

# Proxy reporting of health information

*Margot Shields*

## Abstract

### *Objectives*

This article documents the extent of proxy reporting in Statistics Canada's National Population Health Survey (NPHS) and explores associations between reporting status and the prevalence and incidence of selected health problems.

### *Data sources*

Data are from the household cross-sectional (1994/95, 1996/97 and 1998/99) and longitudinal (1994/95 to 2000/01) components of the NPHS. Supplemental data are from the 2000/01 Canadian Community Health Survey.

### *Analytical techniques*

Estimates of health conditions from the two cross-sectional files that are produced for each NPHS cycle were compared. The file with the lower proxy reporting rate was expected to yield higher prevalence rates. Multivariate analyses of the longitudinal data were used to examine associations between changes in reporting status and the incidence of the selected conditions.

### *Main results*

Compared with the 1998/99 General file, in which proxy reporting was more common, the 1998/99 Health file yields higher estimates of certain health conditions. Declines in proxy reporting rates over time are generally associated with greater increases in estimates. Analyses based on the longitudinal file suggest that the incidence of some conditions may also be subject to a proxy effect.

## Key words

chronic conditions, people with disabilities, health surveys, longitudinal studies

## Author

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

An important decision that must be made in the design and implementation of a health survey is whether to accept proxy responses. While common sense suggests that it is best to question people directly about their health, many surveys allow one knowledgeable respondent to answer on behalf of others.

The reasons for accepting proxy responses fall into two categories: necessity and convenience. "Proxy by necessity" refers to situations in which individuals selected to be interviewed are unable to respond on their own behalf because of physical or mental conditions. Excluding such people from a health survey would bias estimates. "Proxy by convenience" refers to the acceptance of proxy responses for people capable of providing their own information. If information about everyone in a household is collected from one person, it is possible to obtain a large sample size with just one contact, thereby improving response rates and reducing costs. Eliminating the need for several call-backs to interview people who are difficult to reach also helps reduce costs.

## Methods

### Data sources

#### National Population Health Survey

Since 1994/95, Statistics Canada's biennial National Population Health Survey (NPHS) has collected information about household and institutional residents in all provinces and territories, except residents of Indian reserves, Canadian Forces bases, and some remote areas.<sup>1,2</sup> This analysis is based on the household population aged 18 or older in the 10 provinces.

For each of the first three cycles (1994/95, 1996/97 and 1998/99), two cross-sectional files were produced: General and Health. The General file contains socio-demographic and basic health information, collected using the General questionnaire, for every member of each participating household. The Health file contains in-depth health information, collected using the Health questionnaire, for one randomly selected member of each household, as well as the information about that person in the General file.

For each cycle, two cross-sectional response rates are calculated: household and person. The *household* response rate is based on the number of households for which at least the General questionnaire was completed for the randomly selected member. The *person* response rate is based on the number of responding households for which the Health questionnaire was completed for the randomly selected member.

	Response rate	
	Household	Person
	%	
1994/95	88.7	96.1
1996/97	82.6	95.6
1998/99	87.6	98.5

A longitudinal file is also produced. In 1994/95, a panel of 17,276 respondents (a subset of the randomly selected household members) was chosen to be followed over subsequent cycles. In cycle 4, the NPHS became strictly longitudinal, and the General and Health questionnaires were combined. This analysis uses the cycle 4 (2000/01) longitudinal "square" file, which contains records for all originally selected panel members about whom information is available in the cycle 1 General file, regardless of whether information about them was obtained in later cycles.

#### Canadian Community Health Survey

Arthritis prevalence rates for 2000/01 are from cycle 1.1 of Statistics Canada's Canadian Community Health Survey (CCHS). The CCHS covers the household population aged 12 or older in all provinces and territories, except residents of Indian reserves, Canadian Forces bases, and some remote areas.<sup>3</sup> Data collection began in September 2000 and continued over 14 months. The sample size was 131,535; the response rate was 84.7%. The CCHS data for this article were reported by 116,171 respondents aged 18 or older in the 10 provinces.

### Analytical techniques

Proxy reporting rates, based on weighted data, are presented for the cross-sectional files (General and Health) for the first three NPHS cycles, as well as for the cycle 4 longitudinal file.<sup>4,5</sup> Sample sizes and unweighted counts by proxy status for all files can be found in Appendix Tables A and B.

To investigate the possibility of a proxy effect on the prevalence of health conditions, estimates from the cross-sectional General and Health files were compared for cycles 1 and 3. It was hypothesized that because of its lower proxy response rate, the cycle 3 (1998/99) Health file would yield higher estimates than the General file. Estimates for the two files were expected to be closer for cycle 1 (1994/95) because of the similar proxy rates.

Multiple logistic regression models were used to explore the possibility of a proxy effect on the incidence of health problems. Each model examined the two-year incidence of selected conditions (new cases in a two-year period) in relation to changes in reporting status in the same period. Based on NPHS respondent selection rules and a review of the literature on proxy reporting, several control variables were included: sex, age, living arrangements, education, household income, employment status, and the Health Utilities Index. All regressions were run on the 1994/95 to 2000/01 longitudinal square file. An incident case was defined as a health problem reported in cycle 2, 3 or 4 from a respondent who had not reported the problem in the previous cycle. For every two-year interval (1994/95-to-1996/97, 1996/97-to-1998/99 and 1998/99-to-2000/01), a new record was created for each respondent who had not reported the condition in the previous cycle. Consequently, one respondent could contribute up to three records to the analyses for each health condition: one for every two-year interval. Approximately 30,000 records were used in each model (ranging from 27,204 to 34,995).

To account for the effects of survey design, the variance on prevalence and incidence rates, on differences between rates, and on odds ratios was calculated using the bootstrap technique.<sup>6-8</sup> In comparing rates across NPHS cycles, bootstrap weights that account for the overlap among samples were used.

