DISCUSSION OF MEDICAL PROVIDER SURVEYS OF NEUROLOGICAL CONDITIONS

Monroe G. Sirken, National Center for Health Statistics

1. INTRODUCTION

The papers presented at this session confirm the well established fact that designing sample surveys to compile national statistics for relatively rare diseases presents difficult methodological problems. Medical provider surveys represent one of the promising survey design strategies [5] that have been developed in recent years to provide national statistics on rare health conditions. The essential features of the medical provider surveys are: (1) The enumeration units are medical providers such as hospitals, physicians in private practice, clinics, etc., and the population elements are the patients with the condition. (2) The patients are enumerated at a sample of medical providers. The providers report about the patients they have treated and identify other potential sources of information for them. (3) Follow-up surveys are conducted, as necessary, with the other sources to supplement or confirm the information about the patients reported by the medical providers in the original sample.

Recently, the National Institute of Neurological and Communicative Disorders and Stroke established a survey program to compile national statistics on three types of neurological conditions: intracranial neoplasms (IN), head and spinal cord injuries (HSCI), and multiple sclerosis (MS). In view of the difficult methodological problems, elaborate pilot studies were undertaken for each type of neurological condition. These studies, described by the three prior papers [1] [5] [7], were designed to investigate the effect of alternative survey strategies for controlling sampling and nonsampling errors and costs in the national survey. All three studies indicate that neurological statistics based on medical provider surveys are subject to large measurement errors as well as large sampling errors. It is the combination of large errors of both types, in fact, that makes these studies particularly difficult to design, since they require methods that will effectively and simultaneously control both error types.

The underlying survey factors responsible for the large errors are essentially the same for the three neurological conditions. Thus, the rareness of each condition makes for large sampling errors since most medical providers are not linked to any persons with the neurological condition. The fact that each condition is pathological as well as rare, I believe, makes for large nonresponse and response errors since it makes the medical providers even often reluctant to divulge the names of their patients for follow-up surveys. Diagnostic problems also contribute to errors of measurement for these conditions. For an absorbing condition, such as MS or IN, there are diagnostic problems associated with ascertaining the presence of the condition and its date of onset. For an nonabsorbing condition, such as minor HSCI, there are additional diagnostic problems associated with determining the absence of the condition and the date of the patients recovery.

The course of medical treatment and care provided to a patient with a serious neurological condition usually involves a series of transactions with a network of medical providers. Thus, the same patient can be enumerated more than once in a medical provider survey. Fortunately, there are a number of ways of coping with this potential source of coverage bias. As a matter of fact, the multiplicity of options available to cope with this problem becomes in itself as important survey design strategy in controlling both the large sampling and measurement errors. Let us consider this problem particularly from the viewpoint of its survey design implications.

2. NETWORK ESTIMATORS

An interesting estimation problem arises in medical provider sample surveys because, as I noted previously, a person with a serious medical condition, such as IN, HSCI, or MS, is usually under concurrent treatment by a network of several medical providers. For example, the network of medical providers to an IN patient will often include a general practitioner, a neurologist, a neurosurgeon as well as one or more hospitals. The basic problem is to appropriately count and weight the persons reported by the medical providers in the sample so that the sample estimator, is unbiased or nearly so.

What appears, at first glance, to be an estimation problem in medical provider surveys can often be converted into a strategy for controlling sampling and nonsampling errors and thereby improving the survey design. The proposed design strategy [6] involves: (1) constructing several alternate estimators, (2) estimating the sample and nonsampling errors and survey costs associated with each estimate and, (3) adopting the estimator that is subject to the smallest mean square error giving due consideration to cost factors.

It is often feasible to construct several unbiased or nearly unbiased estimators for a medical provider survey by exercising the available survey design options. Three design factors are particularly relevant: (1) sampling frames, (2) counting rules, and (3) counting rule weights. The first of these factors defines the numbers and kinds of medical providers that are eligible to serve as enumeration units in the survey. The second factor determines the number and kinds of medical providers that are eligible to report each patient in the population with the neurological condition. The third factor specifies how the enumerated patients are weighted to adjust for the number and kinds of medical providers that are eligible to report them consistent with the sampling frame and counting rule options adopted in the survey.

