

# Comparison of sampling methods for hard-to-reach francophone populations: yield and adequacy of advertisement and respondent-driven sampling

Emmanuel Ngwakongnwi, Kathryn M King-Shier, Brenda R Hemmelgarn, Richard Musto, Hude Quan

## ABSTRACT

**Background:** Francophones who live outside the primarily French-speaking province of Quebec, Canada, risk being excluded from research by lack of a sampling frame. We examined the adequacy of random sampling, advertising, and respondent-driven sampling for recruitment of francophones for survey research.

**Methods:** We recruited francophones residing in the city of Calgary, Alberta, through advertising and respondent-driven sampling. These 2 samples were then compared with a random subsample of Calgary francophones derived from the 2006 Canadian Community Health Survey (CCHS). We assessed the effectiveness of advertising and respondent-driven sampling in relation to the CCHS sample by comparing demographic characteristics and selected items from the CCHS (specifically self-reported general health status, perceived weight, and having a family doctor).

**Results:** We recruited 120 francophones through advertising and 145 through respondent-driven sampling; the random sample from the CCHS consisted of 259 records. The samples derived from advertising and respondent-driven sampling differed from the CCHS in terms of age (mean ages 41.0, 37.6, and 42.5 years, respectively), sex (proportion of males 26.1%, 40.6%, and 56.6%, respectively), education (college or higher 86.7%, 77.9%, and 59.1%, respectively), place of birth (immigrants accounting for 45.8%, 55.2%, and 3.7%, respectively), and not having a regular medical doctor (16.7%, 34.5%, and 16.6%, respectively). Differences were not tested statistically because of limitations on the analysis of CCHS data imposed by Statistics Canada.

**Interpretation:** The samples generated exclusively through advertising and respondent-driven sampling were not representative of the gold standard sample from the CCHS. Use of such biased samples for research studies could generate misleading results.

**Emmanuel Ngwakongnwi**, PhD, was, at the time this study was conducted, a doctoral student in the Department of Community Health Sciences at the University of Calgary, Calgary, Alberta. He is now a Research Officer at the University of Calgary in Qatar, Doha, Qatar. **Kathryn M. King-Shier**, RN, PhD, is a Professor in the Faculty of Nursing and the Department of Community Health Sciences, University of Calgary, Calgary, Alberta. **Brenda R. Hemmelgarn**, MD, PhD, is a Professor in the Departments of Medicine and of Community Health Sciences, University of Calgary, Calgary, Alberta. **Richard Musto**, MD, is a Clinical Associate Professor, Department of Community Health Sciences, University of Calgary, Calgary, Alberta. **Hude Quan**, MD, PhD, is a Professor in the Department of Community Health Sciences, University of Calgary, Calgary, Alberta.

**Competing interests:** Emmanuel Ngwakongnwi, Kathryn King-Shier, Brenda Hemmelgarn, Richard Musto, and Hude Quan have all received grants from the Canadian Institutes of Health Research for studies other than the one reported here. No other competing interests declared.

**Funding:** This study was funded by the Canadian Institutes of Health Research (CIHR) Catalyst Grant on Official Language Minorities. Emmanuel Ngwakongnwi was supported by the Frederick Banting and Charles Best CIHR Doctoral Student Award. Kathryn King-Shier, Brenda Hemmelgarn, and Hude Quan were supported by salary awards from Alberta Innovates: Health Solutions.

**Correspondence:** Dr. Emmanuel Ngwakongnwi, University of Calgary in Qatar, PO Box 23133, Al Rayyan Campus, Doha, Qatar; [engwakon@ucalgary.ca](mailto:engwakon@ucalgary.ca)

➤ **REPRESENTATIVE SAMPLING IS IMPERATIVE IF STUDY** results are to be generalizable beyond the study sample. If an appropriate sampling frame is lacking, researchers may find it difficult to generate a random

sample that is representative of the target population. In Canada, the term “minority francophone” refers to people residing in primarily English-speaking provinces whose mother tongue or first official language is

