of Statistics (ABS), as in many statistical organisations, the pressure to reduce resources spent on collections becomes greater each year. In this environment a major focus and a third use of evaluation studies should be to determine the most cost effective survey strategies. Although most survey design texts emphasise the importance of both sampling errors and non-sampling errors, relatively few have attempted "optimum" design using both survey error models (encompassing sampling and non-sampling errors) and cost models. Of course, optimisation of sam-

[1] Australian Bureau of Statistics, P.O. Box 10, Belconnen ACT 2616, Australia.
[2] Department of Statistics, P.O. Box 2922, Wellington, New Zealand.

pling error for a given survey budget is a common technique and described in even the most basic sampling texts. The process of determining the relative importance of sampling error and different types of non-sampling error, and then optimally allocating resources to minimise total error (usually mean squared error) is far less common. One early attempt at a total survey design approach was reported in Anderson et al. (1979). They compared the total survey error for three different approaches using re-interview data from a health survey.

Of course, such optimisation studies are not possible without cost data. This has also been a neglected field. Cost models tend to have been specified in terms of the number of sample units, number of units selected at each stage of selection in a multi-stage design, etc. This may allow for the optimisation of the sample design (i.e., minimisation of sampling variance) but not the optimisation of the survey design (i.e., minimisation of mean squared error). A difficulty with cost models is the lack of continuity that often exists, particularly when significantly different methodologies are considered (e.g., telephone versus face to face interviewing).

The next section of this paper describes a study using information on the variable "total sales" using evaluation studies of the 1984–85 CIS. The study attempted to use available data to determine a near optimal allocation of resources from the survey, taking into account coverage error, quality of the sampling frame, sampling error, alternative enumeration strategies, and non-response error.

As the study was a feasibility study, some simplifications were made, and where data from the earlier evaluation studies were no longer available, informed estimates were used. However, the study does provide a realistic example of the gains to be had, in accuracy or cost terms, from understanding the extent to which resources spent on error reduction tasks actually reduce the overall resulting error in the estimates.

The third section of the paper briefly discusses some of the difficulties of implementing a Total Survey Design approach.

2. A Study of Optimal Allocation of Resources to Minimise Total Error

As noted above, evaluation studies are often used to determine effective methods to reduce non-sampling error. A question that frequently arises in practice is how much of the resources available for a collection should be spent on each facet of the collection. For example, how much should be spent on setting up a good quality frame, how much, if any, on pilot testing, how much on field enumeration in preference to mail or telephone enumeration, on intensive non-response follow up, etc. Each of these "error reduction" tasks requires resources and the problem is to minimise the total overall error given the resource constraints. This section illustrates how various evaluation studies run in conjunction with the 1984–85 CIS were used to estimate the cost effectiveness of a variety of options in terms of the resultant level of total error (i.e., mean square error).

2.1. 1984–85 CIS Evaluation Studies

The CIS was first conducted in 1978–79, was repeated for 1984–85 and 1988–89, and is to be conducted approximately every five years. Similar surveys of other sectors (e.g., manufacturing, retail, services) are conducted in intervening years. Lessons learned on the cost effectiveness of the design of the 1984–85 CIS could be applied in subsequent surveys not only of the construction industry but of other industry sectors.

Table 1. Options for resource expenditure considered in study

Coverage check and units survey	Taxation frame survey	Enumeration	Follow up for non-response
Coverage Check 1. No additional checks	1. Select sample from taxation frame (additional cost of setting up framework $0.1 m)	*Large Units*: 1. Mail with personal visit, intensive follow up during editing (Cost: $72.22 per unit)	1. Follow up appropriate to achieve 75% response overall (72.2%, 75%, 90% for small, medium and large units respectively) (save $0.85 per unit enumerated)
2. Registers of Builders (ROB) (Cost: $0.052 m)	2. No taxation frame survey (No cost)	2. Mail without personal visit, some follow up at editing (Cost: $30.00 per unit)	2. Follow up appropriate to achieve 93% response overall (92.5%, 93%, 96% for small, medium and large respectively (Cost: no savings)
3. ROB + Telecom Files (Cost: $0.184 m)		*Medium Units*: 1.Mail with occasional follow up at editing (Cost: $24.89 per unit)	
4. ROB + Telecom + Industry Association Membership Lists (Cost: $0.211 m)		*Small Units*: Data collection cost of (a) $0.5 m (b) $1.0 m (c) $1.5 m with options of	
Units Survey 1. No Units Survey 2. Conduct Units Survey (Cost: $0.6 m)		1. Field enumeration, no follow up at editing (Cost: $78.13 per unit) 2. Mail enumeration no follow up at editing (Cost: $20.00 per unit) 3. Taxation Data Substitution (Cost: $0.05 m)	

