NATIONAL NETWORK SURVEYS OF DIABETES

Monroe G. Sirken, Barry I. Graubard, Miriam J. McDaniel
National Center for Health Statistics

INTRODUCTION

Historically, estimating accurately the number of persons with a rare health condition by means of household sample surveys has been difficult. In conventional household surveys, each person with a condition can be reported by one and only one household, namely the household in which he resides. These surveys require large sample sizes in order to produce reliable estimates (i.e., small relative standard errors) of rare population quantities. Also, many conventional surveys tend to have substantial underreporting, especially when the health condition is a sensitive matter. About a decade ago, Sirken [5] developed a way of designing household sample surveys that, in most cases, produces considerably lower sampling variances for estimating rare quantities than conventionally designed surveys. Later work has indicated that these types of surveys may actually reduce certain types of reporting biases that are found in conventional surveys [9]. These sample surveys have been called multiplicity surveys or network surveys in the literature. A network household survey is different from a conventional household survey in that a person can be reported at more than one household. Originally, network surveys were developed to estimate the prevalence of rare diseases in a human population [1]. Since then they have been used to estimate drug use prevalence [7], birth, death and marriage rates [3], and completeness of death registration [9], and are presently being tested to estimate undercoverage in the 1980 census [10].

This paper concerns itself with the problem of measuring the prevalence of diabetes in the United States, using a national household sample survey, the 1976 Health Interview Survey (HIS). A pretest of this survey was conducted in 1975 [8]. We will compare three estimators of diabetes prevalence - two network estimators and a conventional estimator. Each estimator corresponds to a different counting rule and a different set of information items that need to be collected in the household interview. The counting rule specifies the conditions that make persons eligible to be enumerated at households. The counting rules adopted in diabetes surveys specify conditions that link persons with diabetes to households where they are eligible to be enumerated. The conventional estimator of diabetes prevalence for household surveys we will call the de jure estimator. It is based on a rule which links diabetics to themselves so that they are the only ones eligible to report diabetics in the survey. We use consanguine relationships to define the counting rules for the network estimators. The de jure-sibling estimator is based on a counting rule which links diabetics to their siblings as well as themselves, that is, diabetics are eligible to be reported by themselves and also by their siblings wherever they may live. The de jure-children estimator is based on a counting rule which links diabetics to themselves and their parents wherever they may live. The two network estimators require collection of additional information in the household interview which is not required by the conventional estimator.

The estimates of diabetes prevalence and their sampling errors based on a conventional estimator and two network estimators derived from HIS are compared in this report.

COUNTING RULE WEIGHTS

The multiplicity of a diabetic person is defined as the total number of persons in the population that are eligible to report him. For example, the de jure-sibling counting rule specifies that the diabetic person is eligible to be enumerated at his de jure household and at the households of his siblings. The number of persons eligible to report him at his de jure household is equal to the number of his siblings living with him plus one because he can report himself. The number of persons eligible to report him at a household which is not his residence is equal to the number of his siblings living there. The other network counting rule, the de jure-children counting rule, specifies that the diabetic person is eligible to be enumerated at his de jure household and at the households of his children.

Again, the number of persons eligible to report him at his de jure household is equal to the number of children living with him plus one. The number of people eligible to report him at another household is equal to the number of his children living there.

Every diabetic person enumerated in the survey is assigned a counting rule weight. This weight is required by the estimator to adjust for the multiplicity of persons eligible to report him. The counting rule weight assigned to a person is a fraction equal to the number of times the person is eligible to be enumerated in the household in which he was enumerated divided by the multiplicity of the person. Thus, the counting rule weight assigned to a person depends on the particular counting rule adopted in the survey.

The weights for conventional counting rules are always equal to one since the conventional counting rule permits every diabetic person to be enumerated once and only once. Since the network counting rule weights are usually unknown prior to the survey, they are determined on the basis of additional information collected from the sample household where the diabetic persons are enumerated.

