

COMPARISON OF SCREENING STRATEGIES FOR THE HEALTH AND ACTIVITY LIMITATION SURVEY

Éric R. Langlet, Statistics Canada

R.H.Coats building, 15N, Tunney's Pasture, Ottawa, Canada, K1A 0T6

KEY WORDS: screening surveys; coverage issues; false positives; false negatives

1. Overview of the problem

The household component of the 1991 Statistics Canada Health and Activity Limitation Survey (HALS) collects information on the nature and severity of disabilities, and information on the barriers which disabled persons face in the conduct of their daily activities. Two questionnaires are used, one for the adults aged 15 and over and one for the children aged 14 and under. This paper will focus on the adult population.

This survey selects a sample among individuals who reported a limitation in the Census ("yes" sample) and a sample from those individuals who did not ("no" sample). Even though the probability for someone in the "no" sample of being limited to the 1991 HALS is relatively small (the false negatives), individuals limited to HALS coming from the "no" sample represent about 50% of the individuals limited to HALS. The sampling of the "no" population, although necessary for bias concerns, is however very costly given the size of this population. The sampling fraction in the 1991 HALS is about 10 times smaller in the "no population" compared to the "yes" population. This feature has a serious effect on the variance of the estimates.

For the 2001 Census, new screening questions more in line with the HALS limitation definition are considered. The degree of correspondence between the two definitions was measured between April and July 1999 through a test involving respondents from the 1998 National Census Test (NCT). In the 1998 NCT, two versions of the long form questionnaire were tested using the old and the new Census activity limitation screening questions. In the test, a sample from each group was selected and interviews were conducted using the screening questions of the 1991 HALS questionnaire. The proportion of false positives and negatives will be compared between individuals who received the old and the new screening questions.

2. The 1991 HALS survey

The 1991 HALS survey is a post-censal survey, that is a survey that uses Census microdata to identify the population of interest. A post-censal survey also integrates the survey operations with those of the Census and is conducted shortly after the Census to preserve the currency of the Census information. The post-censal survey approach has proved in the past to be a cost-effective means of collecting data for a rela-

tively small and scattered subgroup of the population and for producing small domain estimates. The Census microdata constitutes a good survey frame containing for each person a large amount of information that can be used in the sampling design and estimation methods. Other examples of post-censal surveys in Statistics Canada include the 1986 HALS and the 1991 Aboriginal Peoples Survey.

In the household component of the 1991 HALS, separate questionnaires have been developed for the children and the adults. In each case, the survey questionnaire begins with a screening section where the nature, cause and severity of disability is defined. Other sections collect information on the impact of disability in the everyday life of the respondent in areas such as employment, education, transportation, social network, leisure, accommodation and finances.

The main problem of a post-censal survey is the identification of the population of interest through the Census questionnaire. This problem can be summarized by the following table.

Census		Post-Censal survey	
		Screened-in	Screened-out
	positive	True positives	False positives
negative	False negatives	True negatives	

The Census, through a limited number of questions, allows the classification of respondents to "positive" and "negative" groups, also called "yes" and "no" populations. Ideally, most people classified as positive should belong to the target population and vice-versa. Then, during the post-censal survey, the interview starts with a set of screening questions, usually much more complex than the Census questions, which allow a more precise determination of whether the person belongs to the post-censal survey target population. The efficiency of a post-censal survey is a function of the concentration of the population in the table diagonal, which implies few false positives and few false negatives. The conformity of the two classifications depends mainly on the complexity of the concept measured. Typically, disability is a concept difficult to measure since it relies for a good part on the perception of the respondent.

For the 1991 HALS, Census questions on activity limitations and long-term disabilities are included on the "long form", which is completed by a one in five households across Canada. These questions are used as

a stratification factor improving the efficiency of identifying the target population in HALS. Any respondent who indicates a "yes" answer to at least one of the following questions will be considered as "positive" and all others as "negative".

