

# Proxy reporting in the National Population Health Survey

Margot Shields

## Abstract

### Objectives

This article examines the extent of proxy reporting in the National Population Health Survey (NPHS). It also explores associations between proxy reporting status and the prevalence of selected health problems, and investigates the relationship between changes in proxy reporting status and two-year incidence of health problems.

### Data source

Cross-sectional results are based on the 1996/97 NPHS Health file and General file. Longitudinal results are based on 1994/95 respondents who were still residing in households in 1996/97.

### Analytical techniques

The extent of proxy reporting in the various NPHS files was computed. Prevalence estimates of selected health problems from the two 1996/97 cross-sectional files were compared. Multivariate analyses were used to estimate associations between proxy reporting status and health problems.

### Main results

For several health conditions, prevalence estimates based on the 1996/97 cross-sectional Health file (where proxy reporting was less common) were significantly higher than estimates derived from the General file. Individuals whose data were proxy-reported in 1994/95 and self-reported in 1996/97 had higher odds of reporting new cases of certain health conditions.

### Key words

chronic conditions, activity limitation, health surveys

### Author

Margot Shields (613-951-4177; shiemar@statcan.ca) is with the Health Statistics Division at Statistics Canada, Ottawa, Ontario, K1A 0T6.

Proxy reporting often poses a dilemma to survey designers. While it is generally believed that information will be more accurate if it is provided directly by the individuals selected for the survey sample, there are often compelling reasons for accepting information provided on their behalf by others (proxy responses). One major reason for doing so is to reduce non-response rates; another is to reduce the costs of data collection.

Although previous research has not produced conclusive evidence,<sup>1</sup> several studies have suggested the possibility of a “proxy effect” for data collected in health surveys.<sup>2-13</sup> That is, individuals providing information on behalf of others may be less likely to report health events than people who give their own information directly. Lower estimates of hospitalization, chronic conditions, activity limitations, emotional problems, and pain have been attributed to proxy reporting.<sup>2-13</sup> For conditions based on a relatively short reference period (two weeks, for example), including acute illness, disability, and dental and doctor visits, findings have been contradictory. Some studies have found that self-respondents are more likely than proxy reporters to report events of this nature,<sup>4,7,8,10,11</sup> while others have not.<sup>6,14,15</sup> And

some comparisons of survey data with medical files and physician reports suggest that even self-respondents may under-report health events.<sup>2,12,13,16,17</sup>

When the National Population Health Survey (NPHS) was designed in the early 1990s, the issue of proxy response was debated extensively. For several reasons, it was decided that the NPHS—like many other surveys—would accept proxy reporting. However, for certain components of the survey, proxy responses are strongly discouraged.

This article evaluates the extent of proxy reporting in the 1996/97 NPHS cross-sectional files. (See *Data source*, *Analytical techniques* and *Limitations*.) The characteristics of individuals with proxy response status (that is, those whose survey information was provided by another household member) are presented. The relationship between proxy reporting and the prevalence of various health conditions is examined to determine if there is a proxy effect. The relationship between new cases

### Data source

The National Population Health Survey (NPHS), which began in 1994/95, collects information about the health of the Canadian population every two years. It covers household and institutional residents in all provinces and territories, except persons living on Indian reserves, on Canadian Forces bases, and in some remote areas. The NPHS is both longitudinal and cross-sectional. Longitudinal panel members will be followed for up to 20 years.

The analysis in this article is based on cross-sectional household data from the second cycle (1996/97) of the NPHS and longitudinal data for the first (1994/95) and second cycles (both for the 10 provinces). The 1996/97 cross-sectional sample is made up of longitudinal respondents and respondents who were selected as part of supplemental samples, or buy-ins, in three provinces. The additional respondents for these buy-ins were chosen using the random digit dialling technique and were included for cross-sectional purposes only.

The general component of the questionnaire was used to collect socio-demographic and some health information for each member of participating households. These data are found in the General file. The health component of the questionnaire was used to collect additional in-depth health information for one randomly selected household member. This additional information, as well as the information collected in the general component pertaining to that person, is found in the Health file.

The 1994/95 provincial, non-institutional sample consisted of 27,263 households, 88.7% of which agreed to participate. After the application of a screening rule (to ensure a more representative sample), 20,725 households remained in scope. In 18,342 of these households, the randomly selected person was aged 12 or older. Their response rate to the in-depth health component was 96.1% or 17,626 respondents. Of these 17,626 randomly selected

respondents, 14,786 were eligible members of the NPHS longitudinal panel. In addition, 468 selected respondents for whom only general information had been collected in 1994/95 and 2,022 randomly selected respondents younger than 12 were also eligible. Thus, a total of 17,276 longitudinal panel members were eligible for re-interview in 1996/97. The remaining respondents to the 1994/95 survey were sponsored by the provincial governments that elected to enlarge the sample size in their province for cycle 1 only. These respondents were not followed up.

A response rate of 93.6% was achieved for the longitudinal panel in 1996/97. Of these 16,168 respondents, full information was available for 15,670; that is, general and in-depth health information for both cycles of the survey.

The one-time participation of additional respondents in cycle 2 for cross-sectional purposes yielded a total of 210,377 respondents to the general component (173,216 aged 12 or older) and 81,804 respondents to the health component (73,402 aged 12 or older). The overall response rate to the health component was 79.0%.

The data were weighted taking into account the sample design, adjustments for non-response, and post-stratification. The cross-sectional analysis is based on data from both the General and Health files for individuals aged 12 or older, weighted to represent about 24.6 million Canadians in the 10 provinces. The longitudinal analysis is based on individuals who were 12 or older in 1994/95, who were still living in a household in 1996/97 (that is, non-institutionalized) and for whom complete data (from the general and health components) were provided in both cycles. This sample of 13,427 was weighted to represent about 23.5 million Canadians.

More detailed descriptions of the NPHS design, sample and interview procedures can be found in published reports.<sup>18-20</sup>

(incidence) of selected health conditions and changes in proxy reporting status between cycles is also investigated, based on longitudinal data from the first two cycles of the NPHS (1994/95 and 1996/97).

### Proxy versus self-response

Proxy reporting (see *Definitions*) is one of several methodological factors that may affect the accuracy of survey results—especially for health surveys, which often ask very personal questions. For a number of reasons,<sup>8,12</sup> data reported by proxy may be inaccurate:

- The reporter may not be fully aware of the health situation of another household member.

For proxy reporters to answer accurately, they must be fully aware of the health situation of the person for whom they are replying. Sometimes, individuals may deliberately conceal health problems or behaviours. For example, a husband may not tell his wife that he has been diagnosed with an ulcer, or a teenager may not want her parents to know that she smokes. Or a proxy reporter may not be aware of certain health information. For instance, the reporter may not know about all physician consultations another household member has had over the past 12 months.

- The reporter may not recall relevant health information.

The ability to recall information depends on its importance to the individual reporting it. Health conditions that are more serious, painful, persistent or potentially life-threatening are, in general, reported more often and more accurately.<sup>2,4,8,17</sup> Clearly, a condition is likely to be more immediate and important to the individual affected than to another household member. However, those other household members would be more likely to be aware of conditions such as diabetes or heart disease, because they are often reminded of them (for instance, they see the family member taking medication) and because such conditions may seriously affect the family member's health. Conditions such as allergies—to penicillin, for example—may be more easily forgotten by proxy reporters.

- The reporter may mislabel health problems. Whether provided by proxy or by the individual affected, information tends to be more accurate if it pertains to conditions that are easily defined and labelled (diabetes and heart disease, for example), versus conditions that are more difficult to describe (such as recurring back trouble or chronic skin diseases).<sup>13</sup> For conditions in the latter category, information reported by proxy tends to be less accurate than self-reports.<sup>13</sup> And conditions that are not directly observable—pain or emotional distress, for example—are generally less adequately measured by proxy.<sup>5</sup>

- The reporter may deliberately not report certain information.

Conditions that are perceived to be very personal or potentially embarrassing tend to be under-reported. For example, several studies comparing survey responses with medical records have found that mental illness is under-reported.<sup>16,17</sup> In such cases, both self- and proxy reports may be subject to under-reporting.

### Why accept proxy responses?

Given the problems that attend proxy data, it might seem that such responses should be avoided. However, there are several compelling reasons for accepting information provided by proxy.

The NPHS collects certain health information for all members of a household so that intra-household relationships of health characteristics can be investigated.<sup>18</sup> But contacting and interviewing each member of every sampled household is time-consuming and expensive, and often requires several follow-up calls. To save time and money, the NPHS, like many other health surveys, allows one knowledgeable household member—a proxy reporter—to answer questions on behalf of all people in the household.