### Limitations

It is generally assumed that differences between self- and proxy reports of health problems reflect under-reporting by proxy reporters. Comparisons with medical records have shown that under-reporting is more common in proxy reports than in self-reports.<sup>9-11</sup> However, while even self-respondents may under-report health and health care events,<sup>9-13</sup> in some cases, self-respondents may over-report. For example, the criteria for chronic conditions 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 interviewers specify these criteria, a respondent who, for instance, has been suffering 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- than proxy respondents, the effect would be over-reporting.

The models in the analysis that consider incident health conditions in relation to changes in reporting status controlled for socio-demographic and other confounders that might be related an individual's reporting status. The results could be misleading if important variables related to morbidity and to reporting status were inadvertently excluded, or are not available from the NPHS.

Although residents of long-term care institutions were interviewed in the NPHS, this analysis was not carried out for that population. Proxy responses were accepted for residents of health care institutions only by necessity, and therefore, it was not possible to assess potential biases related to proxy reporting.

Although reasons for accepting proxy responses are often compelling, the result may be lower estimates of some health problems. Proxy reporters tend to have less knowledge about another's health than that person him- or herself. Lower estimates of chronic conditions, disability, activity restriction, pain, medication use, physician visits, and hospitalization have been found for proxy reports, indicating a downward bias.<sup>10,11,14-22</sup> However, comparisons with administrative data such as medical records suggest that even self-respondents may under-report health events.<sup>9-13</sup>

Research on people who are elderly or frail suggests that such a bias may be in the opposite direction. That is, proxies responding for these groups may report more impairment than self-respondents.<sup>21,23-28</sup> Over-reporting is especially common when proxy responses are accepted for residents of health care institutions.

Based on cross-sectional and longitudinal data from Statistics Canada's National Population Health Survey (NPHS), this article investigates the possibility of a "proxy effect" (biased estimates) as a result of proxy reporting. The article focuses on the population aged 18 or older who were living in private households at the time of the interview (see *Methods* and *Definitions*).

## Guidelines

When the NPHS was first conducted in 1994/95, proxy responses by necessity and by convenience were both accepted, depending on the information sought.

Originally, the NPHS had two questionnaires: General and Health. The General questionnaire was used to collect socio-demographic information and basic health information (chronic conditions, long-term disability, two-week disability, and health care utilization) for every member of each household. Because the information was generally factual and objective, proxy by convenience was accepted.

As well, one member of each household was chosen at random to respond to the Health questionnaire, which covered topics such as smoking, physical activity, medication use, social support and mental health. Because the questions

were detailed and often personal, proxy responses were accepted only by necessity. Some highly subjective and personal questions were skipped if the responses were being provided by a proxy reporter.

These guidelines were in effect for the first two cycles (1994/95 and 1996/97). In cycle 3 (1998/99), a modification was introduced because of the longitudinal component—the subset of the randomly selected respondents who are followed over time. As a result of concern that variations in reporting status across cycles might confound measures of change, cycle 3 interviewers were instructed to collect information directly from longitudinal panel members for **both** the General and Health questionnaires. If longitudinal panel members could not be contacted throughout the entire collection period, interviewers could accept a proxy response for the General questionnaire, but for the Health questionnaire, it was proxy by necessity only. (For other household members, the previous rules still applied.)

In cycle 4 (2000/01), when the NPHS became strictly longitudinal, the General and Health questionnaires were combined, and proxy responses were accepted only by necessity.

## Proxy reporting rates

For each of the first three NPHS cycles, two cross-sectional files were created: General and Health. The General file has a record for every member of each responding household, and contains information collected with the General questionnaire. The Health file has a record for each randomly selected respondent (just one record per household), and contains the Health questionnaire information for that person along with information collected about him or her in the General questionnaire.

Because the information in the Health file comes from two questionnaires that have different guidelines for proxy reporting, records on this file can have a "mixed" reporting status. For example, if a married man was the selected respondent, he may have answered the Health questionnaire himself, but his wife may have provided answers on his

behalf for the General questionnaire. Therefore, two proxy reporting rates are calculated for the Health file: one for the Health questionnaire and one for the General questionnaire.

In the General file, the proxy reporting rates for the General questionnaire were fairly stable across survey cycles at approximately 35% (Table 1). By contrast, in the Health file, the proxy rate for the information derived from the General questionnaire file dropped substantially from 31% in 1994/95 to 15% in 1998/99. Proxy reporting rates for the Health questionnaire were under 5% for all three cycles.

Table 1  
Percentage of proxy responses to National Population Health Survey, cross-sectional files, by sex, household population aged 18 or older, Canada excluding territories, 1994/95, 1996/97 and 1998/99

	General questionnaire		Health questionnaire
	General File	Health File	Health File
	%	%	%
<b>Total</b>			
1994/95	36.0	31.1	4.9
1996/97	37.9	23.7	2.3
1998/99	34.5	14.6	2.2
<b>Men</b>			
1994/95	51.1	45.6	7.1
1996/97	50.5	33.4	3.0
1998/99	47.9	22.0	3.1
<b>Women</b>			
1994/95	21.4	17.2	2.8
1996/97	25.9	14.4	1.6
1998/99	21.6	7.6	1.4

Data sources: 1994/95 to 1998/99 National Population Health Survey, cross-sectional sample, General and Health files

### Effect on prevalence of health problems

With their different proxy reporting rates, the General and Health files offer an opportunity to investigate the “proxy effect.” Although the number of records on each file varies considerably, both files have been weighted so that they can be used to produce estimates for exactly the same populations. As well, each record in the Health file is automatically included in the General file, and the data for the

other records in the General file were collected using the same methods, at the same time, and by the same interviewers. However, the proxy reporting rates differ: in 1998/99, 35% for the General file and 15% for the Health file; in 1994/95, 36% and 31%, respectively. If people responding on their own behalf are more likely to report health problems, estimates for 1998/99 based on the Health file, with its low proxy rate, should be higher than those based on the General file. By contrast, estimates from the two files for 1994/95, when the proxy reporting rates were similar, might be expected to be closer.

