

Population Surveys of Health Services Utilization Linked to Health Care Provider Surveys

Monroe G. Sirken
National Center for Health Statistics

Abstract

Whenever health care utilization statistics are needed for small domains of encounters between populations and health care providers and/or the variables of interest are diagnoses, medical test results, charges for services, etc. the linked population/provider survey deserves serious consideration attention as a design alternative to the stand-alone population survey. In the stand-alone population survey, household respondents are the only information sources, and the variables of interest relate primarily, but not necessarily exclusively, to the characteristics of the health care users. In the linked population/provider survey, information is obtained from household respondents and from their health care providers. Household respondents report their encounters and identify their health care providers, and usually report the same variables for their encounters as they would in the stand-alone population survey; providers report diagnoses, charges, etc. of their encounters with households in the population survey and/or samples of their encounters with all households. The paper compares the relative success of three proposed sample designs of the linked population/provider survey in addressing the limitations of the stand-alone population survey.

1. Introduction

Health care utilization statistics on the volume of transactions between users and providers of health care services are often collected in the stand-alone population sample survey. In the stand-alone survey, household respondents are the sole sources of information and they report selected variables of interest about their encounters with health care providers. The National Health Interview Survey of the National Center for Health Statistics is a well-known example of a stand-alone population survey of health services utilization ().

Stand-alone population surveys of health services utilization have several desirable features. They produce quality utilization statistics by the characteristics of health care users (viz. socio-economic and demographic status, etc) that can be aggregated at several usage levels including

transactions, persons and households. Though stand-alone population surveys excel in meeting these kinds of substantive data needs, they are often marginal in meeting other data needs. They are relatively poor sources of statistics on the services and charges of health care providers, and low transaction sample yields often present estimation problems especially for small and elusive populations such as those likeliest to underutilize health care services.

This paper compares the utility of stand-alone population sample surveys and the linked population/provider surveys of health services utilization where utility refers to the survey's capability to meet the data needs of health services utilization within prescribed limits of sampling and measurement errors. The linked survey seeks to improve the utility of the stand-alone population survey by collecting supplementary information about the encounters of the health care providers that have encounters with households in the population survey.

Section 2 describes the major features of the linked population/provider survey and proposes three designs for sampling encounters of health care providers. Section 3 lists the basic data needs for health services utilization statistics, and discusses the limited utility of the stand-alone population survey in meeting the basic data needs. Section 4 compares the variables of interest about the encounters that household respondents report in the stand-alone population survey and the linked population/provider survey. Section 5 compares the potential of the three proposed designs for sampling encounters of health care providers to improve the utility of stand-alone population surveys of health care utilization.

2. Design of Linked Population/Provider Surveys of Health Services Utilization

The linked population/provider survey is a two-phase sample survey. In phase 1, household respondents in the population survey report the variables of interest for their transactions with health care providers and identify their health care providers. In phase 2, a follow-on survey is

conducted with health care providers that have encounters with households in the population survey, and the providers report the variables of interest about their health care services for specified samples of their encounters with households that they may or may not have had encounters in the population survey.

Three proposed sample designs for selecting encounters of the health care providers of households in the population survey are shown in Exhibit 1.

Design A is equivalent to the well known procedure of matching the encounters reported in the population survey the records of their health care providers. For example, the transactions of households in the National Medical Expenditure Panel Survey (MEPS) are matched with the records of their health care providers to verify and supplement the medical expenditures reported by MEPS households (). Designs B is equivalent to a two-stage population sample survey design in which households are first stage selection units and the encounters of providers that have encounters with households are the second stage selection. Design C is a hybrid of Designs A and B. The unbiased estimator and sampling variance of the linked population/provider survey based on Design B is available (). Design C is being proposed here for the first time.

3. Needs for Health Care Utilization Statistics

The utility of the population survey of health care utilization ultimately depends on its capability to meet the basic needs for health care statistics within the desired ranges of precision. Exhibit 2 lists four basic data needs for health care utilization statistics.

Judged on the basis of the four Data Needs shown in Exhibit 2, stand-alone population surveys of health care utilization have important but limited utility. Household respondents are preferred sources of information when the variable of interest relate to the characteristics of health care users and relatively poor information sources when the variables of interest relate to the services of health care providers. Hence, from the viewpoint of response errors, stand-alone population surveys are good sources of health care statistics that meet Data Need 1 and poor sources of statistics of health statistics that meet Data Needs 2, 3, and 4. From the viewpoint of sampling errors, the stand-alone

survey has limited utility in meeting any of the four Data Needs especially for small and elusive populations such as those representing people that underutilize health care services.

4. Information Collected in Stand-alone Population Surveys and Linked Population/Provider Surveys

Before discussing the impact of Designs A, B, and C of the linked population/provider survey on the utility of the stand-alone population survey, it is useful to compare the kinds of information reported by household respondents in these surveys. These comparisons are shown in Exhibit 3.

Household respondents report the attributes of health care users in the stand-alone population survey and in the linked surveys based on Designs A, B, and C. Services of healthcare providers would probably be reported in linked survey based on Design B and D to assist in the matching operations and probably would not be reported in the stand-alone population survey and the linked survey based in Design C. Household respondents identify their health care providers in linked survey Designs A, B and C, and not in the stand-alone population survey. Confidentiality waivers are needed in Linked Designs B and D because encounters of household respondents are matched with their health provider records.