The nature of the information to be collected is an important consideration in deciding whether to accept proxy responses. It is generally believed that opinions, attitudes, and questions of a subjective nature are best answered directly by the individual selected for the survey, while questions that solicit factual information about specific health conditions may be answered by proxy reporters.

Proxy information is also accepted because some people may be difficult to contact. If proxy responses are not accepted for such individuals, a survey risks a higher non-response rate. Furthermore, the characteristics of these “hard-to-contact” individuals are often very different from those of the general population, thereby also introducing the possibility of inaccurate estimates.

Finally, some individuals selected for the survey may be unable to provide their own information. Thus, many health surveys accept proxy information about individuals who cannot respond because of a physical or mental condition. And parents usually respond on behalf of their children.

### NPHS proxy response rules

The NPHS questionnaire has two major sections: the general and health components. Socio-

demographic and some health information is collected for each member of participating households using the general component. The health component is used to collect additional in-depth health information for one randomly selected household member.

The proxy response rules differ for the general and health components of the questionnaire (see *Definitions*). For the general component, where the information sought is, for the most part, objective and factual, interviewers are instructed to obtain the information about each person in the household from one knowledgeable member. The *Interviewer's Manual* suggests that, in many cases, this person “should be the mother, since she knows more about health status and utilization of health services in the family.”

## Analytical techniques

All analyses are based on weighted data. The percentages of individuals with proxy reporting status are presented for the population aged 12 or older, by sex, for both the cross-sectional files (General and Health) and the longitudinal file.

Multiple logistic regression was used to explore the relationship between proxy reporting status and various health outcomes. Based on NPHS respondent selection rules and a review of the literature on proxy reporting, several socio-demographic factors that are considered to be related to proxy reporting status were included in the models as control variables: sex, age group, presence of other adults in the household, marital status, education, employment status, income and self-reported health. With data from the cross-sectional Health file, multiple logistic regression was used to model the relationship between proxy reporting status in the general component and the reported prevalence of various health conditions. A separate model was fitted for each health condition considered.

Since the sex distribution of the proxy reporters differed by the sex of the individual for whom they reported, the regression models were recalculated, this time taking the sex of the proxy reporter into consideration. Three categories were considered for proxy reporting status: self-response, female proxy reporter, and male proxy reporter. The third category was used as the reference.

To explore the possibility of a proxy effect in incidence estimates of health conditions (new cases of a health condition in the two years between survey cycles), multiple logistic regression models were used to relate the incidence of a selected number of new health

problems to changes in proxy reporting status. A separate model was fitted for each new health problem considered. In this case, proxy reporting status across survey cycles was defined as follows: proxy response in 1994/95 and 1996/97; proxy response in 1994/95 to self-response in 1996/97; self-response in 1994/95 to proxy response in 1996/97; and self-response in 1994/95 and 1996/97. The control variables entered into the models included the ones used in the cross-sectional analyses, as well as other variables to reflect changes between the two years. (See Appendix Table G for a complete list of the variables included in the longitudinal models.)

Because the two-year incidence rates for many health conditions were relatively low, sample sizes were often too small to ensure reliable results in the regression models. Only conditions for which there were at least 30 new cases for each of the four patterns of proxy reporting status were considered: other allergies, arthritis, back problems (excluding arthritis), activity restriction and long-term disability or handicap.

For all analyses, the bootstrap procedure was used to estimate sampling error in the estimation of coefficients of variation, in tests of significance on differences between rates, and in the calculation of confidence intervals for odds ratios.<sup>21-23</sup> This procedure yields an unbiased estimate of variance. Bonferroni tests of significance of differences between rates were used to take multiple comparisons into account. Critical values were adjusted in testing, according to the number of comparisons being made.

The health component is completed for one randomly selected household member, and the importance of having that individual provide information directly (non-proxy) is stressed in the *Interviewer's Manual* and during training. The detailed—and often very personal—nature of the information sought makes it essential that questions be answered directly by the selected individual. However, if that person is younger than 12, the information is collected from a parent. Proxy response is also accepted if the selected person is unable to answer because of special circumstances (for example, disability). Questions that are highly subjective and personal (mental health and social support, for example) are left unanswered if the only alternative would be to accept information from a proxy reporter. (See *Proxy reporting for the health component*.)

These rules were used for the first two cycles of the NPHS (1994/95 and 1996/97). In cycle 3 (1998/99), a slight modification was introduced for the general component of the questionnaire. Since

randomly selected respondents are followed over time to produce the longitudinal file and the main purpose of the longitudinal file is to measure change, there was some concern that changes in proxy reporting status across cycles might distort measures of change. Therefore, for cycle 3, interviewers were instructed to collect information directly from the longitudinal panel member (non-proxy) for the entire questionnaire (the general and health components). That is, the rules previously used for the health component would also apply to the general component for longitudinal respondents. For the general component, however, proxy response could still be accepted to avoid non-response. The original rules (relying on a knowledgeable source for information pertaining to all household members and discouraging proxy reporting on behalf of the randomly selected member for the health component) still apply to households and individuals selected for the top-up sample (those added solely to maintain representative cross-sectional files).

## Limitations

This analysis compares estimates of reported health problems between self- and proxy responses in a multivariate model, controlling for other socio-demographic confounders possibly related to an individual's proxy response status. Although every effort was made to consider all relevant variables, the results could be misleading if other variables related to morbidity and proxy reporting status were inadvertently excluded, or are not available from the National Population Health Survey (NPHS).

Other methods could have been used to assess the potential of a proxy effect on estimates of health conditions based on NPHS data. One would be to match the information with medical records and look for differences between self- and proxy-reported information. However, such a comparison would not only be complicated and expensive, but would also require assumptions about the completeness and accuracy of, and the success of matching, the medical records. A second possibility would be to conduct a controlled experiment in which data would be collected twice for some individuals: once from a knowledgeable household member and again directly from the individual. This would allow a comparison

of proxy and self-reports. It would be necessary to ensure that each reporter was unaware of the other's responses. Again, this approach would be complicated and expensive, and would place a heavier burden on participating households.

It is assumed that the differences in the reporting of health problems between self- and proxy responses are caused by under-reporting by proxy reporters. While research based on comparisons with medical records has shown that under-reporting is more common for proxy reports,<sup>2,12,13</sup> it is possible that, in some cases, the problem may be over-reporting by self-respondents. For example, the criteria for chronic conditions to be reported in the NPHS are: "long-term conditions that have lasted or are expected to last six months or more and that have been diagnosed by a health professional." Although reporters are reminded of this by interviewers, some individuals may report conditions that do not fully meet the criteria. For instance, a person who has been suffering from back pain for several months may report a back problem even if it has not been diagnosed by a health professional. If this is more likely to happen for self-responses, the effect would be over-reporting of morbidity.

## Definitions

In the National Population Health Survey (NPHS), *proxy responses* are those obtained for a particular household member from another knowledgeable member of the household (the proxy reporter); for example, a parent may provide answers for a child, or a wife may respond on behalf of her husband.

*Self-responses* are those obtained directly from the individuals selected for the survey.

The *general component* of the NPHS contains questions on demographic and socioeconomic characteristics, as well as limited health information such as two-week disability, health care utilization, restriction of activities, and chronic conditions. This information is obtained for each person in the household from a knowledgeable household member.

The *health component* contains detailed questions on topics such as general self-perceived health, health status, medication use, smoking, alcohol consumption, mental health, social support, blood pressure check, height and weight, physical activity, and injuries. Given the detailed nature of the information requested, the randomly selected individual in the household usually answers on his or her own behalf (self- or non-proxy response).

To measure *chronic conditions*, individuals were asked about any “long-term conditions that have lasted or are expected to last six months or longer and that have been diagnosed by a health professional.” Interviewers read a list of conditions and then asked about “any other long-term condition that has been diagnosed by a health professional” (such as cystic fibrosis or multiple sclerosis).

Two *other health conditions* were included in this analysis. Activity restriction due to a long-term physical or mental health problem is measured by a derived variable based on a positive response to any of the following questions, which were asked about every person in the household: “Because of a long-term physical or mental condition or a health problem, are you limited in the kind or amount of activity you can do: at home? at school? at work? in other activities?” Long-term disability or handicap was determined by asking, “Do you have any long-term disabilities or handicaps?”

Five *age groups* were defined for this analysis: 12 to 17, 18 to 24, 25 to 44, 45 to 64, and 65 or older.