THE HEALTH INTERVIEW SURVEY

The information which we will use was obtained from the 1976 Health Interview Survey (HIS). HIS, one of the major components of the

National Health Survey Program, is an annual survey that assesses the health status of the population on the basis of comprehensive interviews that are conducted weekly in a national household sample survey. The 1976 HIS was based on a sample of 41,559 households containing 113,178 persons. Annually, revisions are made in the supplements to the HIS. The diabetes supplement in the 1976 HIS contained a set of questions intended to produce familial aggregation of diabetes in the United States. The set of questions, listed in the Appendix, was completed for each member of a household selected into the sample.

The only information required for the de jure estimator is whether each resident of the sample household is diabetic, that is Q.1 (See Appendix). However, each of the network estimators requires two additional pieces of information. First, information is needed to count the diabetic relatives each household resident is eligible to report for the de jure-sibling estimator (Q.3b) and for the de jure-children estimator (Q.4b, Q.5a and Q.5b). Secondly, information is needed to produce the counting rule weights for each reported diabetic, Q.3a for the de jure-sibling estimator and Q.2 and Q.3a for the de jure-children estimator.

The standard HIS interviewing procedures including the standard respondent rules were applied in the diabetes supplement. According to the standard HIS respondent rules, the preferred household respondents are: (1) the parents or guardians of the children under 17, (2) the person himself, if he is older than 18, and (3) either the parent or the person himself if the person is 17 or 18. In the event that the preferred respondent is not at home when the interviewer visits, the HIS respondent rule specifies that any other related adult in the household is eligible to serve as the proxy respondent for the absent person.

FINDINGS

Table 1 presents three sets of diabetes prevalence rates - the first is based on the de jure estimator, the second is based on the de jure-sibling estimator, and the third is based on the de jure-children estimator. The numerators of the three types of estimators for diabetes rates are based on their respective estimators. The denominator of each set of rates is based on the de jure estimator. Complete demographic information (i.e., age, race, and sex) was collected for each resident of a sample household during the household interview. However, the only

Table 1. Diabetes Prevalence Rates (Per 1000 Population), Their Relative Standard Errors and Design Effects by Type of Estimator, and by Age, Color and Sex

Population Characteristics	Population Size (per 1000)	Rate per 1000 Population			Relative Standard Errors (in percent)			Design Effects	
		De Jure[1] Estimator	De Jure-Sibling Estimator	De Jure-Children Estimator	De Jure Estimator	De Jure-Sibling Estimator	De Jure-Children Estimator	De Jure-Sibling Estimator	De Jure-Children Estimator
All Persons	210,629	19.7	18.9	24.0	2.3	1.5	1.4	2.1	2.8
Under 25 years	91,816	1.9	2.5	1.7	11.2	6.8	11.3	2.7	1.0
25-44 years	53,766	10.6	12.6	16.3	5.5	3.9	4.0	2.0	1.9
45-64 years	43,250	42.0	40.1	54.1	3.3	2.4	1.8	1.9	3.5
65 years and over	21,798	72.6	61.9	76.9	2.8	2.4	1.8	1.4	2.6
White	182,838	19.4	19.0	24.1	2.3	1.6	1.3	2.1	3.2
Under 25 years	77,305	2.0	2.7	1.8	11.6	7.0	11.7	2.7	1.0
25-44 years	47,012	9.8	12.4	15.3	5.6	3.8	4.5	2.2	1.6
45-64 years	38,754	39.5	38.3	52.9	3.3	2.4	1.8	1.9	3.6
65 years and over	19,768	70.5	60.8	75.7	3.1	2.7	1.9	1.4	2.9
Other	27,791	21.7	18.4	23.1	6.3	4.3	5.7	2.2	1.2
Under 25 years	14,511	1.7	1.2	1.3	26.0	25.5	29.3	1.0	0.8
25-44 years	6,754	15.8	14.1	23.1	15.2	11.7	11.3	1.7	1.8
45-64 years	4,496	63.3	55.6	64.3	8.4	7.2	6.0	1.4	2.0
65 years and over	2,030	93.1	72.7	88.4	9.2	8.5	7.5	1.2	1.5
Male	101,619	18.0	--	21.6	3.5	--	2.3	--	2.3
Under 25 years	46,061	1.9	--	1.8	15.9	--	16.1	--	1.0
25-44 years	25,965	8.2	--	14.5	9.3	--	6.4	--	2.1
45-64 years	20,631	42.9	--	53.6	4.4	--	2.5	--	3.1
65 years and over	8,961	71.9	--	69.6	4.9	--	4.0	--	1.5
Female	109,011	21.2	--	26.2	3.1	--	2.0	--	2.4
Under 25 years	45,755	2.0	--	1.6	12.4	--	11.7	--	1.1
25-44 years	27,801	12.8	--	17.9	7.0	--	5.1	--	1.9
45-64 years	22,618	41.2	--	54.6	4.3	--	2.6	--	2.6
65 years and over	12,836	73.1	--	81.9	4.2	--	2.3	--	3.4

