DESIGN EFFECTS IN RETROSPECTIVE MORTALITY SURVEYS

Monroe G. Sirken and Patricia N. Royston
National Center for Health Statistics

1. INTRODUCTION

Recently, we conducted an experiment to measure some error effects of selected design factors in retrospective household sample surveys on dual system estimators of mortality. These estimators have been used to estimate the number of deaths [6] and the completeness of death registration [4]. The dual system mortality estimator implies two operational stages:

- Conducting a survey in which household respondents report retrospectively deaths that occurred during a prior calendar period.
- Matching the deaths enumerated in the survey against a file of registered deaths.

Our experiment was designed to measure the error effects of survey design factors on the number of deaths enumerated in the survey that were matched with their death certificates.

The experiment was based on a sample of deaths registered in North Carolina. First, an address frame of the households that would be eligible to report these deaths in a retrospective survey was compiled. Second, retrospective surveys were conducted on subsamples of addresses. Finally, the reported survey deaths were matched against the file of registered deaths in North Carolina. The objectives and procedures of each stage of the experiment are summarized in Exhibit 1.

2. DESIGN FACTORS AND OPTIONS

Design factors represent the manipulatable features of the survey design. There are many design factors, including the sample design, estimator, data collection method, questionnaire design, etc. Each factor usually presents several design options, including the null option, and each option has its cost and error effects. In this frame of reference, the survey design problem may be stated as follows: to select an option for every design factor such that the set of options selected is best in the sense that it produces a smaller mean square error for fixed costs than would be produced by any other option set.

The first and second columns of Exhibit 2 list and define respectively the five design factors that were investigated in the survey experiment. The particular options that were investigated for each factor are listed in the third column. In view of our particular interests in the counting rule strategy [5], the counting rule and the counting rule weight were the principal design factors that we investigated in the experiment. In this connection, a recent paper by Nathan [3] is noteworthy. These factors are probably less familiar to the reader than the other listed design factors. Hence, the information presented for counting rules and counting rule weight in Exhibit 2 is amplified below.

Counting Rules

In retrospective mortality surveys, counting rules specify the conditions that make decedents eligible to be enumerated at households. Five different counting rule options are listed in Exhibit 2. For instance, option 1.1 links decedents to their last places of residence. In compliance with this rule, the household respondent would be asked: "Did anyone die during the [reference period] while he was living here?" Option 1.1 is a conventional rule since it has the property of making a decedent eligible to be enumerated at one and only one household. The remaining options 1.2-1.5 listed in Exhibit 2 are multiplicity rules since they do not uniquely link decedents to one and only one household. In compliance with each of these options, the appropriate question becomes: "Does anyone living here have a [specified relative] who died during the [reference period]?"

It is noteworthy that the counting rule could be based on any subset of the five options listed in Exhibit 2 including options taken 1, 2, 3, 4 and 5 at a time. Summing these subsets, we obtain a total of 31 different counting rule possibilities.

Options 1.2-1.5 were used to link middle aged decedents to the residences of surviving relatives. Somewhat different options were tested for decedents in the youngest and oldest age groups. Children under 17 years were linked to the residences of their mother, and their maternal grandparents, aunts, and uncles. Decedents 85 years and older were linked to residences of their spouse, children and siblings.

The experiment also investigated counting rules that specify the proximity of the decedents' last residence to the residences of their surviving relatives. There were three options: the decedent and his surviving relative lived in the same County, in the same State, or in the U.S. To apply these geographic rules requires collecting information on the location of last residence for every decedent enumerated in the survey.

Counting Rule Weights

Survey estimators based on multiplicity counting rules adjust for the multiple chances of a decedent being enumerated by appropriately weighting each household that reports him in the survey. These weights are called counting rule weights. The survey estimator is unbiased if the sum of weights assigned to the households eligible to report the same decedent is equal to one.

Exhibit 1.