ACTIVITY LIMITATIONS		At home? 03 <input type="radio"/> No, not limited 04 <input type="radio"/> Yes limited
18. Is this person limited in the kind or amount of activity that he/she can do because of a long-term physical condition, mental condition or health problem:	At school or at work? 05 <input type="radio"/> No, not limited 06 <input type="radio"/> Yes, limited 07 <input type="radio"/> Not applicable	
See Guide	In other activities, e.g. transportation to or from work, leisure time activities? 08 <input type="radio"/> No, not limited 09 <input type="radio"/> Yes limited	
19. Does this person have any long-term disabilities or handicaps?	10 <input type="radio"/> No 11 <input type="radio"/> Yes	
See Guide		

A sample from both groups is taken and selected persons are interviewed using either the adult or children HALS questionnaire. In each questionnaire, an initial screening section identifies disabled persons who are then asked, in subsequent sections, questions concerning the barriers faced by disabled persons in their daily activities. Screening for the adults is carried out using a modified version of the Activities of Daily Living (ADL) developed by the World Health Organization in 1982 for physical disabilities. Physical disabilities covered by these questions are hearing, seeing, speaking, mobility and agility. Mental disability screening questions are also included covering learning disabilities, mental handicaps and psychiatric disabilities. Any respondent having a positive answer to one of the screening questions will be considered as disabled.

The following table presents the relationship between the Census disabled population and the 1991 HALS disabled population.

Adult population		Post-Census survey		Total
		Disabled	Not disabled	
Census	Positive	80%	20%	10%
	Negative	47%	3%	90%
		10%	90%	
Total		53%	97%	
		17%	83%	100%

The percentage of positives in the Census is 10% as compared to 17% for the HALS disabled population. Almost 80% of the adults classified as positive in the Census were classified as disabled in HALS (true positives). However, even if only about 10% of the persons classified as negative were actually disabled (false

negatives), this group represents almost 53% of the HALS disabled population. Consequently, sampling only from the "yes" stratum would seriously bias the survey results. The sampling from the "no" stratum, although necessary for bias concerns, is however very costly. Indeed, in 1991, a sample of about 36,000 "yes" was selected compared to 113,000 "no". Given the very large size of the "no" population compared to the "yes" population, the sampling rate in the "no" has to be much smaller than in the "yes". As a result, the weights in the "no" sample are about 10 times larger on average than in the "yes" sample. This feature, although necessary for cost-efficiency concerns, has a serious impact on the variance of the estimates.

Although the "no" sample contributes for 53% of the disabled population, most of the individuals in that group are mildly disabled. The following table compares for the disabled population the severity of the disability between the "yes sample" and the "no sample".

	"yes sample"	"no sample"
Mild	29%	67%
Moderate	40%	27%
Severe	31%	6%

Hence, for users interested mainly in the moderately and severely disabled individuals, there is a much higher correspondence between the two sets of screening questions (Census and HALS).

3. The 1999 HALS test

In preparation for the 2001 HALS, new Census screening questions, more in line with the HALS definition of disability were considered. Some qualitative studies were conducted before coming up with the new questions. In these qualitative studies, a sample of individuals who answered only one "yes" to the 1996 Census screening questions (over a possibility of four) was selected. The interview started by asking again the Census screening questions followed by the 1991 HALS screening questions. Many of the selected respondents answered "no" to the Census screening questions this time. The focus of the qualitative studies was on respondents who answered "no" to the Census screening questions and "yes" to the HALS screening questions. This sample selection method was an efficient way to eventually discover false negative respondents. Some of the reasons mentioned for this apparent discrepancy ("no" to Census and "yes" to HALS) was the fact that, in the Census, the answer to each screening question had to be "yes" or "no". There was no "sometimes". Also the words "limited", "disabilities" and "handicap" had a scarecrow effect on the respondents.

These studies led to the replacement of question 19 on long term disabilities and handicaps by a summary

of the ADL and the second question was rephrased by using the term “activity reduction” as opposed to “activity limitation”. Moreover, for each question, the category “sometimes” was added. The new questions appear below.