Individuals were asked for their current *marital status*. Those who chose “now married,” “common-law” or “living with a partner” were

grouped as “married.” Individuals who answered “single” were classified as “never married,” and responses of “widowed,” “separated” or “divorced” were categorized as “previously married.”

A derived variable, based on the ages of household members, was used to establish the presence of *other adult(s) aged 18 or older in household*.

*Education* was grouped into three categories, based on the highest level attained: high school graduation or less, some postsecondary, and postsecondary (college, trade school or university) graduation.

Individuals who were currently working at a job or business were considered to be *employed*.

*Household income group* was defined based on the number of people in the household and total household income from all sources in the 12 months before the interview.

Household income group	People in household	Total household income
Lowest	1 to 4	Less than \$10,000
	5 or more	Less than \$15,000
Lower-middle	1 or 2	\$10,000 to \$14,999
	3 or 4	\$10,000 to \$19,999
	5 or more	\$15,000 to \$29,999
Middle	1 or 2	\$15,000 to \$29,999
	3 or 4	\$20,000 to \$39,999
	5 or more	\$30,000 to \$59,999
Upper-middle	1 or 2	\$30,000 to \$59,999
	3 or 4	\$40,000 to \$79,999
	5 or more	\$60,000 to \$79,999
Highest	1 or 2	\$60,000 or more
	3 or more	\$80,000 or more

*Self-perceived health* was assessed with the question, “In general, would you say your health is: excellent? very good? good? fair? poor?” In the longitudinal analyses, an individual was classified as having improved general health if the 1996/97 rating was better than that given in 1994/95 (for example, an individual reported fair health in 1994/95, then good health in 1996/97). Likewise, an individual was classified as having had a decline in health if the 1996/97 rating was worse than that reported in 1994/95.

## Cross-sectional files

The information collected using the general and health components forms two cross-sectional data files, named the General file and the Health file. The General file comprises the data collected using the general component of the questionnaire. This file contains separate records for *all* members of participating households, based on the socio-demographic and basic health information provided for everyone by one knowledgeable household member. The Health file contains *one* record per household with more in-depth health information about the randomly selected household member, along with the information collected about that person in the general component.

Because the proxy response rules differ for the two survey components, some records on the Health file can show a “mixed” proxy status. For example, in a household comprising a married couple and one child, the mother could be identified as the “knowledgeable household member,” completing the general component for herself, her spouse and her child. The husband could be randomly selected to complete the health component, providing his own information. His record on the Health file would therefore have proxy responses for the general component and self-responses for the health component.

## Comparison of cross-sectional files

The General and Health files can each be used to produce estimates for any items included in the general component of the questionnaire. The sample sizes, however, differ. Estimates produced using data from the General file have the advantage of a larger sample size, since they are based on information pertaining to all members of selected households. By contrast, Health file estimates are based on one randomly selected person per household.

The proxy reporting rates for the two files are quite different. For the General file, 42% of responses for the population aged 12 or older were completed by proxy. In contrast, the proxy reporting rate for the general component of the Health file was 28% (Table 1).

Table 1  
Percentage of proxy responses to general component, General and Health files, 1996/97 National Population Health Survey, by sex and age, household population aged 12 or older, Canada excluding territories

	Proxy responses, general component			
	General file		Health file	
	'000	%	'000	%
<b>Total</b>	<b>24,595</b>	<b>42.2</b>	<b>24,595</b>	<b>27.5</b>
<b>Sex</b>				
Male	12,099	53.9†	12,099	36.5†
Female	12,495	31.0	12,495	18.7
<b>Age group</b>				
12-17	2,445	81.6‡	2,435	61.7‡
18-24	2,689	51.4‡	2,699	29.6‡
25-44	9,709	36.3	9,709	22.4
45-64	6,335	39.1§	6,335	25.2§
65+	3,416	29.6††	3,416	19.9††

**Data source:** National Population Health Survey, cross-sectional sample, General and Health files, 1996/97

**Notes:** Bonferroni significance tests were used to adjust for multiple comparisons. Estimates on the General and Health files are post-stratified to agree with census projections by sex and age group. Since the age groups used in this analysis are not identical to those used in the post-stratification, the population estimates from the two files differ slightly in some cases (12-17 and 18-24 age groups). See Appendix Table A for unweighted sample counts.

† Significantly higher than females ( $p \leq 0.05$ )

‡ Significantly higher than each older age group ( $p \leq 0.05$ )

§ Significantly higher than 45-64 age group ( $p \leq 0.05$ )

†† Significantly lower than each younger age group ( $p \leq 0.05$ )

Patterns by sex and age group in the two files were similar, however. Information for males was more likely than that for females to have been reported by proxy. Among age groups, responses for 12- to 17-year-olds were most likely to have been reported by proxy; those for people aged 65 or older, least likely. This is not surprising, given the proxy reporting rule requiring interviewers to ask one knowledgeable household member to provide information about all household members for the general component of the questionnaire. For children aged 12 to 17, this individual would likely be a parent. The relatively low proxy reporting rate for people aged 65 or older may reflect the fact that seniors are more likely than younger people to live alone. In such cases, no one else would be available in the household to provide a proxy response.

Because each file is weighted separately to represent the same total Canadian population by sex and age group, estimates of health characteristics

derived from either one should be very close. But, if it is true that individuals responding on their own behalf are more likely to report health events, estimates based on the Health file may be higher (and probably more accurate), since it has a lower proxy reporting rate.

### Health file yields higher estimates

A comparison of prevalence estimates of various health outcomes reveals that the Health file does

yield significantly higher estimates than the General file for many conditions: food allergies, other allergies, asthma, arthritis or rheumatism, back problems (excluding arthritis), high blood pressure, migraine, sinusitis, thyroid disorder, activity restriction, and long-term disability or handicap (Table 2). For sinusitis and thyroid disorder, the prevalence estimates for females, but not for males, are greater in the Health file than in the General file.

Table 2

**Prevalence estimates of chronic and other health conditions, general component, General and Health files, 1996/97 National Population Health Survey, by sex, household population aged 12 or older, Canada excluding territories**

	Both sexes		Males		Females	
	General file	Health file	General file	Health file	General file	Health file
	%		%		%	
<b>Chronic conditions</b>						
Food allergies	6.3	6.8*	4.7	5.0	7.9	8.5
Other allergies	20.0	22.4****	16.5	18.6****	23.4	26.0****
Asthma	6.8	7.2*	5.7	6.0	7.9	8.4
Arthritis or rheumatism	12.6	13.8****	8.8	9.7**	16.3	17.8****
Back problems (excluding arthritis)	12.8	14.2****	12.2	13.4**	13.4	15.0****
High blood pressure	9.6	10.1*	8.3	8.7	10.8	11.4*
Migraine	6.8	7.8****	3.5	4.3**	10.1	11.2**
Chronic bronchitis or emphysema	2.6	2.8	2.2	2.3	3.0	3.3
Sinusitis	4.2	4.6*	3.1	3.4	5.2	5.7*
Diabetes	3.2	3.2	3.4	3.5	3.1	2.9
Epilepsy	0.5	0.6	0.6	0.7	0.5	0.6†
Heart disease	3.9	3.9	4.1	4.0	3.7	3.7
Cancer	1.4	1.5	1.2	1.2	1.7	1.8
Stomach or intestinal ulcers	2.5	2.7	2.4	2.5	2.6	2.9
Effects of a stroke	0.8	0.9	0.8	1.0	0.8	0.8
Urinary incontinence	1.4	1.5	1.0	1.0	1.8	2.0
Bowel disorder	1.5	1.5	1.0	1.0	1.9	2.1
Alzheimer's disease/Other dementia‡	0.3	0.3	0.3	0.3	0.3	0.3
Cataracts‡	2.9	3.0	2.1	2.3	3.7	3.6
Glaucoma‡	1.2	1.2	1.0	1.1	1.4	1.4
Thyroid disorder	3.2	3.5*	1.0	1.2	5.3	5.8*
Other§	5.5	5.6	4.9	5.2	6.0	6.0
At least one chronic condition	54.1	57.7****	49.1	52.8****	59.0	62.4****
<b>Other health conditions</b>						
Activity restriction due to long-term physical or mental health problem	11.9	13.0***	10.7	11.8**	12.9	14.1**
Long-term disability or handicap	10.7	11.6**	10.4	11.1*	10.9	12.1**

**Data source:** National Population Health Survey, cross-sectional sample, General and Health files, 1996/97

**Notes:** Prevalence estimates were calculated excluding missing values. The percentage of missing values was less than one-quarter of a percentage point for each condition. Unweighted sample counts of the number of records with a report of each health condition (by proxy response status) for the General and Health files appear in Appendix Table B. A one-tailed significance test was used in comparing the prevalence estimates; based on the literature, it was hypothesized that the General file would yield lower prevalence estimates because it has a higher proxy reporting rate.