Table 1 outlines a number of different options of how the CIS budget could be expended. Indicative costs (in 1988–89 dollars) are also stated. Considering all "reasonable" combinations, 176 options are available. The variants are discussed in more details in the following sections.

2.1.1. Coverage checks

The 1984–85 CIS was a dual frame list survey. The first frame was the ABS register of businesses (derived from employer registrations) where a State × Industry × Size (employment) stratification was used. The second frame was one year out of date taxation with a Type of Business (company, partnership, etc) × Industry × Size (total business income) stratification. There was considerable overlap between these two frames.

A major coverage check was undertaken prior to the survey to update the construction sector on the ABS register, as its coverage was not good, particularly for businesses without employees. The sources accessed were lists of registered builders, Telecom files (Telecom being the national telephone company), and membership lists for key construction industry associations. A source indicator was applied to each unit and added to the register to enable an evaluation of the cost effectiveness of the various sources.

In addition, a complete enumeration of all units in the construction sector of the register (referred to below as a units survey) was undertaken prior to the survey proper to determine units no longer in business (defunct), out of scope, duplicates, etc. The enumeration also obtained revised industry and size information and information to ensure the organisational structure of large businesses was correctly represented. This was then used as the main frame for the stratified sample selection of units to be included in CIS. The units survey is a short form survey of the full construction population and its cost ($0.6 m) reflects the size of the population and the processing required to update the frame. Questions arise as to whether the coverage check and the units survey were worthwhile activities.

2.1.2. Taxation frame survey

The 1984–85 CIS design involved the selection of a sample of units on the taxation frame which was then matched to the register. Those not on the register were enumerated along with the sample of units from the ABS business register. Despite the extensive coverage enhancement work, the taxation frame contributed about 41% of the estimate of the total establishments (mostly small non-employers) operating during the year, 7% to the estimate of number of employees at 30 June 1985 and 10% to the estimate of total sales. With potential non-coverage biases of this size, it is clearly necessary to sample the taxation frame as well as the business register frame.

2.1.3. Enumeration errors

To improve the quality of response for the larger and more complex units, the largest 170 businesses in the construction sector received a personal visit to explain the survey and concepts being measured. Processing of forms for the largest units included intensive editing and personal follow-up with these businesses.

Large measurement error problems were anticipated in a sector dominated by very small businesses. We expected those completing forms for these businesses to have a poor understanding of financial records and many to have a poor command of English (being born in non-English speaking countries) in relation to form completion. For this reason, small units (employment less than eight persons or "non-companies" with total business income less than $500,000) were field enumerated by specially trained interviewers in the 1984–85 CIS.

The largest single item of expense for the 1984–85 CIS was $1.5 m for the enumeration of small units, including $1.0 m paid to field agents. For this reason a special study of its effectiveness was undertaken. To do this, an additional sample of about 1000 units was drawn and enumerated by mail. Response rates were compared with those for medium sized CIS units. Also, relative data quality was assessed. Experienced processing staff clerically examined each form to provide a subjective opinion, and put the data through one edit cycle.

The overall finding was that the rate of response for small units enumerated by mail compared favourably with the rate of response for medium units although the final level of response after extensive follow-up was lower. This is not surprising as the check sample of small units was voluntary. Also, based on the number of telephone queries required to clarify information provided, the quality of data for small units appeared to be on a par with that for medium units in the CIS proper.

Similar comparisons were made for field enumerated small units. As expected, the response rate was much higher for field enumeration and data quality also seemed to be superior. In particular, the number of edit queries for field enumerated businesses was much lower. So we have the situation where the quality of data from a mail survey of small units appears to be equivalent to the quality of data from a mail survey of medium units, but the quality of field enumerated data appears to be superior to both. However, it does raise the question as to whether the additional cost of the field phase is justified.