5. Effects of Linked Population/Provider Survey Designs on Utility of Stand-alone Surveys of Health Services Utilization

The utility of stand-alone population surveys in meeting the basic needs for health care utilization statistics is limited by relatively large response errors and sampling. Large response errors are due to incomplete and inaccurate reporting of the services of health care providers by household respondents. Large sampling errors are due to small yields of transactions. Exhibit 4 makes rough comparisons of the relative success of linked population/provider survey designs A, B and C in controlling response and sampling errors in stand-alone population surveys. The assessment assumes that the variables of interest about the services of health care providers are reported by household respondents in the stand-alone population survey and linked population/provider surveys Designs A, B, and C .

Design A substantially improves response errors except for statistics associated with Data Need 1 and does not affect the sampling errors of stand-alone population surveys.

As a matter of fact the linked household/provider survey based on Design A and the stand-alone population survey of equivalent household sample size yield precisely the same transactions and have precisely the same sampling errors.

Design B substantially improves the response errors of statistics of the stand-alone population survey associated with Data Need 2, has little or no effect on other the response errors, but substantially improves sampling errors. The sampling error gains are most likely and predictable when relatively few households have multiple transactions, and ρ , the ratio of the within to the sum of the within and between components of provider variance, is significantly greater than zero. If every household in the universe has a single transaction, for example, the reduction in sampling variance of statistics associated with Data Need 2 is about $(1 - n_p / n_B)\rho$ where n_p is the number of transactions of households in the stand-alone population survey and n_B is the size of the transaction sample in the follow-on survey with health care providers.

Design C substantially improves the response errors as well as the sampling errors of stand-alone population surveys. However, matching transactions reported in the population survey with their medical records exposes Designs A and C to two potentially serious sources of coverage bias. Transactions can be missed due to attrition in the matching process and/or to failure of household respondents to sign confidentiality waivers. In MEPS, for example, percent of the households failed to sign confidentiality waivers and percent of the transactions of households that signed waivers were not matched with their provider records ().

6. Summary and Concluding Remarks

This paper compares the utility of stand-alone population surveys and linked population/provider surveys of equivalent household sample size in meeting the basic data needs for health services utilization statistics with precision, where precision depends on response errors and sampling errors. In the stand-alone population survey, household respondents report the variables of interest about

their encounters with health care providers. In the linked population/provider survey, the information reported by household respondents in the stand-alone survey is supplemented by the health care providers that have encounters with the households in the stand-alone population survey. Health care providers report the variables of interest for specified samples of their transactions with all households. The paper proposes and compares three designs for selecting samples of the encounters of the providers that have encounters with households in the stand-alone population survey, and describes the potential of each proposed design option to improve the utility of the stand-alone population survey of health services utilization.

However, comparing utility of surveys of equivalent household sample size is not the same as and perhaps not as relevant as comparing the utility of surveys of equivalent survey costs. Removing the constraint of equivalent household sample size, offers opportunities to exploit the potential advantages of two-stage sampling in linked survey Designs B and C by optimizing first and second stage sample sizes for fixed total survey costs. On the other hand, the stand-alone population survey and linked survey Design A are single stage sample surveys. Consider briefly the prospects of linked survey Designs B and C respectively improving the utility of the stand-alone population survey and linked survey Design A of equivalent total survey costs.

In the absence of research on the relative costs of stand-alone population surveys and linked surveys of health services utilization, relative efficiencies of these surveys are not discussed this paper. Nevertheless, existing research findings comparing survey precision based on equivalent household sample sizes are encouraging. They indicate that sampling errors are virtually always less for two-stage designs (e.g. linked survey Designs B and C) than one-stage sample designs (e.g. stand-alone population survey and linked survey Design A) if sufficiently large second-stage encounter samples are selected in Designs B and C. Some directions for future research are suggested below.

After conducting this research, it would be feasible to make realistic comparisons of the efficiencies of (1) the two-stage linked survey Design B relative to the single-stage linked population/provider survey, and (2) and the two-stage linked survey design C relative to the single-stage linked survey Design A.

Acknowledgements

The views expressed in this paper are solely those of the author and do not necessarily represent the official views or positions of the National Center for Health Statistics.

Exhibit 1. Three Designs for Sampling Transactions of Health Care Providers
Of Households in the Population Survey

<u>Sample Design</u>	<u>Description</u>
Design A	Encounters of households in the population survey are matched with their health care provider records.
Design B	An encounter samples is independently selected for each provider with sample size proportional to the number of its transaction with households in the population sample survey.
Design C	Two transaction samples (Designs A and B) are selected for each provider.

Exhibit 2. Data Needs of Health Care Utilization Statistics

<u>Data Need</u>	<u>Description</u>
1	Statistics by the characteristics of health care users (viz. socio-economic and demographic status, etc.)
2	Statistics by the services of health care providers (viz. health care test results, diagnoses, charges, etc.).
3	Statistics by the characteristics of health care users and services of health care providers
4	Statistics meeting Data Needs 1, 2, and 3 that can be aggregated at several usage levels (viz. transaction, person, household)

Exhibit 3. Information Reported by Household Respondents in Stand-alone Population Surveys and Linked Population/Provider Surveys

<u>Kinds of information Reported</u>	<u>Stand-alone Survey</u>	<u>Linked Survey</u>		
		A	B	C
Attributes of health care users	Yes	Yes	Yes	Yes
Services of health care providers	Doubtful	Probably	Doubtful	Probably
Identities of health care providers	No	Yes	Yes	Yes
Confidentiality waivers	No	Yes	No	Yes

Exhibit 4. Relative Success of Linked Population/provider Survey Designs A, B and C in Improving the Utility of Stand-alone Population Surveys

<u>Type of Errors</u>	<u>Design A</u>	<u>Design B</u>	<u>Design C</u>
Response Errors			
Data Need 1	some	little	some
Data Need 2	substantial	substantial	substantial
Data Need 3	substantial	little	substantial
Data Need 4	substantial	none	substantial
Sampling Errors	none	substantial	substantial