† Coefficient of variation between 16.6% and 25.0%

‡ Only asked for population aged 18 or older

§ Any other long-term chronic condition; for example, cystic fibrosis or multiple sclerosis

\* Significantly higher than General file estimate ( $p \leq 0.05$ )

\*\* Significantly higher than General file estimate ( $p \leq 0.01$ )

\*\*\* Significantly higher than General file estimate ( $p \leq 0.001$ )

\*\*\*\* Significantly higher than General file estimate ( $p \leq 0.0001$ )

In keeping with earlier research,<sup>2,4,8,13</sup> estimates were closer for conditions that proxy reporters would be more likely to notice and less likely to mislabel: diabetes, epilepsy, heart disease and cancer, for example.

This comparison of estimates from the general component of the General and Health files suggests the possibility of a proxy effect, resulting in underestimates of certain health conditions. Because the General file has a higher proxy reporting rate, estimates based on the file would tend to be more subject to this proxy effect. Nonetheless, the effect on the Health file itself (for which the proxy reporting rate was 28%) may not be negligible.

### Characteristics of individuals with proxy response status

In exploring the association between health variables and proxy reporting, it must be kept in mind that proxy response to the NPHS general component is by no means random. Interviewers were given precise instructions about who could report for others. The population deemed to be “knowledgeable household members” did, for the most part, report their own information, while others in the household had their information provided by this proxy reporter. To understand if proxy responses affect the measurement of health outcomes, the characteristics of people for whom information is provided by proxy must be taken into account.

A multiple logistic regression model was set up using proxy response (“yes” or “no”) as the dependent variable. The model is based on proxy response to the general component of the Health file because this file is used most often in analytical studies.

Proxy reporting was related to sex and age (Table 3). Males had significantly higher odds of having proxy responses provided for them than did females. And younger people (12 to 17 and 18 to 24) of both sexes had higher odds of having their information provided by proxy, compared with individuals aged 25 to 44. Although seniors (65 or older) have a lower proxy rate than any younger age group (Table 1), the odds of proxy response were

Table 3  
Adjusted odds ratios relating selected characteristics to proxy response in general component, Health file, 1996/97 National Population Health Survey, household population aged 12 or older, Canada excluding territories

	Proxy response <sup>†</sup>	
	Adjusted odds ratio	95% confidence interval
<b>Sex</b>		
Males	2.5****	2.3, 2.7
Females <sup>‡</sup>	1.0	...
<b>Age group</b>		
12-17	4.3****	3.6, 5.2
18-24	1.4****	1.2, 1.6
25-44 <sup>‡</sup>	1.0	...
45-64	1.1	1.0, 1.2
65+	1.0	0.9, 1.2
<b>Marital status</b>		
Married	1.1	1.0, 1.4
Never married	1.2	0.9, 1.4
Previously married <sup>‡</sup>	1.0	...
<b>Other adult(s) aged 18 or older in household<sup>§</sup></b>	115.7****	85.5, 156.6
<b>Education</b>		
High school graduation or less	1.3****	1.2, 1.5
Some postsecondary	1.0	0.9, 1.1
Postsecondary (college, trade or university) graduation <sup>‡</sup>	1.0	...
<b>Employed<sup>§</sup></b>	1.1*	1.0, 1.2
<b>Household income</b>		
Lowest/Lower-middle <sup>‡</sup>	1.0	...
Middle	1.1	1.0, 1.3
Upper-middle	1.1	1.0, 1.3
Highest	1.4***	1.1, 1.6
<b>Self-perceived health<sup>††</sup></b>		
Excellent <sup>‡</sup>	1.0	...
Very good	1.1	1.0, 1.2
Good	1.1*	1.0, 1.3
Fair	1.3****	1.2, 1.5
Poor	1.8****	1.4, 2.2

**Data source:** National Population Health Survey, cross-sectional sample, Health file, 1996/97

**Notes:** Based on 73,241 records, of which 20,451 were proxy responses to the general component; 161 records were removed from the analyses due to missing values. Missing categories for education, work status and income were included in the model to maximize sample size, but their respective odds ratios are not shown. Because of rounding, some confidence intervals with 1.0 as the lower limit were significant.

<sup>†</sup> Reference category is self-response.

<sup>‡</sup> Reference category, for which odds ratio is always 1.0

<sup>§</sup> Reference category is absence of characteristic; for example, reference category for “Other adult(s) aged 18 or older in household” is “no other adult(s) aged 18 or older in household.”

<sup>††</sup> Collected as part of the health component where proxy reporting was very low (2.6%); used to predict proxy response to the general component

\*  $p \leq 0.05$

\*\*\*  $p \leq 0.001$

\*\*\*\*  $p \leq 0.0001$

... Not applicable

not significantly lower for this age group, compared with 25- to 44-year-olds. Older individuals are more likely to live alone; therefore, including marital status and the presence of other adult(s) aged 18 or older in the regression model renders the odds ratio for seniors not statistically different. If these variables are removed from the model, older individuals have decreased odds of proxy reporting (data not shown). Not surprisingly, the odds ratio for the presence of other adults was extremely high. If there are no other adults in the household, proxy reporting is rare, since no one is available to act as proxy reporter.

When all other variables were taken into account, proxy reporting was also associated with lower levels of education. However, NPHS selection procedures partially account for this association. The household members for whom information was proxy-reported (especially children, who were still in school) tended to have less education than the reporting person, resulting in the observed relationship.

The association between higher income and proxy reporting is not surprising, given that households with two or more adults were more likely to be in the highest or upper-middle income group, compared with households with only one adult (data not shown), where proxy reporting is less common.

Some research has suggested that individuals whose information is reported by proxy are in better health.<sup>9</sup> (That is, because they are in better health, they are more likely to be busy and away from home; therefore, someone else provides their information.) It was possible to control for this potentially confounding factor in this analysis by using the information on self-perceived health collected in the health component, where the proxy reporting rate was extremely low (2.6%). And counter to the conjecture made in the previous study,<sup>9</sup> NPHS data showed that individuals in poor, fair, or good health had higher odds of having their information (for the general component) reported by proxy, compared with those in excellent health.

## “Proxy effect” and prevalence estimates

To investigate a potential proxy effect on estimates of various health conditions, a series of multiple

Table 4  
Adjusted odds ratios relating chronic and other health conditions to reporting status in the general component, Health file, 1996/97 National Population Health Survey, population aged 12 or older, Canada excluding territories

	Self-response <sup>†</sup>	
	Adjusted odds ratio	95% confidence interval
<b>Chronic conditions<sup>‡</sup></b>		
Food allergies	1.3***	1.1, 1.5
Other allergies	1.6****	1.4, 1.7
Asthma	1.3***	1.1, 1.6
Arthritis or rheumatism	1.9****	1.6, 2.1
Back problems (excluding arthritis)	1.5****	1.3, 1.6
High blood pressure	1.2**	1.1, 1.4
Migraine	1.4***	1.2, 1.7
Chronic bronchitis or emphysema	1.6***	1.2, 2.0
Sinusitis	1.5****	1.2, 1.8
Diabetes	1.1	0.9, 1.4
Epilepsy	1.6	0.9, 2.7
Heart disease	1.1	0.9, 1.4
Cancer	1.1	0.8, 1.5
Stomach or intestinal ulcers	1.8****	1.4, 2.3
Effects of stroke	0.8	0.6, 1.2
Urinary incontinence	1.4*	1.0, 1.8
Bowel disorder	1.5**	1.1, 1.9
Cataracts	1.3*	1.0, 1.7
Glaucoma	1.0	0.7, 1.4
Thyroid disorder	1.6****	1.3, 2.0
Other <sup>§</sup>	1.4***	1.1, 1.6
At least one chronic condition <sup>‡</sup>	1.6****	1.4, 1.7
<b>Other health conditions</b>		
Activity restriction due to long-term physical or mental health problem	1.3****	1.1, 1.4
Long-term disability or handicap	1.2**	1.1, 1.4

**Data source:** National Population Health Survey, cross-sectional sample, Health file, 1996/97

**Notes:** Presents results of the 24 separate regression models based on 73,402 records on the Health file. Sample counts of the number of records for which each health condition was reported appear in Appendix Table B. In each case, a small percentage (< 0.5%) of records was dropped because of missing values. Each regression includes selected control variables, but only the odds ratios for proxy reporting status are presented. (See Appendix Table F for a complete list of independent variables.) Results for complete models are available on request. Because of rounding, some confidence intervals with 1.0 as the lower limit were significant.