ACTIVITIES OF DAYLY LIVING	
20. Does this person have any difficulty hearing, seeing, communicating, walking, climbing stairs, bending, learning or doing any similar activities?	01 <input type="radio"/> Yes, sometimes 02 <input type="radio"/> Yes, often 03 <input type="radio"/> No
21. Does a physical condition or mental condition or health problem reduce the amount or the kind of activity this person can do:	04 <input type="radio"/> Yes, sometimes 05 <input type="radio"/> Yes, often 06 <input type="radio"/> No
(a) at home?	07 <input type="radio"/> Yes, sometimes 08 <input type="radio"/> Yes, often 09 <input type="radio"/> No
(b) at work or at school?	11 <input type="radio"/> Yes, sometimes 12 <input type="radio"/> Yes, often 13 <input type="radio"/> No
(c) in other activities, for example, transportation or leisure?	

It is expected that these new questions should produce a higher rate of false positives and a smaller rate of false negatives. A substantial reduction of the false negatives could possibly eliminate the need to sample from the “no” stratum. Also, even if using the new questions would produce a non-negligible false negative rate, it is possible that the Census questions would identify an important subset of the survey disabled population, in particular individuals who are moderately and severely disabled. Typically, the group of mildly disabled individuals, which the Census may not identify, is of less interest to the users in general.

The expected increase of false positives with the new questions is not a big concern especially if no sampling is required in the “no stratum”. Under this scenario, substantial savings could be made by only increasing the “yes” sample appropriately. Even if sampling is still required in the “no stratum” for 2001, a higher degree of correspondence between the Census and the HALS screening questions could lead to a more efficient sampling plan. Under this second scenario, savings would be much less, however.

The 1999 HALS test selects respondents from the 1998 National Census Test (NCT). In the 1998 NCT, two versions of the long form questionnaire (2B1 and 2B2) were tested and a new version of the short form questionnaire (2A) was also tested. Individuals in 13 different geographical sites were systematically assigned to either the 2A, 2B1 or the 2B2. In the 2B1, the old Census activity limitation screening questions were used while the new questions were used for the 2B2.

In the 1999 HALS test, a sample from the 2B1 and the 2B2 was selected and interviews were conducted by Computer Assisted Telephone Interviewing (CATI) using only questions of the 1991 HALS screening section identifying the disabled population, the nature and the severity of their disability. It should be noted that the 1991 HALS screening section also contained some follow-up questions for each positive answer to a given screening question. That is, individuals reporting a particular difficulty were asked at what age they first had the difficulty, what was the main condition or health problem causing the difficulty and what was the cause of this condition. Moreover, the 1991 HALS questionnaire contained very extensive follow-up questions for individuals selected in the screening section. As will be seen, these differences can influence very significantly the disability rate.

In order to obtain approximately 10,000 interviews, with an expected response rate of 80%, a sample of 12,500 individuals was allocated in different strata in order to estimate the expected proportions of false positives and negatives with a desired coefficient of variation (CV). Having achieved the CV requirement, the sample size in each subgroup (or stratum) was inflated proportionally in order to obtain a total sample size of 12,500. Each stratum is then sorted by a number of characteristics and a random number. Then, a systematic sample of the required sample size is selected within each stratum. The stratification was done by type of Census form (2B1 vs. 2B2), by adult vs. children and by answer to the Census activity limitation questions (“yes” vs. “no”).

Having fixed a CV within each stratum, the required sample size can be easily determined assuming that the sample design within each stratum is equivalent to a simple random sample. Under these conditions, if n denotes the sample size, N the population size, the CV of an estimated proportion \hat{p} is given by the following formula:

$$CV(\hat{p}) = \frac{SE(\hat{p})}{\hat{p}} = \sqrt{\frac{N-n}{N(n-1)} \frac{(1-\hat{p})}{\hat{p}}}$$

where $SE(\hat{p})$ denotes the standard error of the estimated proportion \hat{p} . Isolating n in the equation, the sample size required is given by

$$n = \frac{\hat{p}CV^2 + (1-\hat{p})}{\hat{p}CV^2 + \frac{(1-\hat{p})}{N}}$$

The following table gives the estimated population size, the NCT sample size and the final HALS test sample size in each stratum.

Stratum			Population size	NCT sample size	HALS test sample size
2B1	adult	Yes	19,971	4,413	1,092
2B1	adult	No	134,594	29,999	2,666
2B1	children	Yes	1,185	262	262
2B1	children	no	32,412	7,262	1,789
2B2	adult	Yes	31,045	6,805	866
2B2	adult	No	121,795	26,859	3,552
2B2	children	Yes	1,642	360	360
2B2	children	no	31,909	7,030	1,912
total			374,553	83,040	12,499

Note that the total population size is in fact about twice the population size in the 13 sites since each of the 2B1 and 2B2 samples are weighted up to the total population.