And in fact, the 1994/95 Health and General file estimates of chronic conditions, long-term disability and two-week disability are very close: the only significant difference between the estimates is for the prevalence of non-food allergies (Table 2). The same is not true for 1998/99: as well as non-food allergies, the Health file estimates are higher for the prevalence of asthma, arthritis, back problems, migraine, stomach or intestinal ulcers, urinary incontinence, thyroid disorder, activity restriction, long-term disability, and cutting down on normal activities in the past two weeks because of illness or injury.

Previous studies have found that proxy reports are most accurate for conditions that are serious, painful, persistent or potentially life-threatening.<sup>9,15,16,19</sup> Consistent with such findings, the 1998/99 General and Health file estimates do not differ significantly for diabetes, epilepsy, heart disease, need for assistance with activities of daily living, and having spent at least a day in bed during the previous two weeks because of illness or injury.

### Changes in prevalence

Before the NPHS became strictly longitudinal, one reason for conducting it every two years was to monitor the prevalence of health conditions over time. However, apparent changes in prevalence might reflect variations in reporting status. As noted, the proxy reporting rate for the General file remained fairly stable, while the rate for the General questionnaire portion of the Health file fell from 31% in 1994/95 to 15% in 1998/99. If a proxy effect does exist, this reduction in the proportion

of proxy responses might result in sharper increases in the prevalence of some conditions in the Health file, particularly conditions that are less serious and/or noticeable. And indeed, data from the Health file show larger increases in non-food allergies, asthma, arthritis, migraine, urinary incontinence, and cut-down day(s) in the previous two weeks (Table 2). The prevalence of activity restriction and long-term disability decreased on both files, but the decrease was smaller for the Health file, again suggesting a proxy effect.

## Men and women

In 1998/99, the difference between the General and Health files in proxy reporting rates was greater for men than for women. For men, the proxy reporting rate for the General file was 26 percentage points higher than the rate for General questionnaire portion of the Health file; for women, the difference was 14 percentage points (Table 1). Therefore, the proxy effect on the prevalence of health conditions might be expected to be stronger for men. Nonetheless, results for men and women were

Table 2  
Estimates of chronic conditions and disability in General and Health files of National Population Health Survey, household population aged 18 or older, Canada excluding territories, 1994/95 and 1998/99

	1994/95		1998/99		Change between cycles
	General file	Health file	General file	Health file	
	%		%		
<b>Chronic conditions</b>					
Non-food allergies	16.5	17.2*	21.3	23.9*	ch
Arthritis	13.7	14.2	14.9	16.8*	ch
Back problems	14.6	15.1	14.3	15.0*	
High blood pressure	9.7	9.7	11.9	12.1	
Migraine	7.3	7.6	7.4	8.2*	ch
Asthma	5.6	5.7	7.0	7.7*	ch
Food allergies	5.4	5.4	6.6	6.8	
Heart disease	4.4	4.2	4.6	4.7	
Thyroid disorder	...	...	4.3	4.6*	...
Diabetes	3.4	3.4	3.9	3.8	
Cataract	2.7	2.7	3.5	3.5	
Stomach/Intestinal ulcers	3.5	3.6	2.7	3.1*	
Chronic bronchitis/Emphysema	3.0	3.3	2.5	2.7	
Urinary incontinence	1.1	1.2	2.0	2.3*	ch
Bowel disorder	...	...	1.7	1.8	...
Cancer	1.7	1.7	1.6	1.6	
Glaucoma	1.1	1.1	1.4	1.3	
Effects of stroke	0.9	1.0	1.1	1.2	
Epilepsy	0.6	0.7	0.6	0.7	
Alzheimer's disease/Other dementia	0.1	0.1 <sup>E1</sup>	0.3	0.3	
<b>Long-term</b>					
Activity restriction	17.3	17.4	14.1	15.0*	ch
Disability	15.4	15.8	12.4	13.6*	ch
Activity restriction or disability	21.1	21.5	17.4	18.9*	ch
Dependency in instrumental activities of daily living	9.1	9.2	11.9	12.3	
Dependency in activities of daily living	1.6	1.5	2.2	2.2	
<b>Two-week</b>					
At least one cut-down day	12.7	12.5	10.8	11.5*	ch
At least one bed-day	6.8	6.8	5.8	5.9	
<b>Proxy reporting rate to General questionnaire (%)</b>	36.0	31.1	34.5	14.6	
<b>Sample size (number of respondents)</b>	41,045	16,291	34,543	14,150	

*Data sources:* 1994/95 and 1998/99 National Population Health Survey, cross-sectional sample, General and Health files

\* Significantly higher than General file estimate for corresponding cycle ( $p < 0.05$ )

ch Change between 1994/95 and 1998/99 in Health file estimates significantly different from change in General file estimates ( $p < 0.05$ )

E1 Coefficient of variation 16.6% to 25.0%

... Not applicable

similar (Table 3). A possible explanation is that previous research has found women to be more accurate than men in reporting health events,<sup>10,14</sup> and men are more likely than women to have a female proxy reporter (see *Who provides the information?*).

Table 3  
Estimates of chronic conditions and disability, by sex, in General and Health files of National Population Health Survey, household population aged 18 or older, Canada excluding territories, 1998/99