<sup>†</sup> Reference category is proxy response.

<sup>‡</sup> Alzheimer's disease was not included, since the proxy reporting rate for this condition was very high.

<sup>§</sup> Any other long-term chronic condition; for example, cystic fibrosis or multiple sclerosis

\*  $p \leq 0.05$

\*\*  $p \leq 0.01$

\*\*\*  $p \leq 0.001$

\*\*\*\*  $p \leq 0.0001$

logistic regression models, based on data from the Health file, was used. For each regression, the outcome was a specific health condition. When all variables thought to be related to an individual's proxy response status were controlled, self-reporting individuals had higher odds of reporting most of the health conditions studied (Table 4), compared with those whose information was provided by proxy. But, consistent with previous research,<sup>2,4,8,13</sup> no proxy effect was found for conditions perceived to be more serious, or those less likely to be mislabelled. In this analysis, no differences were found for diabetes, epilepsy, heart disease, cancer, effects of a stroke or glaucoma.

### Who provides the information?

The characteristics of the people who provided information on behalf of others (proxy reporters) differed based on the age and sex of the individual for whom the information was being provided (Table 5). For both males and females aged 12 to 24, the proxy reporter was most often a parent (86% and 84%, respectively). And when the proxy reporter was a parent, it was usually the mother (78%) (data not shown).

Proxy reporters for individuals aged 25 or older were typically spouses. For men, 83% of proxy

reporters were their wives; for women, 66% of proxy reporters were their husbands. Largely as a result of this, 93% of proxy reports for men were provided by women, and 79% of reports for women were provided by men. Reports from an "other relative" were more common for females (25%) than for males (10%).

Some studies have suggested that women may be more inclined than men to report health conditions, both for themselves and for others.<sup>12,17,24,25</sup> To explore this issue, all the regression models for the various health conditions were rerun, taking into account the sex of the proxy reporter. Thus, three categories were considered for proxy reporting status: self-response; proxy response, female proxy reporter; and proxy response, male proxy reporter. The third category was used as the reference.

Compared with individuals whose information was reported by men, those whose information was reported by women had increased odds of reporting the following conditions: food allergies, other allergies, back problems (excluding arthritis), high blood pressure, chronic bronchitis or emphysema, sinusitis, effects of a stroke, urinary incontinence, cataracts, activity restriction due to a long-term physical or mental health problem, long-term disability or handicap, and "at least one chronic

Table 5  
Population with proxy response status in general component, by characteristics of proxy reporter, Health file, 1996/97 National Population Health Survey, household population aged 12 or older, Canada excluding territories

	Total	Sex of proxy reporter		Proxy reporter's relationship to individual with proxy response			
		Male	Female	Spouse	Parent	Other relative	Non-relative
	'000	%		%			
<b>All individuals with proxy response</b>	<b>6,757</b>	<b>27</b>	<b>73</b>	<b>52</b>	<b>33</b>	<b>13</b>	<b>2</b>
<b>Sex and age</b>							
Males 12-24 with proxy response	1,247	25	75	3 <sup>†</sup>	86	9	2
Females 12-24 with proxy response	1,054	25	75	3	84	11	2
Males 25+ with proxy response	3,170	7	93 <sup>‡‡</sup>	83 <sup>‡‡</sup>	5	10	2
Females 25+ with proxy response	1,286	79 <sup>‡</sup>	21	66	6	25 <sup>‡</sup>	3 <sup>‡</sup>

**Data source:** National Population Health Survey, cross-sectional sample, Health file, 1996/97

**Note:** Unweighted sample counts appear in Appendix Table C.

<sup>†</sup> Coefficient of variation between 16.6% and 25.0%

<sup>‡</sup> Significantly higher than value for males aged 25+ ( $p \leq 0.05$ )

<sup>‡‡</sup> Significantly higher than value for females aged 25+ ( $p \leq 0.05$ )

### Proxy reporting for the health component

Proxy reporting for the health component of the National Population Health Survey was strongly discouraged because of the detailed, subjective and personal nature of the information. Thus, the extent of proxy reporting for this component on the 1996/97 cross-sectional Health file was very low (2.6%). Unlike the general component, the percentage of proxy responses for the health component was highest for the oldest age groups: 7% for people aged 75 to 84 and 17% for those aged 85 or older. The elderly were more likely to have "special circumstances" due to a physical or mental health problem that made it impossible for them to answer on their own behalf. Proxy reporting was also high (9%) for children aged 12 to 14. Some parents objected to having their children interviewed. In such cases, interviewers did not press for an interview with the children and accepted a parent's proxy response.

#### Percentage of proxy responses to health component, 1996/97 National Population Health Survey, by sex and age, household population aged 12 or older, Canada excluding territories

	Population '000	Proxy response %	Unweighted sample counts of records with proxy response
<b>Total</b>	<b>24,595</b>	<b>2.6</b>	<b>1,707</b>
<b>Sex</b>			
Male	12,099	3.4 <sup>†</sup>	1,094
Female	12,495	1.8	613
<b>Age group</b>			
12-14	1,151	8.8 <sup>‡</sup>	139
15-17	1,284	2.6 <sup>§</sup>	71
18-44	12,408	1.4	479
45-64	6,335	2.1	352
65-74	2,096	3.8 <sup>††</sup>	279
75-84	1,060	7.2 <sup>‡</sup>	263
85+	261	17.2 <sup>‡‡</sup>	124

**Data source:** National Population Health Survey, cross-sectional sample, Health file, 1996/97

**Note:** Because of rounding, detail may not add to totals. Bonferroni significance tests were used to adjust for multiple comparisons.

<sup>†</sup> Significantly higher than percentage for female ( $p \leq 0.05$ )

<sup>‡</sup> Significantly higher than each age group, 15-to-74 range ( $p \leq 0.05$ )

<sup>§</sup> Coefficient of variation between 16.6% and 25.0%

<sup>††</sup> Significantly higher than 18-44 and 45-64 age groups ( $p \leq 0.05$ )

<sup>‡‡</sup> Significantly higher than each younger age group, 15-to-84 range ( $p \leq 0.05$ )

condition" (Table 6). Again, these associations were found when controlling for all of the variables believed to be related to an individual's proxy response status.

Research consistently finds higher morbidity rates for women compared with men.<sup>26</sup> While various factors have been suggested as potential explanations for these differences (for example, women are more likely to go to a doctor and are therefore more likely to be diagnosed), the higher rates for women may be partly an artifact of proxy reporting, since women are more likely than men to self-report, and self-reporters tend to report more health conditions.<sup>1,8,26</sup> But with NPHS data, it is difficult to assess whether the proxy effect would be more pronounced for males or for females. The proxy reporting rate for males was close to double that for females (37% compared with 19%; Table 1), suggesting that a proxy effect (of lowering estimates) would have more impact on the estimates for males. However, proxy reports for females tended to be provided by a male proxy reporter (Table 5), and male proxy reporters have decreased odds of reporting morbidity, which might yield lower estimates of morbidity for females.

#### Proxy reporting, longitudinal Health file

In some cases, proxy reporting status changed across cycles; for example, data were provided by proxy in one cycle, and in the other cycle, the longitudinal panel member provided his or her own information.

Close to three-quarters (73%) of females, but just 42% of males, in the longitudinal sample self-reported for the general component in both cycles (Table 7). As well, males were more likely than females to move from self- to proxy status (9% versus 4%) between 1994/95 and 1996/97, or to have their information reported by proxy in both cycles (18% compared with 5%).