The above method of comparing enumeration errors is far from satisfactory. A more objective method should have been used such as re-interviews or comparisons from an independent but reasonably reliable source, e.g., taxation data. Some re-interview data were available for the larger units, but it was insufficient to provide reliable quantitative estimates. A sample based study of moderate size should be adequate to give a useful measure of the response bias.

A possible method of further reducing enumeration costs is to directly substitute taxation data for small units data. This approach was evaluated for a variety of potential cut-off levels and the difference between the full survey estimate and an estimate derived using "partial substitution" of taxation data was calculated. If tax data were substituted for all units with wages and salaries less than $200,000 the resulting estimate of total turnover is 5% lower than the ABS survey estimate. (There have been investigations underway into whether the ABS data or taxation data are likely to be the more accurate, but for the purpose of this paper, it is assumed that ABS data are more accurate.)

2.1.4. Follow-up for non-response

The effect of following up non-response was closely monitored in terms of response rates for the survey. All units in the completely enumerated sector were intensively followed up to obtain completed returns. Where a return was still not obtained it was clerically estimated on the basis of all information available about the unit. Follow-up was also undertaken for smaller units but imputation for non-response was based on stratum means. Although it is clearly desirable to increase response levels, there is a question as to the extent that non-response follow-up is cost effective.

2.2. Setting up a model relating total error to resources used

In order to allocate resources represented as a vector (r) to minimise total error (E), it is necessary to relate total error to resources used in each error reduction task: $E = f(\mathrm{r})$. The problem is then to minimise E subject to $C(\mathrm{r}) = \mathrm{c}$, a vector of resource constraints (money, time, specialist resources, respondent load, etc).

The problem with modelling the function f, was tackled by expressing the total error in terms of key intermediate parameters, p, which are themselves related in a relatively straightforward way to resources used on various collection tasks. These parameters include the number of operating units in population on each of the two frames (tax and ABS register), the percentage of population common to the two frames, the percentage of defunct units, the coefficient of variation of live units, the mean of live units, the sample size, the percentage reporting error per unit for each enumeration option, the percentage imputation bias for non-response, and the proportion of the in-scope population not covered by either frame.

If the total error can be written as a function of the parameters p, it can then be estimated for given allocations of resources on the various tasks. Some components of the total error could be estimated as continuous functions of resources used on the various collection tasks. For example, non-response imputation bias might easily be modelled against staff days spent on intensive follow-up and processing action. However, in this study total error has been estimated for point estimates using different combinations of resource allocations on the various error reducing activities. The approach becomes more obvious in Section 2.5.

2.3. Total error as a function of key parameters

The total error can be divided into a variance component and a bias component. This section describes how both variance and bias were estimated in terms of the key parameters described in the previous sub-section.

First, the sampling variance is a key contributor to the total error. Hartley (1962) provides estimates of the variance for dual frame estimators, and for this study it has been assumed that an optimum allocation of sample to each frame was used. To allow the variance to be estimated for the different resource allocation options, the sampling variance was expressed in terms of the means and coefficient of variation of the live population units, and the percentage of defunct units in the population. It has been assumed that the component of the variance contribution from other sources, e.g., response variance, which would differ from one option to another, was negligible compared to sampling variance.

The bias component can also be written in terms of parameters. The simplified model estimates bias as the sum of undercoverage bias, collection mode bias for the responding sector, and non-response bias for the non-responding sector. It has been assumed that all biases are in the same direction, which is negative.

Coverage bias was estimated as a percentage of the total estimate depending on the extent of coverage enhancement achieved by the coverage check option. A coverage bias of -2% of the estimate was assumed for the option with maximum coverage enhancement (even when the taxation framework is used). For options with less than maximum coverage enhancement, data on units added through each source were used to provide an estimate of undercoverage in the absence of that source. Collection mode bias was estimated from the percentage of the stratum responding and the stratum average bias per respondent for the enumeration method.

The non-response bias was modelled according to the non-response rate and the stratum average bias per non-respondent. The bias depended on a qualitative assessment although in future evaluation studies, some attempt should be made to measure non-response bias. The percentage imputed depended on the extent of effort devoted to follow-up.