	Men		Women	
	General file	Health file	General file	Health file
	%		%	
<b>Chronic conditions</b>				
Non-food allergies	17.5	19.9*	25.0	27.8*
Arthritis	10.5	12.5*	19.1	20.9*
Back problems	13.7	14.1	14.8	15.8*
High blood pressure	9.8	10.0	13.9	14.0
Migraine	3.3	3.9*	11.3	12.3*
Asthma	5.8	6.6*	8.1	8.7*
Food allergies	5.1	5.3	8.1	8.3
Heart disease	4.9	4.9	4.2	4.4
Thyroid disorder	1.4	1.7*	7.1	7.5
Diabetes	4.3	4.3	3.6	3.3
Cataract	2.6	2.9*	4.3	4.1
Stomach/Intestinal ulcers	2.5	2.9*	2.9	3.3*
Chronic bronchitis/Emphysema	2.2	2.4	2.8	3.1
Urinary incontinence	1.2	1.4*	2.7	3.0*
Bowel disorder	1.2	1.3	2.2	2.2
Cancer	1.4	1.4	1.7	1.8
Glaucoma	1.2	1.0	1.6	1.7
Effects of stroke	1.1	1.2	1.0	1.1 <sup>E1</sup>
Epilepsy	0.6	0.6 <sup>E1</sup>	0.6	0.8 <sup>E1</sup>
Alzheimer's disease/Other dementia	0.2 <sup>E1</sup>	0.3 <sup>E1</sup>	0.3 <sup>E1</sup>	0.3 <sup>E1</sup>
<b>Long-term</b>				
Activity restriction	12.7	13.7*	15.3	16.2*
Disability	12.2	13.5*	12.6	13.7*
Activity restriction or disability	16.2	17.9*	18.6	20.0*
Dependency in instrumental activities of daily living	8.4	8.9	15.3	15.6
Dependency in activities of daily living	1.8	2.0	2.5	2.5
<b>Two-week</b>				
At least one cut-down day	8.8	8.9	12.7	14.1*
At least one bed-day	4.7	4.6	6.9	7.2
<b>Proxy reporting rate to General questionnaire (%)</b>	47.9	22.0	21.6	7.6
<b>Sample size (number of respondents)</b>	16,519	6,446	18,024	7,704

*Data source:* 1998/99 National Population Health Survey, cross-sectional sample, General and Health files  
\* Significantly higher than General file estimate ( $p < 0.05$ )  
<sup>E1</sup> Coefficient of variation 16.6% to 25.0%

### Seniors

Some studies have suggested that the proxy effect for seniors may be in the opposite direction—that is, proxy reporters for elderly people are more likely to report impairment than seniors would themselves.<sup>21,23,24,26-28</sup> Estimates of the prevalence of health conditions based on NPHS data, however, do not support this finding. When there were significant differences between the 1998/99 General and Health files, the Health file estimates were higher (Table 4). The reason may be that the NPHS sample

### Who provides the information?

The characteristics of proxy reporters (the people providing information on behalf of others) for the General questionnaire of the National Population Health Survey were closely associated with the age and sex of proxy respondents (the people about whom information was provided).

Information about young proxy respondents (ages 18 to 24) tended to come from a parent: in about 60% of cases, their mother, and in an additional 15%, their father.

Information about proxy respondents aged 25 or older was typically provided by a spouse: for more than 80% of male respondents, their wife; for 64% of female respondents, their husband. For an additional 17% of female proxy respondents, information was provided by "other female relatives," usually a daughter (data not shown).

Distribution of proxy respondents to 1998/99 National Population Health Survey, General questionnaire, by age, sex and relationship to proxy reporter, household population aged 18 or older, Canada excluding territories

	Ages 18-24		Ages 25+	
	Men	Women	Men	Women
	%		%	
<b>Total</b>	100.0	100.0	100.0	100.0
Spouse	9.9	10.6	82.2	63.6
Mother	59.9	59.0	4.6	5.2
Father	15.3	14.5	1.2	1.3
Other relative				
Female	6.3	8.0	5.9	16.7
Male	5.4	3.2	4.2	10.4
Non-relative	3.3	4.7	2.0	2.9

*Data source:* 1998/99 National Population Health Survey, cross-sectional sample, General file  
*Note:* Because of rounding, columns may not add to 100.0%

pertains to the household population. Previous studies often concerned institutional residents whose proxy reporters are typically younger people who, viewing chronic conditions from their own perspective, may be more likely to report a senior as being impaired.<sup>23</sup> Proxy reporters for elderly NPHS respondents, however, tended to be of a similar age: close to three-quarters were that person's spouse (data not shown).

Table 4  
Estimates of chronic conditions and disability in General and Health files of National Population Health Survey, household population aged 65 or older, Canada excluding territories, 1998/99

	General file	Health file
	%	
<b>Chronic conditions</b>		
Arthritis	41.6	44.9*
High blood pressure	35.7	36.5
Cataract	17.6	18.1
Heart disease	17.5	17.3
Back problems	16.6	17.8*
Non-food allergies	15.4	15.9
Diabetes	11.7	11.5
Thyroid disorder	8.8	9.2
Urinary incontinence	7.0	7.5
Food allergies	6.2	6.1
Chronic bronchitis/Emphysema	5.8	5.9
Asthma	5.7	6.3
Glaucoma	5.4	5.1
Stomach/Intestinal ulcers	5.2	6.0*
Cancer	4.6	4.4
Effects of stroke	4.2	4.4
Migraine	3.3	3.8
Bowel disorder	3.0	3.2
Alzheimer's disease/Other dementia	1.1	1.3 <sup>E1</sup>
Epilepsy	0.7 <sup>E1</sup>	1.1 <sup>E2*</sup>
<b>Long-term</b>		
Activity restriction	29.2	29.8
Disability	25.0	26.6*
Activity restriction or disability	34.6	36.2*
Dependency in instrumental activities of daily living	37.2	38.2
Dependency in activities of daily living	6.9	7.2
<b>Two-week</b>		
At least one cut-down day	12.9	12.8
At least one bed-day	6.4	6.1
<b>Proxy reporting rate to General questionnaire (%)</b>	26.6	15.0
<b>Sample size (number of respondents)</b>	4,728	2,851

Data source: 1998/99 National Population Health Survey, cross-sectional sample, General and Health files

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

E1 Coefficient of variation 16.6% to 25.0%

E2 Coefficient of variation 25.1% to 33.3%

## Effect on incidence of health problems

For every NPHS cycle, a longitudinal file was created, which includes the General and Health questionnaire data for members of the longitudinal panel. The proxy reporting rate for the General questionnaire portion of the longitudinal file dropped from 33% in 1994/95 to 14% in 1998/99. In 2000/01, when the NPHS became strictly longitudinal and proxy responses were accepted only by necessity, the rate declined to 4% (Table 5).