Stages in the Design of the Survey Experiment: Objectives, Procedures and Products

Stage	Objective	Procedures	Types of Errors Measured
1	To compile an address frame of households eligible to report deaths	Selected a sample of death records from the files of registered deaths; queried death record informants to obtain the addresses where the deaths would be enumerable in a survey.	Sampling errors Counting rule bias
2	To conduct retrospective surveys based on households selected from the address frame	Conducted mail and personal interview surveys; conducted reinterviews with adults who did not originally respond for themselves.	Nonresponse bias Response bias
3	To match the deaths enumerated in the survey to their registered death certificates	Matched the deaths enumerated in the survey against the complete file of registered deaths using the Health Department's standard matching procedures.	Matching bias

Exhibit 2.

Survey Design Factors: Definitions and Options

Design Factors	Definitions	Options
1. Counting Rule	Defines the households where the deaths are eligible to be enumerated in the survey	1.1 Last residence of decedent 1.2 Residence of surviving spouse 1.3 Residence of surviving parents 1.4 Residences of surviving children 1.5 Residences of surviving siblings
2. Counting Rule Weight	A weight assigned to every household for every death it is eligible to report	2.1 Inverse of the number of households eligible to report the death 2.2 Fraction of the eligible relatives residing in the household
3. Data Collection Method	The method of querying the households in the survey	3.1 Mail survey 3.2 Personal interview survey
4. Respondent Rule	Defines the persons that are eligible to respond in the survey	4.1 Related adults are eligible to respond for one other 4.2 Adults are eligible to respond for themselves
5. Length of Reference Period	The elapsed time between the date of the person's death and date the household eligible to report the death is surveyed	5.1 Within 6 months 5.2 Within 9 months 5.3 Within 12 months

(See the appendix for the formulation of the estimator and the derivation of the unbiasedness conditions.) The information needed to calculate the counting rule weights is obtained in the survey from the household that reports the decedent.

We refer to the counting rule options 2.1 and 2.2 that are listed in Exhibit 2 as the unit and the element counting rule weights respectively. Although both options are shown in the appendix to satisfy the unbiasedness conditions, the questions asked in the survey to obtain the information needed to calculate them is somewhat different. They have one question in common, namely

"How many [specified relatives] does the decedent have?"

The unit weight requires two additional questions:

"What are their names?"

"Which of them are living together?"

The element weight requires only one question in addition to the common question:

"How many of the [specified relatives] are living in this household?"

Our experience has been that it takes less time and effort to collect the information for the element than for the unit weight.

3. ERROR EFFECTS OF DESIGN FACTORS

For every design option listed in Exhibit 2, the experiment investigated their separate and combined effects on the five types of errors listed and defined in Exhibit 3. In addition to sampling errors, four types of bias errors were investigated in the experiment. (The stage of the survey experiment that measured these errors is shown in the right hand column of Exhibit 1.) It is noteworthy that many types of errors were not measured at all by the experiment. Excluded, for example, were nonsampling variance and bias errors due to erroneously enumerating or matching deaths.

Design factors are selective in their error effects. Exhibit 4 identifies the types of errors that are affected by each of the design factors. Thus, the counting rule is the only factor that affects all five types of errors. Counting rule weights affects four types of errors. Each of the other design factors affect three types of errors, but only two of them affect the same types of errors.

The error effects of the design factors are not independent. For instance, the effect of a self respondent rule may be quite different when combined with a conventional counting rule than when combined with a counting rule that links deaths to the households of surviving relatives. The findings of the experiment will make it possible to compare the error effects of about 500 different option sets for the five design factors.

4. CONCLUDING REMARKS

We have formulated a strategy for designing efficient data systems, which involves selecting the set of design options that minimizes the sampling and measurement errors. Somewhat similar strategies have been proposed by Dalenius [1] and Nathan [2]. To implement the strategy proposed here requires a matrix of information on the cost and sampling and nonsampling error effects of design factors and their options. We have proposed a partial structure of the design matrix for retrospective mortality surveys and have described a survey experiment that was conducted to measure some of the sampling and nonsampling error effects of selected options for a few design factors. We have implied some of the cost effects in terms of the supplementary information required by some of the options. The remaining structure of this design matrix needs to be defined, and more experiments need to be conducted to compile information for the additional option sets.

Although the design matrix for retrospective mortality surveys may be applicable to some other types of retrospective surveys, it is not likely to be applicable to most data systems. It is timely to begin to formulate the design matrices for different types of data systems, and to construct these matrices from information that is already available or by designing the necessary experiments. If nature is kind, we will discover some generalities in the cost and error structures so that by means of a relatively small number of design matrices of reasonable size we will be able to handle many different types of data systems.