[1]These estimates are smaller than the official National Center for Health Statistics estimates of diabetes prevalence rates because the official NCHS estimates included invididuals who were borderline diabetics.

demographic information collected about diabetic parents was sex (since information about diabetic status was collected separately for mother and father), and the only demographic information collected about diabetic siblings was race. Since only full siblings of enumerated persons were eligible to be reported, the race of the enumerated person was the same as his siblings. The above information was the only demographic information that could be collected about diabetic relatives because of limitations imposed on the length of the HIS questionnaire. Nevertheless, by assuming that siblings are in the same broad age group as the person who reports them we were able to obtain estimates of diabetes prevalence rates by age and race using the de jure-sibling estimator. Also, by assuming that ages of the mother and father were 25 and 27 years greater than the age of the child who reported them, respectively, and by assuming that the race of the mother or father was the same race of the child who reported them, we were able to obtain estimates of diabetes prevalence rates by age, race, and sex using the de jure-children estimator.

Another qualification needs to be stated concerning the estimates in Table 1. The de jure estimator produces estimates from HIS data of the number of noninstitutionalized diabetics. However, the de jure-sibling estimates included the diabetics in the institutionalized population who have noninstitutionalized siblings, and the de jure-children estimator includes the diabetics in the institutionalized population who have noninstitutionalized children. Thus, each of our estimators provides estimates of the number of diabetics for somewhat different populations. This problem could have been avoided if the HIS questionnaire had specified that only noninstitutionalized diabetic relatives were eligible to be reported. Since there are no estimates of the number of diabetics in institutions who have noninstitutionalized children or noninstitutionalized siblings, we cannot correct the two network estimators. Since a much greater proportion of people 65 and over than younger people live in institutions, we examined estimates produced by NCHS [4] of the number of diabetics 65 and over living in nursing homes in 1974. Since not all of these persons have noninstitutionalized siblings or parents these estimates would provide an upper bound on this source of bias for the network estimates. NCHS estimates that 101,498 diabetics 65 and over resided in nursing homes. Since this number represents only about 7 percent of the number of diabetics 65 and over estimated from the de jure estimator, it is not large enough to affect the analysis that follows.

The estimated diabetes prevalence rate for the national population is 19.7 per 1,000 based on the de jure estimator or about 4 percent more than the estimated rate (18.9 per 1,000) based on the de jure-sibling estimator. The difference is within the sampling error range. However, the diabetes prevalence rate for people over 65 years of age based on the de jure-sibling estimator is significantly lower than the rate based on the de jure estimator, indicating possibly that there was underreporting of diabetic siblings. None of the differences for the other estimates based on the de jure estimator and the de jure-sibling estimator is statistically different.

The estimated diabetes prevalence rate based on the de jure-children estimator is 24 per 1,000. It is significantly larger than the comparable estimate based on the de jure estimator. We suspect that the estimate obtained from the de jure estimator is lower than the estimate obtained from the de jure-children estimator because a large number of diabetics who fail to report themselves may be reported by their children. This view is supported by the findings of a record check study [2] in which about 20 percent of diabetics identified from the medical records were not reported by their households in the survey.