In order to estimate the required proportions, several weight adjustments were done. An important one was the weight adjustment for total non-response in the NCT. The overall response rate for the 2B1 was 67.7% and 67.2% for the 2B2. Weights were adjusted for different response rates by NCT site (the site was the only information available for non-respondents). This adjustment is obviously insufficient. It is expected that non-respondents contain a higher proportion of younger and more mobile individuals, a subgroup which should have a lower disability rate. For this reason, weights were also post-stratified for the age and sex distribution of the 1996 Census for the same 13 NCT sites. It is also obvious that the NCT population is not representative of the Canadian population. These factors should be taken into account when comparing results from the test to the 1991 HALS survey.

4. Results from the 1999 HALS test

The percentage of individuals in the adult population classified as "positive" is 12.2% for the 2B1 and 18.8% for the 2B2. Hence, the new version of the Census activity limitation questions successfully identified a larger number of individuals with a positive response. This is mainly attributed to the new "scaled-response" (yes sometimes, yes often, no) as well as the new "catch all" question 20. In fact, the 18.8% of adults classified as "positive" split into 7.4% with at least one "often" and 11.4% with only "sometimes" answers. Whether these individuals correspond to the HALS target population is a different story. It is expected that the new format of the questions will capture more mildly disabled individuals, a category that was previously coming mainly from the "no sample" in 1991.

The following table gives the percentage of positive responses to the Census activity limitation questions for the 1998 NCT, the 1996 Census for individuals falling in the 1998 NCT sites, the 1996 full Census and the 1991 full Census.

	1998 NCT 2B2	1998 NCT 2B1	1996 Census NCT sites	1996 Census Canada	1991 Census Canada
Adults	18.8%	12.2%	12.2%	11.8%	10.0%

The percentage of positives for the 1998 NCT 2B1 (12.2%) is exactly the same as the one for the 1996 Census for the same selected sites. This is partially due to the fact that weights were post-stratified to the age and sex distribution of the 1996 Census for the same 13 NCT sites. This percentage is slightly higher than the one for the Canadian population for the 1996 Census (11.8%), which in turns was higher than the corresponding percentage for the 1991 Census.

Before presenting the disability rates obtained in the test and comparing them to the 1991 HALS, it is important to understand that the disability rate can vary widely for different similar surveys under slightly different conditions. The following table gives the disability rates obtained in different Statistics Canada (STC) surveys for the adult population.

Disability rates in different STC surveys for the adult population				
1983 TEST	1984 CHDS	1985 GSS	1985 CALIB	1986 HALS
19.5%	12.8%	31.7%	21.0%	12.5%

The 1983 TEST was a test prior to the 1984 Canadian Health and Disability Survey (CHDS). The 1985 CALIB was a calibration test prior to the 1986 HALS. The main difference between each test and its corresponding survey is that the tests only included the screening questions while the actual surveys had a large number of follow-up questions for the screened-in individuals. This may cause some form of interviewer bias. Indeed, the interviewer knowing that, for each screened-in individual, the interview will continue for an extra 45 minutes may want to verify that the one or two "yes" reported by an individual are actually real "yes". All surveys are using about the same set of screening questions except for the 1985 General Social Survey (GSS). This was a health survey containing only a subset of the HALS screening questions.

In addition to the inclusion or not of follow-up questions, several other sources of differences can be identified between these surveys. The context is an important factor. For instance, in the GSS, the screening questions are preceded by a few health questions. Answers to the screening questions could very well have been affected by the respondent being sensitized to his/her health problems before being asked these screening questions. Also, the introduction, which was not the same across surveys, is very important in defining the survey context. The period of reference is another source of difference. All surveys except the GSS specified to report only difficulties expected to last

6 months or more. The GSS, on the other hand, was simply asking what people could do on an average day. Whether proxy interviews are allowed in the survey is also important knowing that, in general, proxy interviews present lower disability rates. The GSS was the only survey not using proxy interviews. The three surveys with the largest disability rates contain proportionally much more mildly disabled individuals than the other ones. The group of mild disabled is particularly sensitive to the differences between the surveys. More information on these surveys can be found in Binder and Morin (1988).