The longitudinal file was used to explore the possibility of a proxy effect in incidence measures of selected health conditions. For this analysis, incidence is defined as a new case reported in the two years between NPHS cycles. (It should be noted that, in some cases, interviewer probing indicated that the condition may have existed before the first

Table 6  
Adjusted odds ratios relating chronic and other health conditions to proxy reporting status and sex of proxy reporter in general component, Health file, 1996/97 National Population Health Survey, population aged 12 or older, Canada excluding territories

	Proxy response, female proxy reporter†	
	Odds ratio	95% confidence interval
<b>Chronic conditions‡</b>		
Food allergies	1.5**	1.2, 1.9
Other allergies	1.5****	1.3, 1.8
Asthma	1.2	1.0, 1.6
Arthritis or rheumatism	1.2	0.9, 1.6
Back problems (excluding arthritis)	1.4***	1.1, 1.7
High blood pressure	1.4**	1.1, 1.7
Migraine	0.9	0.6, 1.3
Chronic bronchitis or emphysema	1.8*	1.1, 3.0
Sinusitis	1.7**	1.2, 2.6
Diabetes	1.1	0.8, 1.6
Epilepsy	0.5	0.2, 1.4
Heart disease	1.3	0.8, 2.1
Cancer	1.0	0.5, 1.9
Stomach or intestinal ulcers	1.2	0.8, 1.8
Effects of stroke	1.9*	1.1, 3.2
Urinary incontinence	1.7*	1.1, 2.6
Bowel disorder	1.1	0.7, 1.7
Cataracts	1.6*	1.1, 2.5
Glaucoma	0.9	0.5, 1.8
Thyroid disorder	1.5	0.9, 2.4
Other§	1.2	0.9, 1.6
At least one chronic condition‡	1.4****	1.2, 1.6
<b>Other health conditions</b>		
Activity restriction due to long-term physical or mental health problem	1.3**	1.1, 1.7
Long-term disability or handicap	1.3*	1.1, 1.7

**Data source:** National Population Health Survey, cross-sectional sample, Health file, 1996/97

**Notes:** Presents results of the 24 separate regression models based on 73,402 records on the Health file. Sample counts of the number of records for which each condition was reported appear in Appendix Table B. In each case, a small percentage (< 0.5%) of records were dropped because of missing values. Each regression includes selected control variables, but only the odds ratios for proxy reporting status are presented. (See Appendix Table F for a complete list of independent variables.) Results for complete models are available on request.

† Reference category is proxy response, male proxy reporter; self-respondents were included in the model, but their odds ratios are not shown.

‡ Alzheimer's disease was not included since the proxy reporting rate for this condition was very high.

§ Any other long-term chronic condition; for example, cystic fibrosis or multiple sclerosis

\*  $p \leq 0.05$

\*\*  $p \leq 0.01$

\*\*\*  $p \leq 0.001$

\*\*\*\*  $p \leq 0.0001$

cycle [see *Changes in reporting status and inconsistencies*]). Again, multiple logistic regression was used to examine the possible associations of a new report of a health condition and proxy reporting status, controlling for other variables related to proxy reporting status and changes in proxy reporting status (see Appendix Table G for the variables that were included in the models). For the three chronic conditions with sufficiently large sample size of new cases (other allergies, arthritis, non-arthritic back problems) and “at least one chronic condition,” individuals who had their information reported by proxy in 1994/95 and then reported their own information in 1996/97 had higher odds of reporting a new case of the condition, compared with those who had proxy response status in both cycles (Table 8). For arthritis and back problems (excluding arthritis), those who self-reported in both cycles also had increased odds of reporting a new case between 1994/95 and 1996/97, compared with individuals whose information was reported by proxy in both cycles.

New reports of activity restriction and disability were not significantly related to changes in reporting

Table 7  
Reporting status in general component, Health file, 1994/95 and 1996/97 National Population Health Survey, by sex, household population aged 12 or older, Canada excluding territories

	Total	Males	Females
		'000	
<b>Total population†</b>	<b>23,519</b>	<b>11,568</b>	<b>11,951</b>
		%	
<b>Reporting status</b>			
Proxy response 1994/95 and 1996/97	11.7	18.4‡	5.1
Proxy response 1994/95 to self-response 1996/97	24.2	31.1‡	17.6
Self-response 1994/95 to proxy response 1996/97	6.4	8.8‡	4.1
Self-response 1994/95 and 1996/97	57.7	41.8	73.2§

**Data source:** National Population Health Survey, longitudinal sample, Health file, 1994/95 and 1996/97

**Notes:** Because of rounding, percentages may not add to 100%. Bonferroni significance tests were used to adjust for multiple comparisons. Unweighted sample counts appear in Appendix Table D.

† Based on respondents aged 12 or older in 1994/95 who were still alive and living in households in 1996/97.

‡ Significantly higher than value for females ( $p \leq 0.05$ )

§ Significantly higher than value for males ( $p \leq 0.05$ )

## Changes in reporting status and inconsistencies

Longitudinal surveys look at the same individuals repeatedly to identify changes in their characteristics over several years. However, if a change is reported for an individual at some point, it is important to know if a “true” change has taken place (a new diagnosis of arthritis, for example) or if an inconsistency has arisen because of a reporting error or some other methodological problem.

The individuals selected for the longitudinal sample of the National Population Health Survey (NPHS) will be interviewed every 2 years for up to 20 years, and inconsistencies—changes that are not “true” changes—may occur. In cycles 1 and 2 (1994/95 and 1996/97), questions about chronic conditions and activity limitations were asked in the general component, where proxy responses were permitted. This may have compounded any problems arising from inconsistent reports.

To help minimize inconsistencies, cycle 1 data were incorporated into the computer-assisted questionnaire for cycle 2. Thus, when a change was reported between cycles for conditions that typically do not change over a two-year period (asthma, arthritis, high blood pressure, migraine, diabetes, epilepsy, ulcers, the effects of a stroke, and activity limitation), interviewers were automatically prompted to ask the reporting individual if there had indeed been a change, to give the reasons for the change, and to verify relevant dates. For example, if a chronic condition was reported for a respondent in cycle 2, but had not been reported in cycle 1, a follow-up question was posed to verify the date of diagnosis. If this date preceded the cycle 1 interview, there was an additional probe to determine if the respondent actually had the condition before the cycle 1 interview. This was the most frequent type of inconsistency on the longitudinal file. A total of 916 such inconsistencies arose for the eight chronic conditions for which probing was conducted. One or more such inconsistencies were reported for 6.3% of the longitudinal panel members (846 individuals). Inconsistencies were most common for reports of arthritis, high blood pressure and migraine (data not shown). Not surprisingly, inconsistencies were more frequent when the reporting status was “proxy response 1994/95 to self-response 1996/97.” But inconsistencies were also relatively frequent for the “self-response 1994/95 and 1996/97” group.

Probing was omitted if it would have compromised the confidentiality of an individual’s medical history. If, for example, a particular chronic condition such as diabetes had been self-reported in 1994/95, but a proxy reporter did not report the same condition in 1996/97, interviewers did not probe.

Reports may be inconsistent for several reasons. The individual reporting the information may have deliberately not reported the condition in 1994/95, or the reporter may have misunderstood the question during one of the interviews (1994/95 or 1996/97). An inconsistency could also have been introduced by a proxy reporter who was unaware of the condition, or by an interviewer who made a recording error. As well, there may have been a recall error about the date of diagnosis.

A possible explanation for the elevated inconsistency rate for the “proxy response 1994/95 to self-response 1996/97” group may be the “proxy effect” discussed in this article. If proxy reporters are less likely to report health events, it would not be surprising if 1996/97 self-reporters indicated that certain conditions had existed in 1994/95, but had not been reported by proxy.

The elevated rate for the “self-response 1994/95 and 1996/97” group is more puzzling. Inconsistencies for this group may involve an inaccurately reported date of diagnosis. For example, in some cases, self-reporters may remember the date when the symptoms first appeared, as it seems more important to them than the date when the condition was formally diagnosed by a health professional.

### Percentage of individuals by reporting status, for whom at least one inconsistency was detected in cycle 2,<sup>†</sup> general component, 1994/95 and 1996/97 National Population Health Survey, household population aged 12 or older, Canada excluding territories

Reporting status	Reports with at least one inconsistency <sup>†</sup>	Unweighted sample count
	%	
<b>Total population</b>	<b>6.3</b>	<b>846</b>
Proxy response 1994/95 and 1996/97	3.3 <sup>‡</sup>	41
Proxy response 1994/95 to self-response 1996/97	7.5*	178
Self-response 1994/95 to proxy response 1996/97	2.8 <sup>‡</sup>	29
Self-response 1994/95 and 1996/97	6.8*	598

**Data source:** National Population Survey, longitudinal sample, Health file, 1994/95 and 1996/97

**Note:** Bonferroni significance tests were used to adjust for multiple comparisons.

<sup>†</sup> Probing in 1996/97 revealed that a chronic condition had existed before cycle 1, but had not been reported in 1994/95.