2.4. *Data availability issues*

This study concentrated on estimating total error for the single variable turnover at the national level. The 1984–85 survey data were used to estimate coefficients of variation for the live units.

As the approach discussed in this paper depends on being able to estimate a number of key parameters that drive the total error formula it is important to look at whether estimates of all these parameters are likely to be available.

For the options set out in Table 1, the only parameters that were not available from the evaluation studies conducted were

i. match rates between the ABS register and the tax frame,
ii. the corresponding match rates prior to the units survey, and
iii. remaining undercoverage bias given access to all evaluation sources (including the tax framework).

Both parameters i and ii could have been obtained fairly easily on a sample basis if the need had been foreseen. Parameter iii could be estimated from external sources. For example, information from the monthly Labour Force Survey would provide an estimate of undercoverage of employment for the sector.

2.5. *Results*

For each of the point estimates derived and for each of the combinations of options in Table 1, measures of total collection cost and total error were obtained. These are shown in Figure 1. The issue of interest is the dispersion of root mean square error (RMSE) values possible for a given cost constraint. The worst option gives RMSE values up to 300% higher than the best for given fixed cost. For this study the only options worth considering are those indicated with circles, being on the boundary of the plot of simulations. In all other cases, a smaller RMSE can be obtained for the same cost or a reduced cost for the same RMSE. The viable options are described in Table 2. The designs used in 1984–85 and 1988–89 are also indicated in Figure 1.

According to this example the option used in 1984–85 has an RMSE of 3.12% and a cost of \$2.63 m. The option used for the 1988–89 CIS involved mail enumeration for small units, no coverage enhancement, no units survey but intensive follow-up for high response rates, with medium sample size. This has an RMSE of 5.98% and a cost of \$1.05 m. According to this study, neither appears to be optimal. Option 4 seems to provide a good balance between cost and error.

- This differs from the 1984–85 design in that there is no coverage check, no units survey and a smaller sample size for small units so that costs would be reduced considerably. Nevertheless there is only a relatively small increase in RMSE, because of the greater emphasis on large units.
- This differs from the 1988–89 design