In the following remarks, I will briefly describe and illustrate the use of each of these design factors in constructing network estimators for sample surveys of medical providers. Within this context, I will comment on the three prior papers presented at this session. But first a few brief remarks about the general role of design factors in designing surveys.

3. DESIGN FACTORS IN MEDICAL PROVIDER SURVEYS

Design factors are the survey features over which the survey statistician exercises control in designing the survey. There are many design factors in the typical survey including the sample selection procedures, estimation methods, data collection methods, questionnaire design, etc. Each factor usually presents several alternative design options and each design option has two kinds of design effects - cost effects and error effects. The error effects include sampling as well as measurement errors. The survey design problem may be stated as follows: To select an option for every design factor such that the selected set of options for all design factors is best in the sense that it produces a smaller mean square error for fixed costs than would be produced by any other possible set of options.

We are now ready to discuss the three design factors that are particularly relevant in designing medical provider sample surveys for neurological conditions. Limiting the discussion to these factors does not necessarily imply, however, that the other factors have lesser importance in designing the survey.

Sampling Frames

A sampling frame is a listing of the eligible enumeration units in the survey. In medical provider surveys, there are several potential sampling frames - one for each type of medical provider. Consequently, there are several alternative options with respect to the number and types of medical provider lists that are selected as the sample frames in the survey. For example, if there are k different lists, the total number of options is equal to the sum of the number of combinations of k frames taken 1 at a time, 2 at a time, ..., k at a time. To select the best option, would involve comparing the error and cost effects of the alternative options.

Key factors in deciding whether the set of frames selected for a survey should contain only one or more than one medical provider listing is the extent to which patients are linked to medical providers that are listed in only one or in more than one frame. Woolsey and Simmons [7], for example, have decided that the IN national survey will be based solely on a hospital frame because the pilot study indicated that surveys of other types of medical providers add very few cases that are not reported in the hospital survey. On the other hand, Kalsbeek [4] and Asmann et al [1] will most likely opt for national surveys of HSCI and of MS respectively that are based on multiple frames of medical providers and possibly other sources of information as well.

It is noteworthy that surveys of the HSCI and MS type, that are based on two or more mutually exclusive frames of medical providers, are not stratified surveys in the conventional sense because some or possibly all patients in the population may be linked to more than one frame. Hartley [3] refers to surveys of this type as multiple frame surveys.

Counting Rules

Counting rules in medical provider surveys specify the conditions that make medical providers eligible to report patients in the survey. The basic condition of eligibility is whether or not the patient was treated by the medical provider. There are, however, many variants of this condition depending on its inclusiveness. The most inclusive form of the condition makes all medical providers who treated the patient eligible to report him. This is the rule that appears to have been adopted in the neurological surveys for each of the neurological conditions IN, HSCI, and MS. Less inclusive versions of the condition, would restrict eligibility to subsets of providers, such as those practicing in specified geographic areas or certified in specified medical specialties. Finally, the most restrictive condition would seek to uniquely link each patient to only one medical provider. This might be done by either restricting eligibility to the first provider who treated the patient for the condition, or to the provider that has the major responsibility for treating the patient.

It is noteworthy that a survey is not necessarily limited to adopting one kind of counting rule. For example, one rule could be used to produce one type of statistics and a different rule to produce other types of statistics. In multiframe surveys of HSCI and MS type, different counting rules could be adopted for linking patients to medical sources that are listed in different frames.

Relatively little is currently known about the design effects of counting rules in medical provider surveys. Generally, but not necessarily consistently, sampling errors appear to increase as the conditions of the rules become more restrictive, but survey costs virtually always decrease with the restrictiveness of the rule. Even less is known about the measurement error effects of counting rules in medical provider surveys. However, counting rules that permit the same person to be linked to more than one enumeration unit offer a strategy for reducing measurement errors when the survey conditions make it difficult to implement a rule which uniquely links each person to only one enumeration unit.

Counting Rule Weights

Several kinds of unbiased estimators have been proposed [2] for sample surveys that are based on counting rules that permit several enumeration units to report the same person. Each

of these estimates uses of system of weights to adjust for differences among the patients in their probabilities of being enumerated in these surveys. Some of these estimators also require matching the enumerated patients to eliminate duplicate reports. The estimator proposed by Asmann et al [1], for example, eliminates duplicate reports but it is a biased estimator nonethe-less since it does not adjust for differences in the patients probabilities of being enumerated.