<sup>‡</sup> Coefficient of variation between 16.6% and 25.0%

\*Significantly higher than value for “proxy response 1994/95 and 1996/97” and “self-response 1994/95 to proxy response 1996/97” ( $p \leq 0.05$ )

Table 8

Adjusted odds ratios relating incident cases of selected health conditions to reporting status in general component between 1994/95 and 1996/97, Health file, National Population Health Survey, population aged 12 or older, Canada excluding territories

	Proxy response 1994/95 to self-response 1996/97 <sup>†</sup>		Self-response 1994/95 to proxy response 1996/97 <sup>†</sup>		Self-response 1994/95 and 1996/97 <sup>†</sup>	
	Odds ratio	95% confidence interval	Odds ratio	95% confidence interval	Odds ratio	95% confidence interval
<b>Chronic conditions</b>						
Other allergies	2.0****	1.4, 2.7	0.9	0.6, 1.5	1.4	1.0, 1.9
Arthritis	2.6**	1.4, 4.6	1.2	0.6, 2.5	2.4**	1.4, 4.2
Back problems (excluding arthritis) <sup>‡</sup>	2.0***	1.4, 2.8	1.2	0.7, 2.1	1.7**	1.2, 2.4
At least one new chronic condition <sup>‡</sup>	2.0****	1.6, 2.5	1.0	0.8, 1.4	1.4**	1.1, 1.7
<b>Other health conditions</b>						
At least one new activity restriction due to long-term physical or mental health problem	1.3	0.9, 1.8	0.9	0.5, 1.4	1.0	0.7, 1.4
Long-term disability or handicap	1.5	1.0, 2.3	1.1	0.6, 2.0	1.3	0.9, 2.1

**Data source:** National Population Health Survey, longitudinal sample, Health file, 1994/95 and 1996/97

**Notes:** Presents the results of 6 separate regression models based on 13,427 records on the longitudinal Health file. In each case, a small percentage of records (< 2%) were dropped because of missing values. Each regression includes selected control variables, but only the odds ratios for proxy reporting status are presented. Unweighted sample counts appear in Appendix Table E. (See Appendix Table G for a complete list of the independent variables.) Results for complete models are available on request.

<sup>†</sup> Reference category is proxy response 1994/95 and 1996/97

<sup>‡</sup> Includes food allergies, other allergies, asthma, arthritis, back problems (excluding arthritis), high blood pressure, migraine, chronic bronchitis or emphysema, sinusitis, diabetes, epilepsy, heart disease, cancer, stomach or intestinal ulcers, effects of a stroke, urinary incontinence, cataracts and glaucoma.

\*  $p \leq 0.05$

\*\*  $p \leq 0.01$

\*\*\*  $p \leq 0.001$

\*\*\*\*  $p \leq 0.0001$

status. Perhaps new cases of such conditions have more immediate consequences for proxy reporters, who are therefore more likely to recall and report them.

### Concluding remarks

This analysis of the first two cycles of the National Population Health Survey (1994/95 and 1996/97) indicates that accepting proxy responses for certain health conditions may have resulted in underestimates of prevalence rates. A comparison of estimates from the Health and General files reveals that the Health file yields significantly higher estimates for many conditions, suggesting a proxy effect. However, as expected, estimates were closer for conditions that proxy reporters would be more

likely to notice and less likely to mislabel: diabetes, epilepsy, heart disease and cancer, for example. Results of a multivariate analysis of the Health file revealed that self-response was associated with the reporting of several health conditions when factors related to an individual's response status were controlled. On the longitudinal file, an association between new reports of certain conditions and changes in reporting status was also found.

Interviewers are now instructed to attempt completion of the general component directly with the longitudinal panel member to avoid proxy response. This should reduce the possibility that changes in reporting status across cycles will affect incidence estimates. ●

## References

- 1 Moore JC. Self/Proxy response status and survey response quality: A review of the literature. *Journal of Official Statistics* 1988; 4(2): 155-72.
- 2 Cannell CF, Marquis KH, Laurent A. A summary of studies of interviewing methodology. *Vital and Health Statistics* 1977; 2(69): 1-16.
- 3 Cartwright A. The effect of obtaining information from different informants on a family morbidity inquiry. *Applied Statistics* 1957; 6(1): 18-25.
- 4 Clarridge BR, Massagli MP. The use of female spouse proxies in common symptom reporting. *Medical Care* 1989; 27(4): 352-66.
- 5 Grootendorst PV, Feeny DH, Furlong W. Does it matter whom and how you ask? Inter- and intra-rater agreement in the Ontario Health Survey. *Journal of Clinical Epidemiology* 1997; 50(2): 127-35.
- 6 Groves RM, Mathiowetz NA. An experiment to measure the effects of respondent rules on health survey responses. *American Statistical Association: Proceedings of the Section on Survey Research Methods*. Washington, DC: American Statistical Association, 1981: 346-51.
- 7 Kovar MG, Wright RA. An experiment with alternate respondent rules in the National Health Interview Survey. *American Statistical Association: Proceedings of the Social Statistics Section*. Washington, DC: American Statistical Association, 1973: 311-6.
- 8 Miller RE, Massagli MP, Clarridge BR. Quality of proxy vs. self reports: evidence from a health survey with repeated measures. *American Statistical Association: Proceedings of the Section on Survey Research Methods*. Washington, DC: American Statistical Association, 1986: 546-51.
- 9 Mosley RR II, Wolinsky FD. The use of proxies in health surveys: Substantive and policy implications. *Medical Care* 1986; 24(6): 496-510.
- 10 Tompkins L, Massey JT. Using a most knowledgeable respondent rule in a household telephone survey. *American Statistical Association: Proceedings of the Section on Survey Research Methods*. Washington, DC: American Statistical Association, 1986: 281-6.
- 11 Koons DA. Quality control and measurement of nonsampling error in the Health Interview Survey. *Vital and Health Statistics* 1973; 2(54): 1-24.
- 12 National Center for Health Statistics. Reporting of hospitalization in the Health Interview Survey. *Vital and Health Statistics* 1965; 2(6): 1-51.
- 13 National Center for Health Statistics. Health Interview responses compared with medical records. *Vital and Health Statistics* 1965; 2(7): 1-40.
- 14 Briscoe ME. Research note: proxy responses in health surveys: a methodological issue. *Sociology of Health and Illness* 1984; 6(3): 359-65.
- 15 Cannell CC, Groves RM, Miller PV. The effects of mode of data collection on health survey data. *American Statistical Association: Proceedings of the Social Statistics Section*. Washington, DC: American Statistical Association, 1986: 1-6.
- 16 Heliövaara M, Aromaa A, Klaukka T, et. al. Reliability and validity of interview data on chronic diseases: The Mini-Finland Health Survey. *Journal of Clinical Epidemiology* 1993; 46(2): 181-91.
- 17 Madow WC. Net differences in interview data on chronic conditions and information derived from medical records. *Vital and Health Statistics* 1973; 2(57): 1-25.
- 18 Tambay J-L, Catlin G. Sample design of the National Population Health Survey. *Health Reports* (Statistics Canada, Catalogue 82-003) 1995; 7(1): 29-38.
- 19 Swain L, Catlin G, Beaudet MP. The National Population Health Survey—its longitudinal nature. *Health Reports* (Statistics Canada, Catalogue 82-003) 1999; 10(4): 69-80.
- 20 National Population Health Survey, 1996/97. *Household Component, User's Guide for the Public Use Microdata Files* (Statistics Canada, Catalogue 82-M0009GPE) Ottawa: Statistics Canada, 1998.
- 21 Rao JNK, Wu CFJ, Yue K. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 1992; 18(2): 209-17.
- 22 Rust KF, Rao JNK. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 1996; 5: 283-310.
- 23 Yeo D, Mantel H, Liu TP. Bootstrap variance estimation for the National Population Health Survey. *American Statistical Association: Proceedings of the Survey Research Methods Section*. Baltimore, August 1999.
- 24 Verbrugge LM. Females and illness: recent trends in sex differences in the United States. *Journal of Health and Social Behavior* 1976; 17: 387-403.
- 25 Waldron I. Sex differences in illness incidence, prognosis and mortality: issues and evidence. *Social Science and Medicine* 1983; 17(16): 1107-23.
- 26 Clarke JN. Sexism, feminism and medicalism: a decade of literature on gender and illness. *Sociology of Health and Illness* 1983; 5(1): 62-82.