Some of these sources of difference were present between the 1999 test and the 1991 HALS and will be explained shortly. The following table gives, for the 2B1, the 2B2 and the 1991 HALS, the rates of disability, false positives and negatives as well as the fraction of the disabled population screened-in by the "yes" sample and the correlation coefficient between the Census and HALS screening questions.

Source	Disability rate	False positives	False negatives	Disabled from yes	Correlation
2B1	42%	11%	35%	26%	0.35
2B2	42%	16%	32%	38%	0.41
1991	17%	21%	10%	47%	0.56

As expected, for the 2B2, the false positive rate is higher and the false negative rate is lower compared to the 2B1. However, both false negative rates are substantially higher than in 1991, which causes a large increase in the disability rate. This large increase comes mainly from the mild disabled as can be seen in the following table.

Severity	HALS 1991	1999 2B1	1999 2B2
Not limited	83%	58%	58%
Limited	17%	42%	42%
mild	8%	29%	29%
moderate	6%	8%	9%
severe	3%	5%	4%

Indeed, the proportion of mild disabled individuals more than tripled between 1991 and the 1999 test while the proportions in the moderate and severe categories increased by about 50%. The next table gives the fraction of each level of severity identified by the Census screening questions (fraction of each severity level coming from the "yes" sample).

Severity	1991 HALS	1999 2B1	1999 2B2
all	47%	26%	38%
mild	28%	12%	23%
moderate	57%	45%	61%
severe	81%	76%	84%

From this table, it can be seen that the 2B2 screens-in more disabled individuals than the 2B1 for all levels of severity and the advantage becomes more pronounced toward the mild disabled category. In fact, for the moderate and the severe categories, the 2B2 screens-in proportionally more individuals than the 1991 HALS, despite the much higher disability rate observed in 1999. It is expected, with a rate closer to the one observed in 1991, that this advantage would become even more pronounced.

A number of reasons can explain the large difference observed between the 1991 HALS and the 1999 test. The difference between the questionnaires is believed to be the major cause of this increase. The 1999 test only contained the screening questions to identify the disabled population, the type and severity of the disability. The 1991 HALS included some follow-up questions for each difficulty reported as well as very extensive follow-up questions for screened-in individuals. These questions are known to be somewhat painful and difficult to ask for individuals having very marginal limitations. As suspected for the 1984 CHDS and the 1986 HALS, the interviewer having gone to this process a few times might change his behavior for individuals appearing to be only marginally disabled. Hence, one or two initial "yes" answers by the respondent might be converted to "no" answers by the interviewer following clarifications with the respondent. It is not completely clear by doing that whether the interviewer reduces artificially the disability rate or whether the interviewer has a better understanding of the survey concepts after a few interviews and does the right thing thereafter.

Among other differences between the two surveys, is the fact that the 1991 HALS was conducted by PAPI rather than CATI in the 1999 test. Whether this fact could influence the disability rate is another question. Another element is the fact the population covered by the 1998 NCT sites is obviously not representative of the Canadian population. Finally, as the population is aging and as the concept of activity limitation becomes more and more accepted in the community, there is a natural increase in the disability rate that is expected from year to year. For instance, there was an increase of 18% in the rate of positive responses to the Census screening questions between the 1991 and the 1996 Census.

Because of the very high disability rate in the 1999 test compared to 1991, a second phase was conducted. In this second phase, a sample of about 1,000 screened-in individuals in the first phase was selected. Since the increase in the disability rate came mainly from the "mild disabled", this group was largely over-sampled in the second phase. Respondents selected were assigned the full 1991 HALS selection portion of the questionnaire. Because of the time constraint, this interview

was conducted by telephone using a paper and pencil (PAPI) questionnaire. In this questionnaire, for each difficulty reported, respondents were asked at what age they first had the difficulty, what was the main condition or health problem which caused the difficulty and what was the cause of the condition. For the very mildly disabled individuals, it was suspected that asking the follow-up questions would cause some of them to realize that their difficulty was not serious enough to be reported. Individuals not reporting a difficulty anymore in phase 2 were asked a follow-up question on the reason for change compared to phase 1. Since all questions in phase 1 are also asked in phase 2, this second phase can also be used to determine the stability of the responses.