French.<sup>1</sup> According to the 2006 census,<sup>2</sup> 6.8 million francophones (22.1% of total population) were residing in Canada. Of this number, 975 390 resided outside of Quebec, mostly in the provinces of Alberta, British Columbia, New Brunswick, and Ontario. French is one of Canada's national official languages, and Canada accepts immigrants with adequate knowledge of French. Thus, minority francophones in the English-speaking provinces may originate from the province of Quebec or from overseas. Health researchers need convenient and feasible mechanisms to recruit participants for studies involving hard-to-reach populations.<sup>3-6</sup> In Canada, one such hard-to-reach population (e.g., for survey studies) consists of minority francophones, as there is no registry of francophones, and neither ethnicity (except for registered Indians under the *Indian Act*) nor language is routinely collected as part of administrative health data.<sup>7</sup>

Advertising is one method used to recruit study participants. Often, researchers will advertise a study in print media, online, through email distribution, and with posters or flyers. Snowball sampling is a second method, whereby initial recruits are asked to name others in the target population who may be interested in the study.<sup>3,8</sup> These chain referrals usually continue until the target sample size is obtained. Respondent-driven sampling is another chain-referral method that incentivizes initial study participants (called "seeds") to recruit others.<sup>9</sup> Each seed is given coupons (usually numbering 3) with unique serial numbers, which they use in recruiting their peers in a first wave of recruitment. First-wave respondents who are willing to participate receive the same number of coupons to recruit second-wave respondents, who in turn recruit third-wave respondents, with expansion of the pool until the desired sample size is reached. With respondent-driven sampling, active seeds are rewarded twice, initially for agreeing to be interviewed and then for recruiting others for the study. The serial numbers on the coupons are used to link recruits to recruiters. As with other chain-referral sampling methods, participants recruited through respondent-driven sampling are not randomly selected, depending instead on the subjective choices of the recruiters. However, respondent-driven sampling has a methodologic advantage over other chain-referral methods in that it employs mathematical weights in the analysis to compensate for the nonrandom nature of the sample.<sup>9,10</sup> This allows for the calculation of relative inclusion probabilities for members of the population and thus the possibility of

deriving unbiased estimators and standard errors of the variables being studied.

Many researchers have used advertising, snowball sampling, and respondent-driven sampling, individually or in combination, to sample hard-to-reach populations. For example, Southern et al.<sup>11</sup> used a combination of paid and unpaid advertising to recruit a sample of Americans living in Canada. Researchers in the United States and Europe have used respondent-driven sampling to recruit injection drug users for their studies.<sup>12,13</sup> Studies in which advertising was used as the sole recruiting method have demonstrated its convenience as a recruitment strategy, despite the limitation of not generating the desired sample size within a certain period.<sup>14-16</sup> The findings of studies based on snowball sampling are often not generalizable because of the nonrandom nature of the resulting sample.<sup>3</sup> Respondent-driven sampling may have important advantages over advertising and snowball methods because participants can be recruited relatively quickly, and the resulting sample is diverse and has less investigator bias.<sup>17</sup> However, respondent-driven sampling requires that the researchers identify productive "seeds" and that 6 or more waves of recruitment be achieved.<sup>9,18,19</sup>

Hard-to-reach populations are often the focus of health care research, because they often have unique health care needs. To obtain valid and generalizable research results for these populations, sampling mechanisms must be optimized. In the study reported here, we assessed the adequacy of advertising and respondent-driven sampling to recruit separate samples of francophones living in Alberta, a majority English-speaking province of Canada. The study posed 2 questions:

- Would advertising and/or respondent-driven sampling as recruitment strategies yield study samples similar to a random sample of the francophone population in Calgary?
- Would respondent-driven sampling yield a sample adequate for generating valid population estimates representative of the francophone population in Calgary?