## Appendix

Table A

**Unweighted sample counts of individuals aged 12 or older with proxy response to general component, by sex and age, General and Health files, 1996/97 National Population Health Survey**

	General file		Health file	
	Total	Proxy response	Total	Proxy response
<b>Total</b>	<b>173,216</b>	<b>78,158</b>	<b>73,402</b>	<b>20,469</b>
<b>Sex</b>				
Male	84,409	48,118	34,265	13,199
Female	88,807	30,040	39,137	7,270
<b>Age group</b>				
12-17	19,836	16,537	5,120	3,541
18-24	19,955	10,747	7,000	2,185
25-44	66,401	25,549	28,900	7,319
45-64	44,762	18,310	19,019	4,981
65+	22,262	7,015	13,363	2,443

*Data source: National Population Health Survey, cross-sectional sample, General and Health files, 1996/97*

Table B

**Unweighted sample counts of individuals aged 12 or older reporting chronic and other health conditions, by proxy response to the general component, General and Health files, 1996/97 National Population Health Survey**

	General file		Health file	
	Total	Proxy response	Total	Proxy response
<b>Total</b>	<b>173,216</b>	<b>78,158</b>	<b>73,402</b>	<b>20,469</b>
<b>Chronic conditions</b>				
Food allergies	11,257	3,826	5,335	1,035
Other allergies	34,708	11,906	16,221	3,234
Asthma	12,347	5,129	5,467	1,354
Arthritis or rheumatism	23,332	6,101	13,063	2,005
Back problems (excluding arthritis)	24,327	7,892	12,097	2,474
High blood pressure	16,336	5,303	8,676	1,731
Migraine	12,442	3,725	5,804	1,059
Chronic bronchitis or emphysema	4,471	1,265	2,429	410
Sinusitis	7,521	1,977	3,788	593
Diabetes	5,380	2,013	2,706	623
Epilepsy	956	380	446	99
Heart disease	6,765	2,345	3,695	772
Cancer	2,563	850	1,359	289
Stomach or intestinal ulcers	4,348	1,371	2,245	422
Effects of stroke	1,533	612	868	231
Urinary incontinence	2,669	694	1,596	246
Bowel disorder	2,862	767	1,520	245
Alzheimer's disease/Other dementia <sup>†</sup>	494	300	245	128
Cataracts <sup>†</sup>	4,250	1,113	2,679	406
Glaucoma <sup>†</sup>	1,736	519	1,013	172
Thyroid disorder	5,422	1,249	2,852	356
Other <sup>‡</sup>	9,171	3,103	4,597	973
At least one chronic condition	94,550	35,261	44,535	10,121
<b>Other health conditions</b>				
Activity restriction due to long-term physical or mental health problem	21,192	7,217	10,982	2,313
Long-term disability or handicap	18,106	6,539	9,446	2,111

*Data source: National Population Health Survey, cross-sectional sample, General and Health files, 1996/97*

<sup>†</sup> Asked only for population aged 18 or older

<sup>‡</sup> Any other long-term chronic condition; for example, cystic fibrosis or multiple sclerosis

Table C

**Unweighted sample counts of individuals aged 12 or older with proxy response to general component, by characteristics of proxy reporter, Health file, 1996/97 National Population Health Survey**

	Sex of proxy reporter				Proxy reporter's relationship to individual with proxy response				
	Total	Male	Female	Missing	Spouse	Parent	Other relative	Non-relative	Missing
<b>All individuals with proxy response</b>	<b>20,469</b>	<b>5,945</b>	<b>14,522</b>	<b>2</b>	<b>12,525</b>	<b>5,338</b>	<b>2,084</b>	<b>519</b>	<b>3</b>
<b>Sex and age</b>									
Males 12-24 with proxy response	3,035	759	2,276	0	181	2,490	264	100	0
Females 12-24 with proxy response	2,691	775	1,916	0	162	2,202	237	90	0
Males 25+ with proxy response	10,164	552	9,612	0	8,766	443	758	197	0
Females 25+ with proxy response	4,579	3,859	718	2	3,416	203	825	132	3

*Data source: National Population Health Survey, cross-sectional sample, Health file, 1996/97*

Table D

**Unweighted sample counts of individuals aged 12 or older with proxy response to general component, by sex, longitudinal Health file, 1994/95 and 1996/97 National Population Health Survey**

	Total	Males	Females
<b>Total</b>	<b>13,427</b>	<b>6,071</b>	<b>7,356</b>
Proxy response 1994/95 and 1996/97	1,314	1,036	278
Proxy response 1994/95 to self-response 1996/97	2,603	1,628	975
Self-response 1994/95 to proxy response 1996/97	768	527	241
Self-response 1994/95 and 1996/97	8,742	2,880	5,862

*Data source: National Population Health Survey, longitudinal sample, Health file, 1994/95 and 1996/97*

Table E

**Unweighted sample counts of individuals aged 12 or older reporting new cases of selected health conditions between 1994/95 and 1996/97, by reporting status, general component, Health file, National Population Health Survey**

	Total	Proxy response 1994/95 and 1996/97	Proxy response 1994/95 to self-response 1996/97	Self-response 1994/95 to proxy response 1996/97	Self-response 1994/95 and 1996/97
<b>Total</b>	<b>13,427</b>	<b>1,314</b>	<b>2,603</b>	<b>768</b>	<b>8,742</b>
<b>Chronic conditions</b>					
Other allergies	1,453	88	349	52	964
Arthritis	766	34	135	32	565
Back problems (excluding arthritis)	1,027	73	235	47	672
At least one chronic condition <sup>†</sup>	4,854	337	982	223	3,312
<b>Other health conditions</b>					
Activity limitation due to long-term physical or mental health problem	1,298	107	240	61	890
Long-term disability or handicap	786	54	155	43	534

*Data source: National Population Health Survey, longitudinal sample, Health file, 1994/95 and 1996/97*

<sup>†</sup> Includes food allergies, other allergies, asthma, arthritis or rheumatism, back problems (excluding arthritis), high blood pressure, migraine, chronic bronchitis or emphysema, sinusitis, diabetes, epilepsy, heart disease, cancer, stomach or intestinal ulcers, effects of a stroke, urinary incontinence, cataracts and glaucoma.

Table F  
Variables included in multiple logistic regression models relating selected health characteristics to reporting status in 1996/97

**Reporting status**  
Self-respondent  
Proxy respondent†

**Sex**  
Male‡  
Female

**Age§**

**Other adult(s) aged 18 or older in household††**

**Marital status**  
Married‡  
Never married  
Previously married

**Education**  
High school graduation or less  
Some postsecondary  
Postsecondary (college, trade or university) graduation‡

**Employed††**

**Household income**  
Lowest/Lower-middle‡  
Middle  
Upper-middle  
Highest

**Self-perceived health**  
Excellent‡  
Very good  
Good  
Fair  
Poor

**Data source:** 1996/97 National Population Health Survey, cross-sectional sample, Health file

†In the first set of regression models (Table 4), proxy response was used as the reference category. In the second set (Table 6), this category was split into two: proxy response, female proxy reporter; proxy response, male proxy reporter. In this second set, proxy response, male proxy reporter was used as the reference category.

‡ Reference category

§ Age was entered into the regression models as a continuous variable.

††Reference category is absence of the characteristic; for example, the reference category for "other adult(s) aged 18 or older in household" is "no other adult(s) aged 18 or older in household."

Table G  
Variables included in multiple logistic regression models relating new cases of selected health conditions to changes in reporting status between 1994/95 and 1996/97

**Reporting status**  
Proxy respondent 1994/95 to self respondent 1996/97  
Self respondent 1994/95 to proxy respondent 1996/97  
Self respondent 1994/95 and 1996/97  
Proxy respondent 1994/95 and 1996/97†

**Sex**  
Male‡  
Female

**Age in 1994/95‡**

**Other adult(s) aged 18 or older in household 1994/95§**  
**Other adult(s) aged 18 or older in household 1994/95 to no other adult(s) aged 18 or older in household 1996/97§**  
**No other adult(s) aged 18 or older in household in 1994/95 to other adult(s) aged 18 or older in household 1996/97§**

**Marital status 1994/95**  
Married‡  
Never married  
Previously married

**Not married 1994/95 to married 1996/97§**  
**Married 1994/95 to not married 1996/97§**

**Education 1994/95**  
High school graduation or less  
Some postsecondary  
Postsecondary (college, trade or university) graduation‡

**Employed 1994/95§**  
**Employed 1994/95 to not employed 1996/97§**  
**Not employed 1994/95 to employed 1996/97§**

**Household income 1994/95**  
Lowest/Lower-middle‡  
Middle  
Upper-middle  
Highest

**Self-perceived health 1994/95**  
Excellent‡  
Very good  
Good  
Fair  
Poor

**Improvement in self-perceived health between 1994/95 and 1996/97§**  
**Decline in self-perceived health between 1994/95 and 1996/97§**

**Data source:** 1994/95 and 1996/97 National Population Health Survey, longitudinal sample, Health file

† Reference category

‡ Age was entered into the regression models as a continuous variable.

§ Reference category is absence of the characteristic; for example, the reference category for "other adult(s) aged 18 or older in household" is "no other adult(s) aged 18 or older in household."