Table 5  
Percentage of proxy responses to National Population Health Survey, longitudinal file, household population aged 18 or older in 1994/95, Canada excluding territories, 1994/95 to 2000/01

	General questionnaire	Health questionnaire
	%	
1994/95	32.7	4.8
1996/97	16.4	1.7
1998/99	13.5	2.2
2000/01	4.0	4.0

Data sources: 1994/95 to 2000/01 National Population Health Survey, longitudinal sample, Health file (square)

The longitudinal file is frequently used to produce two-year incidence rates of health conditions. However, changes in reporting status from one cycle to the next could affect these estimates.

To quantify changes in reporting status over time, a set of records was created for each two-year interval over the four NPHS cycles (1994/95-to-1996/97, 1996/97-to-1998/99, and 1998/99-to-2000/01). When these three sets of records were compiled, in the majority of cases, the reporting status for the General questionnaire did not change from one cycle to the next. In close to three-quarters of cases (73.5%), respondents provided their own information over two consecutive cycles (self-self), and for a small percentage of cases (6.1%), cycle-to-cycle information was provided by proxy (proxy-proxy) (Table 6). However, in about 5% of cases, reporting status changed from self to proxy, and in 15%, from proxy to self.

If self-respondents have a greater tendency to report health problems, people whose status changed from proxy in one cycle to self in the next

## Definitions

In the National Population Health Survey (NPHS), *proxy responses* are those obtained for a particular person from another household member (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 questionnaire* contains questions about demographic and socio-economic characteristics, as well as basic health information such as two-week disability, activity restriction, and chronic conditions. This information is obtained for each person in the household from one member.

The *Health questionnaire* contains questions on topics such as medication use, smoking, alcohol consumption, mental health, social support, height and weight, physical activity, and injuries. Given the detailed and often personal nature of this information, the selected respondent usually answers on his or her own behalf.

To measure *chronic conditions*, individuals were asked about long-term conditions that had lasted or were expected to last six months or longer and that had been diagnosed by a health professional. Interviewers read a list of conditions.

*Activity restriction* due to a long-term physical or mental health problem was based on a positive response to any of the following: "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* was determined by asking, "Do you have any long-term disabilities or handicaps?"

Respondents were classified as dependent in *instrumental activities of daily living* if they needed help preparing meals, shopping for groceries or other necessities, doing normal everyday housework, or doing heavy household chores such as washing walls or yard work.

To determine dependency in *activities of daily living*, respondents were asked if, because of a health problem, they needed help with personal care such as washing, dressing or eating, or with moving about inside the house.

*Two-week disability* was measured in terms of *bed-days* and *cut-down days* over the past two weeks. Respondents were asked about days they had stayed in bed because of illness or injury (including nights in hospital) and about days they had cut down on normal activities because of illness or injury.

*Living arrangements* were defined as living alone or living with others.

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

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

*Household income* 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 or 2	Less than \$15,000
	3 or 4	Less than \$20,000
	5 or more	Less than \$30,000
Lower-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

*Daily smokers* were defined as those who indicated that they smoked cigarettes daily.

The *Health Utilities Index* (HUI3) is a summary measure that incorporates functional health and societal preferences for health states.<sup>29,30</sup> Based on responses to 30 questions about eight aspects of functional health (vision, hearing, speech, mobility, dexterity, emotions, cognition, and pain and discomfort) together with a valuation component, an overall score, or index, is produced for each individual. Perfect health is rated 1.000, and death, 0.000; negative scores reflect health states considered worse than death. Possible HUI3 values range from -0.360 to 1.000. The HUI3 score was used as a continuous variable.

To measure *psychological distress*, respondents were asked to reply to six statements on a five-point scale: all of the time (score 4), most of the time (3), some of the time (2), a little of the time (1), or none of the time (0). "During the past month, about how often did you feel:

- so sad that nothing could cheer you up?"
- nervous?"
- restless or fidgety?"
- hopeless?"
- worthless?"
- that everything was an effort?"

The responses were summed, with higher scores indicating more distress (Cronbach's alpha = 0.77). Respondents with a score that fell in the upper quartile of the distribution (5 or more) were considered to be experiencing high psychological distress.

*Weight* was defined in terms of body mass index (BMI), which was calculated by dividing weight in kilograms by the square of height in metres. BMI was grouped into two categories: obese (30 or more) and not obese (less than 30). BMI was not calculated for pregnant women.



Table 6  
Distribution of cycle-to-cycle reporting status to General questionnaire, National Population Health Survey, longitudinal file, household population aged 18 or older in 1994/95, Canada excluding territories, 1994/95 to 2000/01

	%
Total	100.0
Self-self	73.5
Proxy-self	15.2
Self-proxy	5.2
Proxy-proxy	6.1

*Data sources:* 1994/95 to 2000/01 National Population Health Survey, longitudinal sample, Health file (square)

(proxy-self) might be more likely to report a new condition than those who reported on their own behalf in both cycles (self-self). Similarly, respondents whose information was provided by someone else in both cycles (proxy-proxy) and those whose reporting status changed from self to proxy (self-proxy) would be less likely report a new condition.

To investigate this possibility, a multiple logistic regression model was run for each chronic condition for which the cross-sectional analyses suggested a proxy effect (non-food allergies, asthma, arthritis, back problems, migraine, ulcers, and incontinence), as well as for an activity restriction or long-term disability. Because reporting status is related to variables such as sex, living arrangements, education, income and employment,<sup>31</sup> these characteristics were included as control variables.

Previous studies that used this approach have been limited by an inability to control for health status.<sup>21,32</sup> For example, some people whose information is provided by proxy may actually have a health advantage, in that they are healthy enough to be busy and away from home. With NPHS data, it is possible to control for this potentially confounding factor by using the Health Utilities Index, a summary measure based on responses to questions about eight aspects of functional health that are asked as part of the Health questionnaire (where proxy response was rare). In the regression models, the Health Utilities Index and two-year changes in the Health Utilities Index were also entered as control variables.