## Methods

**Design and setting.** We generated samples of francophones through advertising and respondent-driven sampling and then compared these samples with a random subsample from the Canadian Community Health Survey (CCHS).<sup>20</sup> The CCHS is a telephone survey based on random digital dialling. We considered the

CCHS a “gold standard” against which to compare our 2 sampling strategies, in terms of socio-demographic characteristics and 3 specific elements of the CCHS: perceived weight, health status, and having a regular medical doctor.

We conducted this cross-sectional survey of francophones in Calgary, Alberta, in 2010. As of 2006, there were 68 435 francophones in Alberta,<sup>21</sup> representing about 2% of the provincial population. Although some small towns and villages (e.g., Beaumont, Brosseau, Grande Prairie, and Lacombe) have clustered populations of francophones, the majority of the province’s francophones reside in the Calgary and Edmonton metropolitan areas. This study was approved by the Conjoint Health Research Ethics Board of the University of Calgary. Informed consent was obtained from participants.

**Data collection.** We derived our survey questions from the 2006 CCHS cycle 3.1 English and French questionnaires.<sup>22</sup> We used the same wording as the CCHS, to maintain comparability across data sources. The survey included questions about socio-demographic characteristics, self-perceived general health, perceived weight, and having a regular medical doctor. Both the French and English versions of the survey questionnaire were pilot-tested.

Potential participants recruited by advertising and respondent-driven sampling were screened against the following eligibility criteria: age 18 years or older, residence in Calgary for at least 12 months, and self-identification as francophone (on the basis of mother tongue or first official language spoken). Those who met the eligibility criteria had the choice of completing the survey by telephone, by mail, or online.

**CCHS recruitment.** The 2006 CCHS cycle 3.1 was a cross-sectional telephone survey administered nationally to household residents aged 12 years or older in all provinces and territories, with the exception of populations on First Nations reserves, on Canadian Forces bases, and in some remote areas.<sup>20</sup> The CCHS uses a multistage survey design that includes stratification and/or clustering of population units before sampling. Survey weights are provided for analysis to account for the complex survey design. The national response rate for the CCHS cycle 3.1 was 84.7%. A detailed description of the sampling frame and design methodology of the CCHS can be found elsewhere.<sup>20</sup> The CCHS was chosen as the gold standard because the sampling

strategy allows for a random selection of participants at smaller geographic units, such as cities or metropolitan areas. From the CCHS survey data, we obtained a random subsample of francophones (i.e., people who reported French as either the mother tongue or the first official spoken language) at least 18 years of age who resided in Calgary.

**Recruitment by advertising.** Internet and email distribution, newspapers, radio and television advertising, and posters and flyers were the mechanisms used to generate the sample of participants recruited by advertising.

We launched email and Internet advertising on Friday, 27 November 2009, via direct email distribution and by posting study information on a dedicated website for this study. More specifically, we sent flyers by email to leaders of agencies and community associations serving the francophone community in Calgary, asking them to forward the message to their various networks. These leaders had participated in a focus group in September 2009 to discuss how they could facilitate access to the francophone community for this study. A reminder email was sent on 23 January 2010, and a final reminder was sent on 4 February 2010. Potential participants were given the options of calling the research office, following a link to access the survey online, or requesting a paper copy of the questionnaire.

For newspaper recruitment, we advertised our study in 2 local monthly French newspapers (in December 2009 and January 2010, respectively) and an English newspaper distributed through the Calgary metropolitan transit system for 3 consecutive days (3 to 5 February 2010).

For radio and television recruitment, we recorded an advertisement that was aired on 9 December 2009 on the 2 public broadcasting networks (Radio-Canada [French] and CBC [English]), on television during the evening news and several times during a popular evening radio program.

On 4, 7, 8, and 20 December 2009, eye-catching posters and flyers were distributed to various locations throughout Calgary, including French-language schools, primary care practices, the French Centre at the University of Calgary, churches frequented by francophones, a francophone job centre, and various francophone service agencies. Each poster had an envelope-sized tri-fold pocket in which wallet-size flyers were available for pick-up by interested individuals.

Public information sessions about the study and how to participate were held at ethnocultural association

meetings and francophone service agencies. Direct recruitment was not undertaken at these events.