ACKNOWLEDGEMENTS

The survey experiment is being conducted for the National Center for Health Statistics by the Research Triangle Institute.

References

[1] Dalenius, Tore, "Designing Descriptive Sample Surveys" in New Developments in Survey Sampling, Wiley-Interscience 1969, pp. 390-416.

[2] Nathan, Gad, "Utilization of Information on Sampling and Nonsampling Errors for Survey Design - Experiences in Israel. Bulletin of the International Statistical Institute, Vienna, 1973.

[3] Nathan, Gad. "Evaluation of Different Counting Rules and Weighting Procedures for Surveys with Multiplicity." Paper presented at the 1976 Meetings.

[4] Sirken, Monroe G., "Design of Household Sample Surveys to Test Death Registration Completeness," Demography, August 1973, Vol. 10, No. 3, pp. 469-478.

Exhibit 3.

Types of Errors and Their Definitions

Type of Error	Definitions
A. Sampling errors	Errors resulting from the sample selection procedure
B. Counting rule bias	Deaths not enumerated because they are not linked to any household by the counting rule
C. Nonresponse bias	Deaths not enumerated due to household nonresponse in the survey
D. Response bias	Deaths not enumerated because the responding households fail to report eligible deaths
E. Matching bias	Survey deaths that fail to match with their death certificates

Exhibit 4.

Error Effects of Design Factors

Design Factor	Type of Error				
	A Sampling error	B Counting rule bias	C Nonresponse bias	D Response bias	E Matching bias
1. Counting rule					
2. Counting rule weight					
3. Data collection method					
4. Respondent rule					
5. Length of the reference period					

[5] Sirken, Monroe G., "The Counting Rule Strategy in Sample Surveys," American Statistical Association Proceedings of the Social Statistics Section, 1974, pp. 119-123.

[6] Sirken, Monroe G., "Dual System Estimators Based on Multiplicity Surveys," Developments in Dual System Estimation of Population Size and Growth, The University of Alberta Press, 1976, in print.

APPENDIX. DERIVATION OF THE ELEMENT AND THE UNIT COUNTING RULE WEIGHTS

Let

D_α ($\alpha = 1, \ldots, N$) represent the N deaths in the population Π

H_i ($i = 1, \ldots, M$) represent the M households in the sampling frame

A survey is conducted based on a sample of H_{i_j} ($j = 1, \ldots, m$) housing units to estimate N. Deaths are enumerated at the m sample households in compliance with a counting rule adopted in the survey.

The counting rule link between a decedent and the household eligible to report him is represented by the indicator variable

$$\delta_{a,i} = \begin{cases} 1 \text{ if a relative of } D_\alpha \text{ resides at } H_i \\ 0 \text{ otherwise.} \end{cases}$$

Thus,

$$S_\alpha = \sum_{i=1}^{M} \delta_{ai} = \text{number of households containing relatives of } D_\alpha.$$

The linear estimate of N,

$$\hat{N} = \frac{M}{m} \sum_{\alpha=1}^{N} \sum_{j=1}^{m} W_{\alpha,i_j} \delta_{\alpha i_j}$$

is unbiased if and only if

$$\sum_{i=1}^{M} \delta_{\alpha i} W_{\alpha i} = 1 \ (\alpha = 1, \ldots, N).$$

The $W_{\alpha i}$'s are the counting rule weights.

If the weights are assigned such that $W_{\alpha i} = W_\alpha$, the unbiasedness condition becomes

$$W_\alpha = \frac{1}{\sum_{i=1}^{M} \delta_{\alpha i}} = \frac{1}{S_\alpha}.$$

The unbiasedness condition is also satisfied if

$$W_{\alpha,i} = \frac{R_{\alpha i}}{R_\alpha}$$

where

$R_{\alpha i}$ = number of D_α's relatives residing in H_i

$R_\alpha = \sum_{i=1}^{M} R_{\alpha i}$ = number of D_α's relatives.

The W_α's and $W_{\alpha i}$'s are referred to as the unit and the element counting rule weights respectively.