Even when the effects of all these factors were taken into account, in several cases, a change in

Table 7  
Two-year incidence rates of selected health conditions and adjusted odds ratios relating reporting status to incidence of conditions, household population aged 18 or older, Canada excluding territories, 1994/95 to 2000/01

	Two-year incidence rate	Two-year reporting status	Adjusted odds ratio	95% confidence interval
	%			
Non-food allergies	12.3	Self-Self†	1.0	...
		Proxy-Self	1.5*	1.3, 1.8
		Self-Proxy	0.8	0.6, 1.1
		Proxy-Proxy	0.9	0.7, 1.2
Back problems	9.1	Self-Self†	1.0	...
		Proxy-Self	1.1	0.9, 1.2
		Self-Proxy	0.7*	0.5, 0.9
		Proxy-Proxy	0.5*	0.4, 0.7
Arthritis	6.2	Self-Self†	1.0	...
		Proxy-Self	1.1	0.9, 1.4
		Self-Proxy	0.5*	0.3, 0.7
		Proxy-Proxy	0.4*	0.3, 0.6
Migraine	3.0	Self-Self†	1.0	...
		Proxy-Self	1.2	0.9, 1.5
		Self-Proxy	0.5*	0.3, 0.9
		Proxy-Proxy	0.6	0.3, 1.1
Asthma	1.9	Self-Self†	1.0	...
		Proxy-Self	1.3	1.0, 1.8
		Self-Proxy	0.9	0.6, 1.5
		Proxy-Proxy	0.5	0.3, 1.1
Urinary incontinence	1.6	Self-Self†	1.0	...
		Proxy-Self	1.0	0.7, 1.5
		Self-Proxy	0.3*	0.2, 0.5
		Proxy-Proxy	0.7	0.4, 1.2
Stomach/Intestinal ulcers	1.4	Self-Self†	1.0	...
		Proxy-Self	1.4	1.0, 1.9
		Self-Proxy	0.5*	0.3, 0.9
		Proxy-Proxy	0.5*	0.3, 0.9
At least one of 7 chronic conditions	24.7	Self-Self†	1.0	...
		Proxy-Self	1.4*	1.3, 1.6
		Self-Proxy	0.7*	0.6, 0.8
		Proxy-Proxy	0.8*	0.7, 0.9
Activity restriction or disability	8.1	Self-Self†	1.0	...
		Proxy-Self	1.2	1.0, 1.4
		Self-Proxy	0.7*	0.5, 1.0
		Proxy-Proxy	0.8	0.6, 1.1

*Data sources:* 1994/95 to 2000/01 National Population Health Survey, longitudinal sample, Health file (square)

*Notes:* Presents results of nine separate regression models, each controlling for sex, age, living alone, education, household income, employment status, Health Utilities Index, and changes in Health Utilities Index (see Definitions).

† Reference category

\* Significantly different from Self-Self ( $p < 0.05$ )

... Not applicable

reporting status from one survey cycle to the next was associated with the incidence of health problems (Table 7). Compared with respondents answering

on their own behalf in consecutive cycles (self-self), those whose reporting status changed from proxy to self (proxy-self) were more likely to report incident cases of non-food allergies. For arthritis, back problems and ulcers, the proxy-proxy group and the self-proxy group were less likely than the self-self group to report incident cases. As well, the self-proxy group was less likely to report migraine, incontinence and disability.

Because the number of incident cases of most conditions was generally small, especially when divided among the four reporting status categories, it is difficult to achieve statistical significance. But when the combined incidence of at least one of the seven conditions is considered, all reporting categories are significant and suggest a proxy effect, with the proxy-self group more likely to report a new condition, and the proxy-proxy and self-proxy groups less likely to do so.

### **Convenience versus necessity**

A health survey obviously cannot exclude people unable to provide their information because of a physical or mental health problem; that is, proxy by necessity. Proxy by convenience, however, can be avoided by increasing resources at the collection stage and allowing adequate time to contact all respondents directly. Therefore, an important element of a proxy effect is the extent to which it is due to proxy by necessity versus convenience.

For the 1998/99 cross-sectional General and Health files, the circumstances under which proxy by necessity was accepted for the General questionnaire were identical. Consequently, systematic differences in estimates of health problems between the two files must be attributable to proxy by convenience (Tables 2 to 4).

On the longitudinal file, it is possible to differentiate between convenience and necessity. Records with a proxy response for the General questionnaire and a self-response for the Health questionnaire were convenience cases, since these individuals were obviously able to respond on their own behalf. Likewise, it can be concluded that records with a proxy response for both questionnaires were out of necessity.

To study the potential impact of proxy by convenience on incidence measures, the regression models were rerun on the longitudinal file, this time excluding proxy by necessity. The results were similar to those for the full models, suggesting that accepting proxy responses by convenience has an effect on incidence measures of some health conditions.

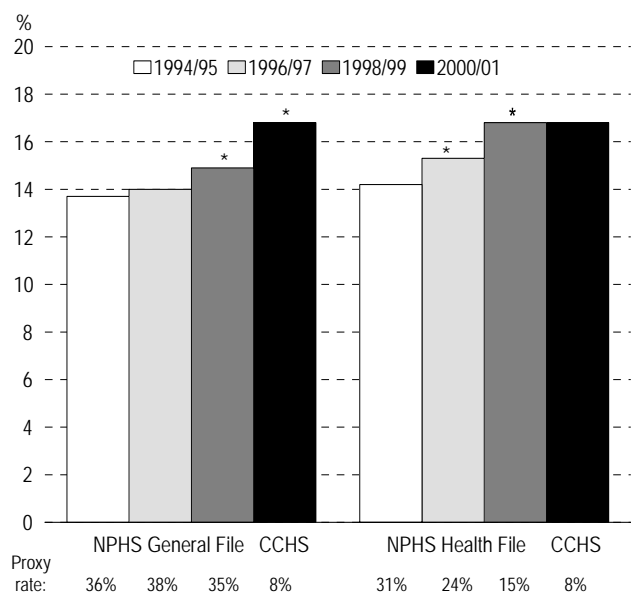
### **Does the proxy effect matter?**

This analysis indicates that accepting proxy responses can affect estimates of health conditions. For prevalence rates, the difference between the 1998/99 General and Health files is generally less than a percentage point. The implication for analysis, however, is that small changes in prevalence across survey cycles must be interpreted cautiously, even when they attain statistical significance. This can be illustrated by examining trends in the prevalence of arthritis.