The network estimator is one of the estimators that counts a patient as many times as he is reported by different medical providers in the survey. The estimator adjusts for the different probabilities of enumerating patients by assigning a counting rule weight to every patient everytime he is enumerated. It assigns a weight of one to those patients that are uniquely linked to a single provider and to patients linked to multiple units, it assigns a non-zero weight everytime they are enumerated. The network estimator is unbiased if the sum of the weights assigned to each of the multi-linked patients is equal to one.

The unbiased network estimator offers many options with respect to the kinds of weights that are assigned to multi-linked patients. For example, Woolsey and Simmons [7] propose to weight every IN patients enumerated in the national hospital sample survey by the inverse of the number of hospitals that the patient is linked to. Thus, in effect, they will be weighting a patient by a uniform weight irrespective of the hospital where he is enumerated. Another possibility would be to assign the person a different weight at each of his eligible hospitals. For example, they might elect to assign weights to a patient on the basis of the relative number of IN discharges from each hospital eligible to report him.

Counting rule weights have design effects on the sampling and nonsampling errors of the survey estimates. Thus, from the viewpoint of minimizing survey errors, assigning sets of uniform weights to IN patients may be a less effective strategy than assigning sets of variable weights. The trick is to determine the best system of weights and to obtain the data needed to calculate the weights. Counting rule weights have design effects on the survey costs, and the cost effects may not be negligible because the information needed to calculate the weights is often not available prior to the survey and hence must be collected as ancillary information in the survey itself.

4. CONCLUDING REMARKS

Except for mortality statistics, we lack national statistics for most rare and relatively rare diseases. Though morbidity and related statistics for some of these diseases are available for restricted subpopulations, it is doubtful that they would be satisfactory for evaluating and planning a national disease oriented research program or a national health delivery program. Precision making at the national level requires estimates of the errors of the national statistics as well as the national statistics themselves.

The expense and difficulty of establishing national data systems based on complete registration of all patients or on complete enumeration of all medical providers has probably been a major deterrent to compiling national statistics on rare diseases. Surveys based on samples of medical providers, of the type discussed at this session, represents another approach to the problem. A data system based on a sample of medical providers has obvious cost advantages over a system that is based on the complete enumerations of all medical providers. However, this fact alone does not settle the matter. We have seen that sample surveys of medical providers present a number of technical problems and that the disease statistics are subject to large sampling and measurement errors. Some solutions to these problems have been proposed at this session, but much work remains to be done.

The National Institute of Neurological and Communicative Disorders and Stroke deserves credit for initiating and sponsoring a survey program on neurological diseases which is giving strong emphasis and support to pilot projects that are designed to test and develop the methodology for national sample surveys of medical providers. The methodological orientation of this program improves immeasurably the chances of collecting viable national statistics on neurological diseases, and it also improves the prospects for developing sample survey methods for collecting national statistics for other types of rare diseases.

References

[1] Asmann, Anthony S., Sherman, Lorna R. and Wilson, Lucy B., "Experiences and Problems in Developing a Sampling Methodology for a Survey of Patients with Multiple Sclerosis and Related Disorders." Paper presented at the annual meetings of the American Statistical Association, 1975.

[2] Birnbaum, Z.W., and Sirken, Monroe G., "Design of Sample Surveys to Testimate the Prevalence of Rare Diseases: Three Unbiased Estimates," Vital and Health Statistics, PHS Publication No. 1000, Series 2, No. 11, National Center for Health Statistics, GPO, 1965, pp. 1-8.

[3] Hartley, H.O., "Multiple Frame Methodology and Selected Applications," unpublished manuscript.

[4] Kalsbeek, William D., "Results From a Pilot Study of Several Methodological Problems Associated with Casefinding Surveys." Paper presented at the annual meetings of the American Statistical Association, 1975.

[5] Sirken, Monroe G., "Survey Strategies for Estimating Rare Health Attributes," Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (1970), pp. 134-141.

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

[7] Woolsey, Theodore D. and Simmons, Walt, "The Use of Medical Records in a Sample Survey of Brain Tumors." Paper presented at the annual meetings of the American Statistical Association, 1975.