**Recruitment by respondent-driven sampling.** We launched respondent-driven sampling on 1 March 2010, 2 weeks after the advertising campaign was completed, to diminish any residual effects from the advertising. Initial seeds were selected from a pool of people who had participated in a previous study involving the same population and who had consented to being contacted for a survey in the future. The study coordinator screened participants who had reported their immigration status and sex in the previous study, who expressed willingness to recruit up to 3 peers for our current study, and who fulfilled the overall study criteria. We selected 8 eligible seeds, 4 immigrants and 4 non-immigrants, with 2 men and 2 women in each group. As noted previously, the success of respondent-driven sampling lies in finding productive seeds and reaching 6 or more waves of recruitment.<sup>9,18,19</sup> In selecting seeds with known attributes, our main assumption was that they would be “active” (i.e., would participate in the study and recruit others) and that the resulting sample would be similar in composition to the francophone population of Calgary. We took a cautious approach, choosing not to sample seeds by their affiliation with ethnocultural organizations or francophone service agencies, so that our final sample would not be limited to members of these organizations. Seeds were purposely<sup>23</sup> selected to ensure balance in terms of sex and immigration status. Each seed who completed the survey received 3 coupons to recruit others. The coupon included instructions on how to access the survey, a request to refer 3 other participants (for which the person making the referral would receive \$10 per eligible participant who completed the survey), and a unique, 4-digit serial number that enhanced the linkage of recruits to recruiters.

**Statistical analysis.** We used descriptive statistics to examine the characteristics of participants reached by the 3 methods of recruitment (CCHS, advertising, and respondent-driven sampling). First, we merged the data for participants in the groups recruited by advertising and respondent-driven sampling. These merged data and the CCHS data were analyzed separately, for the following 2 reasons: first, analysis and reporting of CCHS data requires application of sampling weights to account for the complex, multistage sampling design,<sup>20</sup> whereas such weighting was not required for the

samples that we recruited; second, the current study used raw data from the CCHS, which could be analyzed only in a secure, monitored environment (the Regional Data Centre in Calgary). In addition, Statistics Canada places some restrictions on reporting findings from CCHS data and making inferences for subpopulations. Consequently, we did not perform significance tests to determine whether the 2 samples (merged and CCHS) were from the same population.

To identify recruitment chains (number of recruits originating from active seeds), we analyzed data from the respondent-driven sample using RDSAT software, version 5.6.<sup>24</sup> We fitted logistic models for each sample to assess the likelihood of having a regular medical doctor after adjustment for age, sex, marital status, education, personal income, general health, and perceived weight. For CCHS data, we computed estimates using the sampling weights provided by Statistics Canada.<sup>25</sup> Data were analyzed with Stata software, version 11 (StataCorp LP, College Station, Texas).

## Results

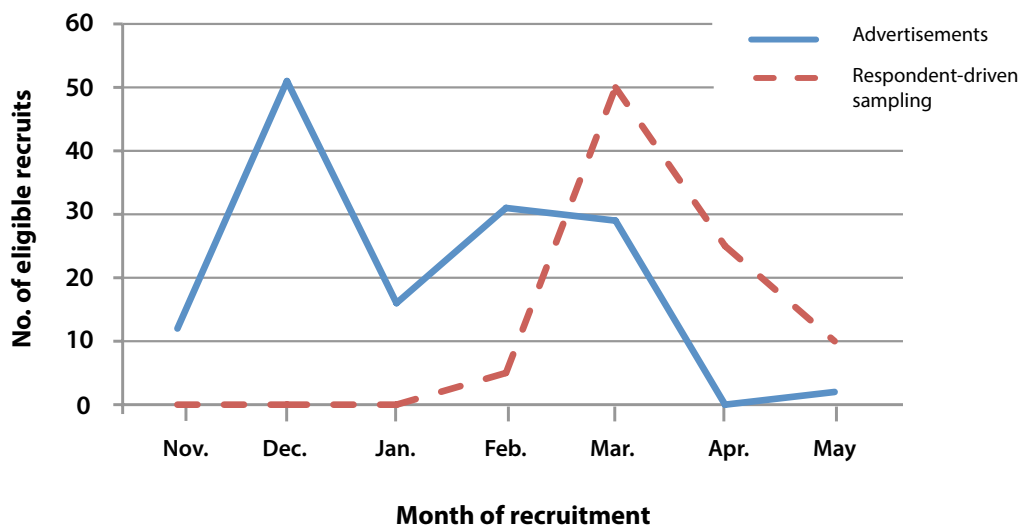
Advertising was conducted over 3 months (from 27 November 2009 to 15 February 2010) and yielded a total of 120 participants. Recruitment of eligible participants peaked in December 2009, declined in January 2010, and then improved in February 2010 (Figure 1). Advertising continued to yield recruits until May 2010, 3 months after the campaign had ended. The most effective source of information for recruitment by advertising was word of mouth (37 [30.8%] of recruits); the lowest yield was from posters and flyers (2 [1.7%] of recruits) (Table 1).