Estimates based on the General file show a gradual increase in the arthritis prevalence rate from 13.7% in 1994/95 to 14.9% in 1998/99: 1.2 percentage points. The increase based on the Health file was greater, at just over 2.5 percentage points, likely owing to the decline in proxy reporting rates in the Health file over this period.

In 2000/01, the Canadian Community Health Survey (CCHS) replaced the NPHS as the source of cross-sectional health data. Proxy responses are accepted in the CCHS only by necessity. When data from the 2000/01 CCHS are compared with data from the 1998/99 NPHS Health file, the prevalence of arthritis remains constant. However, when the CCHS data are compared with the 1998/99 NPHS General file, the increase in the prevalence of arthritis over the two years is close to two percentage points (Chart 1). This apparent increase is likely attributable to the much lower proxy reporting rate for the CCHS: 8% compared with 35% for the 1998/99 NPHS General file. These findings suggest a need for care in reporting changes over time in the prevalence of arthritis, particularly when such increases coincide with sharp decreases in proxy reporting rates.

Chart 1  
Prevalence rates of arthritis, household population aged 18 or older, Canada excluding territories, 1994/95 to 2000/01



Data sources: 1994/95, 1996/97 and 1998/99 National Population Health Survey (NPHS), cross-sectional samples, Health and General files; 2000/01 Canadian Community Health Survey (CCHS)  
\* Significantly greater than previous period

The NPHS longitudinal file is often used to study relationships between risk factors and the incidence of health conditions. An important concern is whether those relationships are altered by reporting status. To investigate this possibility, incident arthritis was examined in two multivariate models: the first controlled for factors that have been shown to be associated with incident arthritis such as age, sex and obesity;<sup>33</sup> the second, for changes in reporting status as well.

The results suggest that although estimates of the two-year incidence of arthritis are affected by reporting status, this does not alter relationships with risk factors. The odds ratios for all variables in the model remained virtually the same when reporting status was taken into account (Table 8). Furthermore, when the second model was rerun, testing for interactions between arthritis risk factors and reporting status, none of the interaction terms was significant.

Table 8  
Adjusted odds ratios relating selected characteristics to two-year incidence of arthritis, without and with control for reporting status, household population aged 18 or older, Canada excluding territories, 1994/95 to 2000/01

	Without control for reporting status		With control for reporting status	
	Odds ratio	95% confidence interval	Odds ratio	95% confidence interval
<b>Reporting status</b>				
Self-Self <sup>†</sup>	...	...	1.0	...
Proxy-Self	...	...	1.2*	1.0, 1.5
Self-Proxy	...	...	0.6*	0.4, 0.8
Proxy-Proxy	...	...	0.6*	0.4, 0.9
<b>Sex</b>				
Women <sup>†</sup>	1.0	...	1.0	...
Men	0.7*	0.6, 0.8	0.7*	0.6, 0.8
<b>Age</b>				
	1.1*	1.0, 1.1	1.1*	1.0, 1.1
<b>Household income</b>				
Lowest/Lower-middle <sup>†</sup>	1.0	...	1.0	...
Upper-middle	1.1	1.0, 1.3	1.1	1.0, 1.3
Highest	1.0	0.9, 1.3	1.0	0.9, 1.3
<b>Doctor contacts in past year</b>				
0 to 6 <sup>†</sup>	1.0	...	1.0	...
7 or more	1.8*	1.6, 2.1	1.8*	1.6, 2.1
<b>Daily smoker</b>				
	1.3*	1.1, 1.5	1.3*	1.1, 1.5
<b>Psychological distress</b>				
Low <sup>†</sup>	1.0	...	1.0	...
High	1.4*	1.2, 1.6	1.4*	1.2, 1.6
<b>Obese</b>				
	1.5*	1.2, 1.7	1.5*	1.2, 1.7

Data sources: 1994/95 to 2000/01 National Population Health Survey, longitudinal sample, Health file (square)

Notes: Missing categories for income, psychological distress and obese variables were included in models to maximize sample size, but coefficients are not shown.

<sup>†</sup> Reference category. When not noted, reference category is absence of characteristic; for example, reference category for "daily smoker" is not daily smoker.

\* Significantly different from reference category (p < 0.05)

... Not applicable

### Concluding remarks

The results of this analysis of data from the National Population Health Survey and the Canadian Community Health Survey suggest that accepting proxy- rather than self-response may result in lower estimates of certain health conditions. Trends in prevalence may also be affected, as declines in proxy reporting rates between 1994/95 and 1998/99

coincided with increases in estimates. Incidence rates may be subject to a proxy effect as well, since changes in reporting status were associated with new reports of health conditions. However, based on the example of arthritis, the proxy effect does not alter associations with known risk factors.

Proxy responses are unavoidable, particularly in a health survey. Excluding people who are unable to respond because of ill health would bias the results. However, the practice of accepting proxy

responses to reduce the number of call-backs, and thereby, survey costs, may unnecessarily introduce bias.