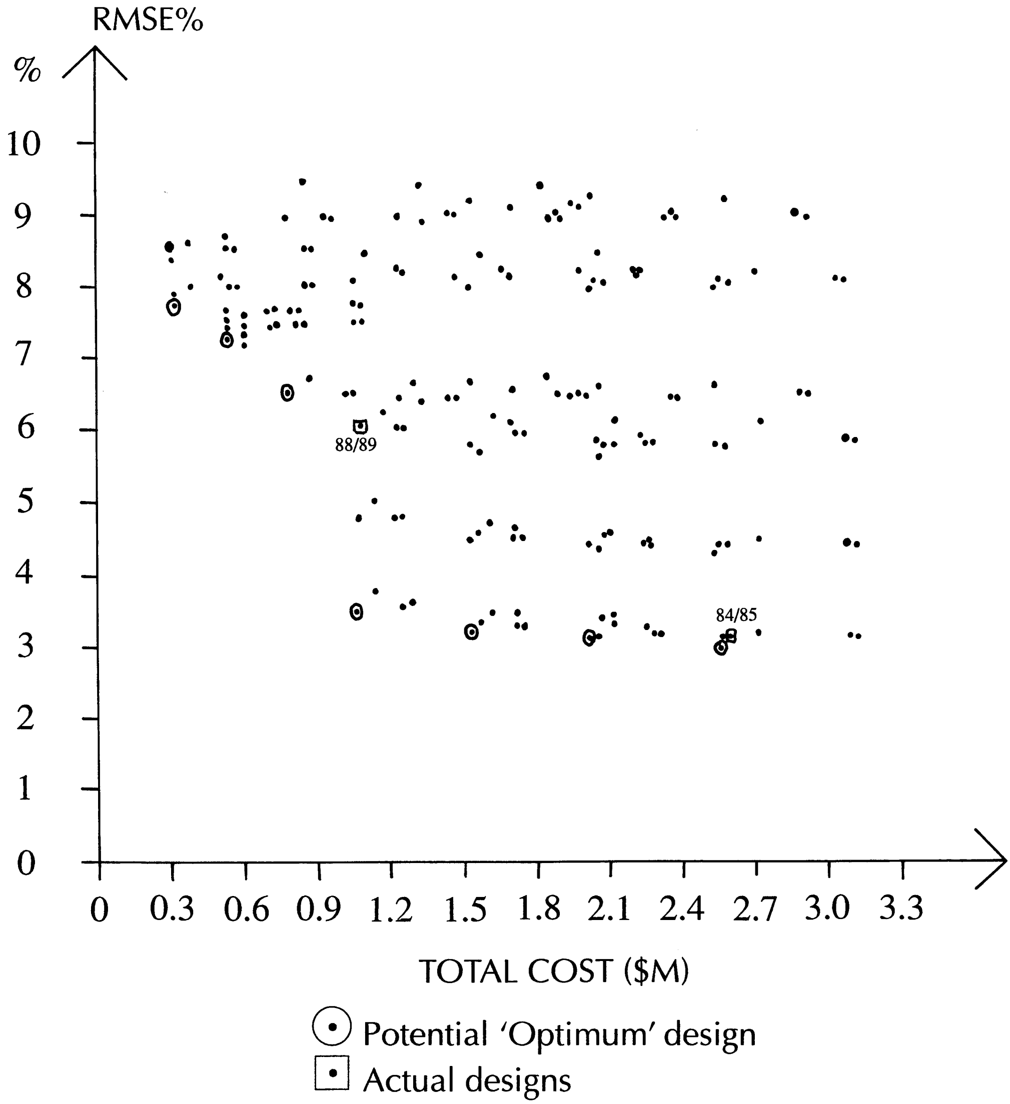

Fig. 1. RMSE% v, total cost for a number of resource allocation options

because field enumeration would be used but the total sample size is smaller. Therefore, there is a trade-off between higher sampling errors from a field enumeration and lower non-sampling errors from mail enumeration whereas total costs remain much the same.

The most significant changes made as a result of this design work was to eliminate coverage checks and units surveys (and to place greater emphasis on the taxation frame) as well as placing relatively more effort on getting accurate responses from the larger units.

A number of criticisms could be made of the assumptions made in the study. However, the main purpose of this paper is to illustrate the technique. The assumptions will be refined in future designs.

2.6. *Sensitivity of results to parameters estimated*

As it will be very difficult to reliably estimate all parameters in the error model, a

Table 2. Cost effective options for resource allocations

Option	RMSE	Cost	Description
1	7.83	330	All mail enumeration, use of taxation data for small units, no coverage enhancement, no units survey, intensive follow-up for high response rate.
2	7.18	530	Personal visit and intensive follow-up for large units, use of taxation data for small units, no coverage enhancement, a units survey and intensive follow-up for high response rates.
3	6.55	780	As for 1 but small mail sample for small units.
4	3.54	1050	Personal visit and intensive follow-up for large units, field interview for small units, no coverage check, no units survey, small sample size.
5	3.23	1550	As for 4 but with medium sample size.
6	3.13	2050	As for 4 but with large sample size.
7	3.05	2580	As for 2 but with small units field enumerated and large sample size.

study was done of the effect on relative efficiencies of marginal changes in some of the more difficult to measure parameters. In particular, options were recalculated with the bias parameters adjusted upwards by 20%. The "optimum" options were unchanged.

Another check of the sensitivity of the results involved the alteration of the non-response bias levels used. The bias levels used in the study were relatively large and negative, based on an assumption of a tendency by processing staff to treat non-responses as nils and hence to underestimate in imputation. The options were recalculated assuming a negligible imputation bias. Not surprisingly this resulted in reduced RMSE for lower follow-up options and for a given field enumeration cost. Nevertheless, option 4 remained an efficient approach.

Even where the designer of a collection has relatively little knowledge of the true value of some parameters used in the model, setting up an optimal allocation using a total survey design approach can indicate the range of values the parameters would need to take for any proposed allocation of resources to be sensible. The likelihood of the parameters falling in this range can then be assessed.