Many changes from "disabled" in phase 1 to "not disabled" in phase 2 were observed. In fact the disability rate went down from 42% in phase 1 to about 30% in phase 2 for both the 2B1 and the 2B2 samples. This figure, however, assumes that all individuals who answered "no" in phase 1 would still answer "no" in phase 2. Since, among selected respondents in phase 2, a few changes were observed in their patterns of "yes" and "no" to the different screening questions, it is quite conceivable that a certain number of individuals saying "no" to all questions in phase 1 would answer at least one "yes" in phase 2. In order to estimate this number, a logistic regression predicting the answer "yes" or "no" in the phase 2 as a function of the disability score obtained in phase 1 (the score is based on the patterns of "yes" and "no" to the different screening questions) could be done. Obviously, this would be a projection (probability of a "yes" in phase 2 given a score of 0 in phase 1) since no individuals with a global "no" response in phase 1 were sampled in phase 2.

Observation of the phase 2 data collection revealed that in general, both the respondent and the interviewer had a better understanding of the questions. Four major reasons for changing from a "yes" in phase 1 to a "no" in phase 2 were observed. About 25% of them indicated that the difficulty was not serious enough to be reported. Another 25% mentioned that they never had a limitation. An additional 15% mentioned that they did not have a limitation at the last interview. These three reasons implicitly assume a misunderstanding from either the respondent or the interviewer in phase 1. Finally, another 25% mentioned an improvement of the situation since phase 1. This reason suggests that the difficulty reported in phase 1 was not a difficulty that lasted or was expected to last for 6 months or more, as specified in the questionnaire. Since phase 1 was conducted in April and phase 2 in July, seasonal problems such as arthritis for instance, are subject to this type of change. Phase 1 was conducted in a CATI mode while phase 2 was conducted by PAPI. Apparently, interviewers that have been exposed to CATI for a long pe-

riod tend to prefer PAPI to CATI. The reason mentioned was the fact that interviewers had to use more their judgment with PAPI. With CATI, the complete flow of the questionnaire is decided for them. Is CATI more subject in some cases to coding errors than PAPI? Without having evidence of this being true, this could be a possibility.

5. Conclusion

The main problem of a post-censal survey is the identification of the population of interest through the Census questionnaire. Since the amount of questions on the Census form to identify the disabled population is usually relatively limited, differences in the populations identified will always persist with a detailed activity limitation survey such as the 1991 HALS. This fact implies almost inevitably the sampling of the Census "no" population. In fact, usually the most detailed the survey questionnaire, the most people with disability will be identified. The sampling of the "no population" is however very costly given the relative sample size required in the "no population" compared to the "yes population", as seen in the 1991 HALS (more than three times higher). The 1999 HALS test attempted to use Census questions identifying a larger number of individuals limited in their activities and questions more in line with the concepts of the 1991 HALS.

Results showed a stronger relationship between the 2B2 screening questions and the HALS screening questions compared to the 2B1 questions. Fewer false negatives and more false positives were observed with the 2B2 than the 2B1 questions. The 2B2 questions screen-in more disabled individuals for all levels of severity and the milder the disability the more pronounced is this advantage. The two groups of positive respondents, that is the true positives and the false negatives, contain proportionally more mildly disabled individuals for the 2B2 than the 2B1.

The results clearly showed the superiority of the new Census screening questions over the old ones in terms of selecting a larger portion of the target population and in terms of missing a portion of the target population which is less critical (the mild disabled population). However, because of the very high rate of false negatives, this study did not permit to determine whether individuals with negative answers to the Census screening questions should be sampled or not. In order to answer this question, a further study under the exact same conditions and the exact same questionnaire as the 2001 HALS should be conducted.

References

Binder D. A. and Morin J. -P. (1988). Use of questions on activities of daily living to screen for disabled persons in a household survey. *The Canadian Journal of Statistics* Volume 16, Supplement, 143-156.