Implementation of new procedures in 2000/01 meant that proxy responses will be accepted only by necessity in all future cycles of the National Population Health Survey, as well as for the Canadian Community Health Survey, which is now the source for cross-sectional estimates. ●

## References

- Swain L, Catlin G, Beudet MP. The National Population Health Survey—its longitudinal nature. *Health Reports* (Statistics Canada, Catalogue 82-003) 1999; 10(4): 69-82.
- 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.
- Béland Y. Canadian Community Health Survey—Methodological overview. *Health Reports* (Statistics Canada, Catalogue 82-003) 2002; 13(3): 9-14.
- Statistics Canada. *National Population Health Survey: 1994-95 Public Use Microdata File* (Catalogue 82F0001XCB) Ottawa: Minister of Industry, 1995.
- Statistics Canada, Population Health Surveys Program. *National Population Health Survey, Cycle 4 (2000-2001), Household Component, Longitudinal Documentation*. Ottawa: Health Statistics Division, Statistics Canada, 2002.
- 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.
- Rust KF, Rao JNK. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 1996; 5: 281-310.
- 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, 1999.
- National Center for Health Statistics. Health Interview responses compared with medical records. *Vital and Health Statistics* 1965; 2(7): 1-40.
- National Center for Health Statistics. Reporting of hospitalization in the Health Interview Survey. *Vital and Health Statistics* 1965; 2(6): 1-51.
- Harlow SD, Linet MS. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *American Journal of Epidemiology* 1989; 129(2): 233-48.
- 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.
- 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.
- Cannell CC, Marquis KH, Laurent A. A summary of studies of interviewing methodology. *Vital and Health Statistics* 1977; 2(69): 1-16.
- Clarridge BR, Massagli MP. The use of female spouse proxies in common symptom reporting. *Medical Care* 1989; 27(4): 352-66.
- 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.
- Koons DA. Quality control and measurement of nonsampling error in the Health Interview Survey. *Vital and Health Statistics* 1973; 2(54): 1-24.
- 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.
- 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 Social Statistics Section on Survey Research Methods* Washington, DC: American Statistical Association, 1986: 546-51.
- Nelson LM, Longstreth WT Jr., Koepsell TD, et al. Completeness and accuracy of interview data from proxy respondents: demographic, medical, and life-style factors. *Epidemiology* 1994; 5(2): 204-17.
- Todorov A, Kirchner C. Bias in proxies' reports of disability: data from the National Health Interview Survey on disability. *American Journal of Public Health* 2000; 90(8): 1248-53.
- White AA, Massey JT. Selective reduction of proxy response bias in a household interview survey. *American Statistical Association: Proceedings of the Social Statistics Section* Washington, DC: American Statistical Association, 1981: 211-6.
- Andresen EM, Fitch CA, McLendon PM, et al. Reliability and validity of disability questions for US Census 2000. *American Journal of Public Health* 2000; 90(8): 1297-9.
- Ball AE, Russell EM, Seymour DG, et al. Problems in using health survey questionnaires in older patients with physical disabilities. Can proxies be used to complete the SF-36? *Gerontology* 2001; 47(6): 334-40.

- 25 Corder LS, Woodbury MA, Manton KG. Proxy response patterns among the aged: effects on estimates of health status and medical care utilization from the 1982-1984 long-term care surveys. *Journal of Clinical Epidemiology* 1996; 49(2): 173-82.
- 26 Epstein AM, Hall JA, Tognetti J, et al. Using proxies to evaluate quality of life. Can they provide valid information about patients' health status and satisfaction with medical care? *Medical Care* 1989; 27(3 Suppl): S91-8.
- 27 Neumann PJ, Araki SS, Gutterman EM. The use of proxy respondents in studies of older adults: lessons, challenges, and opportunities. *Journal of the American Geriatrics Society* 2000; 48(12): 1646-54.
- 28 Rothman ML, Hedrick SC, Bulcroft KA, et al. The validity of proxy-generated scores as measures of patient health status. *Medical Care* 1991; 29(2): 115-24.
- 29 Furlong WJ, Feeny DH, Torrance GW, et al. *Multiplicative Multi-attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report*. Working Paper 98-11. Hamilton, Ontario: McMaster University Centre for Health Economics and Policy Analysis, 1998.
- 30 Furlong WJ, Feeny DH, Torrance GW. Health Utilities Index (HUI): Algorithm for determining HUI Mark 2 (HUI2)/Mark 3 (HUI3) health status classification levels, health states, health-related quality of life utility scores and single-attribute utility score from 40-item interviewer-administered health status questionnaires. Dundas, Ontario: Health Utilities Inc., 1999.
- 31 Shields M. Proxy reporting in the National Population Health Survey. *Health Reports* (Statistics Canada, Catalogue 82-003) 2000; 12(1): 21-39.
- 32 Mosely RR, Wolinsky FD. The use of proxies in health surveys. Substantive and policy implications. *Medical Care* 1986; 24(6): 496-510.
- 33 Wilkins K. Incident arthritis in relation to excess body weight. *Health Reports* (Statistics Canada, Catalogue 82-003) 2003; 15(1): 39-49.

## Appendix

Table A  
Number of respondents aged 18 or older in General and Health files, by reporting status, National Population Health Survey, 1994/95, 1996/97 and 1998/99

	Total	With proxy response
<b>1994/95</b>		
General file	41,045	15,455
Health file		
General questionnaire	16,291	4,158
Health questionnaire	16,291	601
<b>1996/97</b>		
General file	153,380	61,621
Health file		
General questionnaire	68,282	16,928
Health questionnaire	68,282	1,497
<b>1998/99</b>		
General file	34,543	12,835
Health file		
General questionnaire	14,150	1,721
Health questionnaire	14,150	252

Table B  
Number of respondents aged 18 or older in Longitudinal file, by reporting status, National Population Health Survey, Longitudinal file, 1994/95 to 2000/01

	Total	With proxy response
<b>Cycle 1 - 1994/95</b>		
General questionnaire	14,117	3,782
Health questionnaire	13,677	506
<b>Cycle 2 - 1996/97</b>		
General questionnaire	12,879	1,846
Health questionnaire	12,763	210
<b>Cycle 3 - 1998/99</b>		
General questionnaire	11,888	1,386
Health questionnaire	11,792	259
<b>Cycle 4 - 2000/01</b>	10,987	417

*Note: In cycle 4, the General and Health questionnaires were combined.*