3. Problems with an Optimal Allocation of Resources Approach

The aim of any study to provide an optimal or near optimal allocation is to save resources while keeping total error low. Unfortunately the data necessary for such a resource allocation can itself be expensive to collect. In a situation of tight resources it is very easy to cut evaluation studies. The example given above indicates this may not be cost effective, particularly in more complex surveys known to be subject to substantial non-sampling error from a number of sources. Clearly the cost effectiveness of evaluation studies aimed at a total survey design approach will be greatest where a variety of realistic options for error reduction exist and where the survey is repeated so that the cost can be written off over a number of applications.

A further problem for most collections is the problem of multiple objectives. An allocation that is optimal with respect to one variable may be far from optimal with respect to another. In the CIS example, optimisation was for one variable, total sales, at the national level. In fact the survey provides statistics on numerous variables at State and Industry levels. This problem of multivariate objectives is encountered already in survey design. It has generally been tackled by deriving allocations under a number of objectives and determining a compromise allocation which is then checked for performance against the various survey objectives. This would also be possible under a total survey design approach although the cost of the analysis will clearly increase.

Fortunately, for most economic collections the variables of interest are generally highly correlated and the main question in terms of resource allocation is the priority to be given to finer level estimates, e.g., State and Industry compared to national totals. At this more detailed level, sampling errors become dominant and sample size becomes relatively more important. This is the main reason why the managers of the 1988–89 CIS decided to use the larger, but less accurate mail collection method in preference to field collection from the small units.

Another aspect of the multiple objective problem is that different types of error may affect users differently, depending on the use to which the estimates are being put. A prime example is that of coverage bias. If users are interested in estimates of level, coverage bias is an important contributor to total error. However, if users are interested in change over time, a constant coverage bias does not contribute to error at all.

Groves (1989) and Fellegi and Sunter (1974) (reported in Groves) describe and elaborate some other problems faced in a total survey design approach.

i. There are practical constraints (e.g., limits on the availability of resources with particular skills) which restrict the number of feasible alternative designs.
ii. The error reduction and cost functions are often complex and non-continuous, making modelling difficult if not impossible. We have attempted to overcome this by the empirically based approach described in this paper.
iii. Survey errors from different sources may interact in unknown ways. As an example, frame improvement operations could also affect the efficiency of the sample design, the cost and ease of data collection, and the efficiency of the estimation process. It is difficult to take account of interactions of this type.
iv. An optimum design for an individual survey may not necessarily be optimum across several surveys. The example given by Fellegi and Sunter (1974), a national field force of interviewers, may not be cost efficient for a single survey, relative to using a centralised telephone interviewing installation, but that may change when considered over several surveys. The analagous situation is whether to undertake coverage checks or not. For CIS it was clearly not optimal to undertake coverage checks, but if you could amortise the costs of coverage checks over several collections using the business register, a different result may arise.

4. Conclusion

Despite the many potential problems with determining an optimal resource allocation for a statistical collection, the major problem still remains one of determining the relationship between resources used on different aspects of the collection and the total error resulting. Because this relationship is so clearly specified for sampling error through the parameter sample size, statisticians have been willing and able to determine optimal allocation of enumeration resources to minimise sampling error in quite complex surveys, with complex enumeration cost structures.

However, this is only tinkering at the edge given that non-sampling errors are as important as they are in most surveys. If we are to progress further on appropriate allocation of resources to minimise total error, effort must be put into measuring and testing relationships between error and resources for the major sources of error. In the absence of suitable models, it is likely that empirical methods provide the only viable method of determining an optimum design. This study clearly demonstrates the value of the Total Survey Design approach. It has been applied to the design of a Construction Industry Survey which involves decisions such as whether to supplement the business register with other coverage sources, whether to clean up the register for defunct and out-of-scope units before enumeration, whether to use a dual frame approach, the extent of non-response follow-up, whether to use field enumeration rather than mail enumeration, and the extent to which taxation sources can be used rather than direct data collection.

5. References

Anderson, R., Kasper, J., Frankel, M.R., and Associates (1979). Total Survey Error. San Francisco: Jossey-Bass.

Fellegi, I.O. and Sunter, A.B. (1974). Balance Between Different Sources of Survey Errors – Some Canadian Experiences. Sankhya, Ser. C, 36, 119–142.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: John Wiley.

Hartley, H.O. (1962). Multiple Frame Surveys. Proceedings of the Social Statistics Section, American Statistical Association, 203–206.

Received March 1990
Revised June 1992