In respondent-driven sampling, the response from the 8 initial seeds was slow but eventually resulted in the recruitment of 35 other seeds, for a total of 43

**Table 1**  
**Information sources for recruitment by advertising in Calgary**

Source	No. (%) of participants n = 120	
Radio	6	(5.0)
Newspaper	7	(5.8)
Poster, flyer	2	(1.7)
Internet	6	(5.0)
Word of mouth	37	(30.8)
Community event	24	(20.0)
Other*	38	(31.7)

\*Church, friends, school.

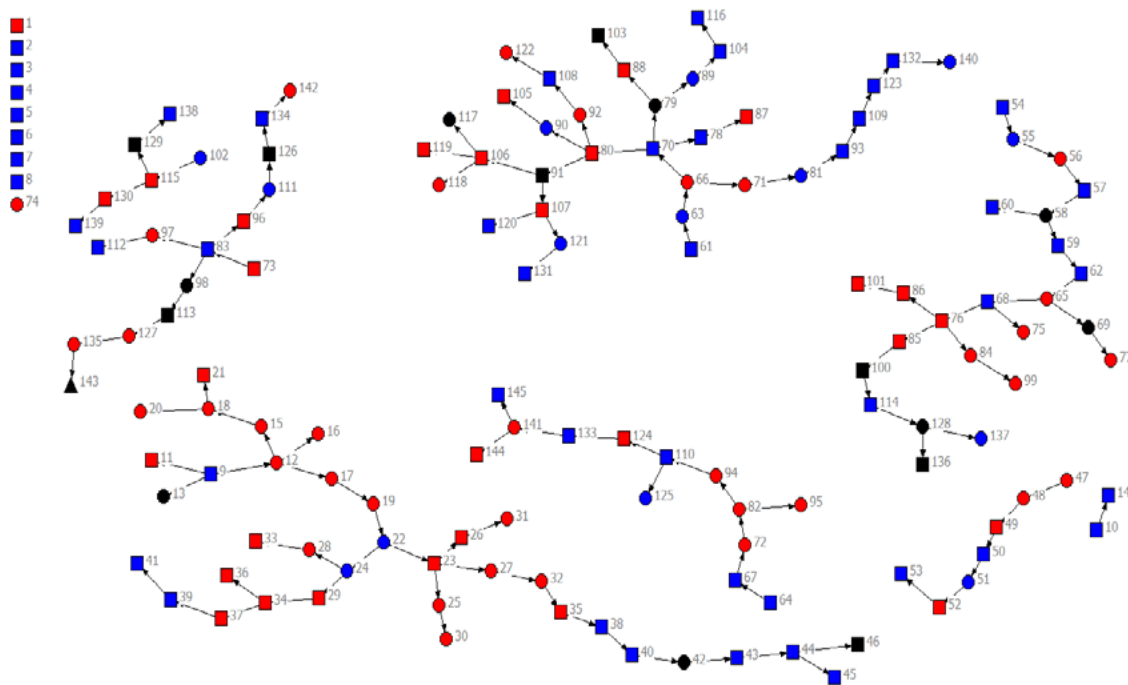


**Figure 1**  
Trends in recruitment by advertising and respondent-driven sampling

seeds. These 43 seeds were involved in 9 waves of recruitment, which yielded a total of 164 participants (see Figure 2). Of these participants, 19 were excluded from the analysis because of missing responses on questionnaire items. There was substantial attrition of seeds: 26 seeds did not complete the survey, and 9 completed the survey but did not recruit any additional participants. Of the 8 active seeds (those who completed the survey and recruited at least 1 other participant), 3 were men

(participant identification [ID] 13, 47, 102) and 5 were women (participant ID 10, 54, 61, 64, 73). Participants were likely to refer people with cultural and immigration status similar to their own.

Of the 265 participants recruited through advertising and respondent-driven sampling, 256 (96.6%) completed the survey in French and 9 (3.4%) in English. Modes of participation were 86 (32.5%) online, 12 (4.5%) by telephone, and 167 (63.0%) by mail. Because



**Figure 2**  
Netdraw visualization of recruitment chains.

The 8 chains represent active seeds, with isolated shapes at the top left corner representing inactive seeds. Circles = men, squares = women, blue = immigrants, red = nonimmigrants, black = individuals for whom immigration status was missing.