From Table 1, one can see that the relative standard errors of the de jure-sibling estimator are consistently smaller than those of the de jure estimator. The relative standard errors of the de jure-children estimator are larger than the de jure estimator in only four out of the 25 population domains in Table 1. The four domains all refer to the under-25 age group. A possible explanation for these exceptions may be the way that the ages of diabetic parents were inferred in the survey.

The last two columns of Table 1 give the design effects for the de jure-sibling estimator and the de jure-children estimator. The design effect of an estimator is defined as the ratio of the rel-variance of the de jure estimator to the rel-variance of the network estimator. Estimates of rel-variance were obtained using the balanced half-sample replication technique. The design effect indicates how much larger a sample of households would be required by the de jure estimator to obtain the precision of the network estimator. Overall, about twice as large a sample would be required for the de jure estimator to achieve the precision of the de jure-sibling estimator.

NONRESPONSE

Since only about three percent of the population has diabetes even relatively small nonresponse rates could cause large biases in diabetes prevalence rates. About 96 percent of the households had completed interviews in HIS. Even though a four percent nonresponse rate is considered low for a household survey, it is not to be ignored in a survey estimating a rare condition such as diabetes. The de jure estimator as well as the two network estimators are subject to this source of bias.

The de jure estimator and the network estimators are not based on the same items of information and hence they are not subject to the same item nonresponse rates. The only item of information needed to estimate diabetes prevalence using the de jure estimator is Q.1, and the nonresponse to this item was negligible in the survey. The network estimators, on the other hand, are based on more items of information than Q.1. The unweighted nonresponse rates for these

other items are presented in Table 2. A hot deck procedure was used to impute for the missing items in deriving the network estimates presented in Table 2.

Q.3b identifies diabetics that are eligible to be counted by the de jure-sibling counting rule. The item nonresponse rate to Q.3b is 2.7 percent; it 4.3 percent for respondents 65 years and over. The bias incurred by such large item nonresponse rates could be great if the siblings involved tended to be diabetics. This does not seem likely, however. For instance, only 3.9 percent of the enumerated diabetics did not know if their siblings were diabetic, which is about the same as the percentage of enumerated nondiabetics that did not provide this information.

Questions 4a, 4b, 5a, and 5b are needed to identify eligible diabetic parents under the de jure-children counting rule. The two most important items, Q.4b and Q.5b, have item nonresponse rates of 2.2 percent and 3.3 percent, respectively. However, for these items also, the item nonresponse rates are smaller for diabetics than for nondiabetics.

Table 2 presents the item nonresponse rates for proxy and self respondents. For females, the nonresponse rates are consistently higher for proxy respondents than for self respondents. For males, their is practically no difference between the item nonresponse rates for proxy respondents and self respondents.

SUMMARY AND CONCLUSIONS

Based on the 1976 HIS, national estimates of diabetes prevalence were larger for the de jure-children estimator than for either the de jure estimator or the de jure-sibling estimator. We suspect that the de jure-children estimator may have less reporting bias than either the de jure or the de jure-sibling estimators. However, in the absence of criterion variables, it is impossible to categorically state that one of the estimators is more valid than the others.

The sampling errors of the estimates of diabetes prevalence based on the network estimator were almost always smaller than those based on the de jure estimator. For most of the population domains considered in this paper, the de jure estimator would require a household sample size nearly 100 percent larger than the de jure-sibling estimator.

ACKNOWLEDGEMENTS

The authors would like to thank Vance Hudgins and Sue Hsiung in the Health Interview Survey Division for their special efforts in computing the estimates and the standard errors presented in Table 1.