of the small number of respondents who participated by phone, responses for phone and Internet participation were combined. Compared with those who responded online or by phone, those who responded by mail were less likely to have college education or higher (77.2% v. 89.8%) but were more likely not to report personal income (28.1% v. 14.3%) and not to have a regular medical doctor (31.1% v. 15.5%) (Table 2). Among online/phone participants, the largest age group consisted of those 30 to 39 years of age (46.2%). Among mail respondents, the 2 largest age groups were those 30 to 39 years and 40 to 49 years (36.1% and 35.5%, respectively).

Of 285 Calgary francophones randomly selected from the CCHS, data for 259 participants were retained for analysis; the other 26 participants were excluded as they did not meet the study criteria. The samples generated through the 3 survey methods differed in terms of proportion of women (43.4% for the CCHS, 73.9% for advertising, and 59.4% for respondent-driven sampling), postsecondary education (59.1%, 86.7%, and 77.9%, respectively), birth place outside of Canada (3.7%, 45.8%, and 55.2%, respectively), missing data for personal income (12.1%, 14.2%, and 30.3 %, respectively), and not having a regular medical doctor (16.6%, 16.7%, and 34.5%, respectively) (Table 3).

## Interpretation

We tested the adequacy of advertising and respondent-driven sampling to recruit separate samples of francophones in Calgary, relative to a random sample from the CCHS (gold standard). Francophones residing primarily in an English-speaking city such as Calgary represent a hard-to-reach population for research purposes. The survey samples recruited in this study differed from the CCHS subsample in terms of basic demographic characteristics (age, sex, education, and place of birth) and thus were not representative of the francophone population in Calgary, as defined by the CCHS subsample. The 2 study samples