Table 2. Item Nonresponse Rates by Color, Age, Diabetic Status, and Respondent Status

Population Characteristics	Questions						
	Q.2 Number of living children?	Q.3a Number of living siblings?	Q.3b Number of living siblings with diabetes?	Q.4a Is mother living?	Q.4b Does mother have diabetes?	Q.5a Is father living?	Q.5b Does father have diabetes?
Total	0.8%	1.3%	2.7%	1.5%	2.2%	2.1%	3.3%
Under 25 years	0.8	1.9	2.8	2.1	3.0	2.8	4.8
25-44 years	0.6	0.7	1.7	1.1	1.8	1.8	2.9
45-64 years	0.7	0.9	2.8	0.9	1.3	1.3	1.6
65 years and over	0.9	1.5	4.3	1.2	1.1	1.3	1.3
Male 20 years and over	0.9	0.9	2.8	1.1	1.8	1.6	2.5
Self respondent	0.7	0.9	2.5	1.1	1.5	1.6	2.3
Nonself respondent	1.0	0.9	3.1	1.1	2.0	1.6	2.7
Female 20 years and over	0.7	0.9	2.1	1.0	1.3	1.5	2.2
Self respondent	0.5	0.7	1.6	0.9	1.1	1.4	1.9
Nonself respondent	1.5	1.8	4.1	1.5	2.3	2.0	3.0
White	0.7	1.3	2.5	1.5	2.1	1.9	2.8
Nonwhite	1.0	1.9	4.2	1.7	2.8	3.3	6.6
Diabetic	0.3	0.8	3.9	0.7	0.9	1.2	1.6
Nondiabetic	0.8	1.4	2.6	1.5	2.2	2.1	3.3

REFERENCES

[1] Birnbaum, Z.W. and Sirken, Monroe G., "Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates," Vital and Health Statistics, Series 2, No.11, October 1965.

[2] Madow, William G., "Net Differences in Interview Data on Chronic Conditions and Information Derived from Medical Records," Vital and Health Statistics, Series 2, No. 57, 1973.

[3] Nathan, Gad, Schmelz U.O., and Kenvin, J., "Multiplicity Study of Marriages and Births in Israel," Vital and Health Statistics, Series 2, No. 70, 1977.

[4] National Center for Health Statistics, "Profile of Chronic Illness in Nursing Homes, United States, August 1973-April 1974," Vital and Health Statistics, Series 13, No. 29, 1977, pp. 25-29.

[5] Sirken, Monroe G., "Household Surveys with Multiplicity," Journal of the American Statistical Association, Vol. 65, March 1970, pp. 257-266.

[6] Sirken, Monroe G., "Design of Household Sample Surveys to Test Death Registration Completeness," Demography, Vol. 10, No. 3, August 1970, pp. 469-478.

[7] Sirken, Monroe G., "Alcohol and Other Drug Use and Abuse in State of Michigan," Office of Substance Abuse Services, Michigan Department of Public Health, April 1975.

[8] Sirken, Monroe G., Inderfurth, G.P., Burnham, C.E., and Danchik, K.M., "Household Sample Surveys of Diabetes: Design Effects of Counting Rules," American Statistical Association Proceedings of the Social Statistics Section, 1975.

[9] Sirken, Monroe G. and Royston, Patricia N., "Reasons Deaths are Missed in Household Surveys of Population Change," American Statistical Association Proceedings of the Social Statistics Section, 1970, pp. 361-364.

[10] Sirken, Monroe G., Graubard, Barry I., and LaValley, Richard W., "Evaluation of Census Population Coverage by Network Surveys," American Statistical Association Proceedings of the Survey Statistics Section, 1978.

APPENDIX

Selected Items of Information on Diabetes Collected for Each Household Member: The 1976 Health Interview Survey Questionnaire

Q.1 Do you have diabetes or sugar diabetes?

Q.2 How many children do you have?

Q.3a How many brothers and sisters do you have?

Q.3b How many of your siblings have diabetes or sugar diabetes?

Q.4a Is your mother still living?

Q.4b Does she have diabetes or sugar diabetes?

Q.5a Is your father still living?

Q.5b Does he have diabetes or sugar diabetes?