Table 2

### Characteristics of respondents, by mode of response

Characteristic	Mode of response; no. (%) of respondents*		p value
	Mail n = 167	Online or phone† n = 98	
Age, yr, mean (range)	38.9 (19–86)	39.8 (18–67)	
Age group, yr	n = 155	n = 91	0.003
18–29	30 (19.4)	14 (15.4)	
30–39	56 (36.1)	42 (46.2)	
40–49	55 (35.5)	16 (17.6)	
≥ 50	14 (9.0)	19 (20.9)	
Sex	n = 165	n = 93	0.20
Male	61 (37.0)	27 (29.0)	
Female	104 (63.0)	66 (71.0)	
Marital status	n = 163	n = 91	0.06
Single	39 (23.9)	11 (12.1)	
Married or common law	115 (70.6)	72 (79.1)	
Separated or divorced	9 (5.5)	8 (8.8)	
Education	n = 167	n = 98	0.028
Primary school or less	13 (7.8)	5 (5.1)	
Secondary or high school	25 (15.0)	5 (5.1)	
College or university	129 (77.2)	88 (89.8)	
Born in Canada	n = 163	n = 93	0.64
Yes	81 (49.7)	49 (52.7)	
No	82 (50.3)	44 (47.3)	
Place of residence	n = 166	n = 97	0.89
Urban	148 (89.2)	87 (89.7)	
Rural	18 (10.8)	10 (10.3)	
Personal income, \$	n = 167	n = 98	0.025
Lowest (< 50 000)	69 (41.3)	38 (38.8)	
Middle (50 000–60 000)	12 (7.2)	15 (15.3)	
Upper middle (60 000–80 000)	14 (8.4)	13 (13.3)	
Highest (> 80 000)	25 (15.0)	18 (18.4)	
Missing	47 (28.1)	14 (14.3)	
General health	n = 167	n = 98	0.78
Excellent	58 (34.7)	29 (29.6)	
Very good	66 (39.5)	40 (40.8)	
Good	36 (21.6)	23 (23.5)	
Fair or poor	7 (4.2)	6 (6.1)	
Perceived weight	n = 156	n = 91	0.24
Overweight	45 (28.8)	35 (38.5)	
Underweight	4 (2.6)	1 (1.1)	
Just about right	107 (68.6)	55 (60.4)	
Has regular doctor	n = 164	n = 97	0.005
Yes	113 (68.9)	82 (84.5)	
No	51 (31.1)	15 (15.5)	

\* Except where indicated otherwise. For some variables, the percentages may not sum to 100 because of rounding.

† Only 12 participants used the telephone to respond. Therefore, online and phone respondents were grouped.

**Table 3**  
**Characteristics of respondents, by sampling method**

Characteristic	Sampling method; no. (%) of respondents*			
	CCHS* n = 259	This study, combined† n = 265	This study, advertising n = 120	This study, RDS, n = 145
<b>Age, yr, mean (range)</b>	42.5	39.2 (18–86)	41.0 (18–86)	37.6 (19–73)
<b>Sex</b>		n = 258	n = 115	n = 143
Male	56.6	88 (34.1)	30 (26.1)	58 (40.6)
Female	43.4	170 (65.9)	85 (73.9)	85 (59.4)
<b>Marital status</b>		n = 254	n = 113	n = 141
Single	23.9	50 (19.7)	12 (10.6)	38 (27.0)
Married or common law	66.7	187 (73.6)	90 (79.7)	97 (68.8)
Separated or divorced	9.4	17 (6.7)	11 (9.7)	6 (4.3)
<b>Education</b>		n = 265	n = 120	n = 145
Primary school or less	8.4	18 (6.8)	6 (5.0)	12 (8.3)
Secondary or high school	32.5	30 (11.3)	10 (8.3)	20 (13.8)
College or university	59.1	217 (81.9)	104 (86.7)	113 (77.9)
<b>Born in Canada</b>				
Yes	96.3	130 (49.1)	65 (54.2)	65 (44.8)
No	3.7	135 (50.9)	55 (45.8)	80 (55.2)
<b>Place of residence</b>				
Urban	94.7	235 (88.7)	108 (90.0)	127 (87.6)
Rural	5.3	30 (11.3)	12 (10.0)	18 (12.4)
<b>Personal income, \$</b>				
Lowest (< 50 000)	51.8	107 (40.4)	50 (41.7)	57 (39.3)
Middle (50 000–60 000)	10.5	27 (10.2)	15 (12.5)	12 (8.3)
Upper middle (60 000–80 000)	13.6	27 (10.2)	14 (11.7)	13 (9.0)
Highest (> 80 000)	11.8	43 (16.2)	24 (20.0)	19 (13.1)
Missing	12.1	61 (23.0)	17 (14.2)	44 (30.3)
<b>General health</b>				
Excellent	26.0	87 (32.8)	38 (31.7)	49 (33.8)
Very good	42.1	106 (40.0)	48 (40.0)	58 (40.0)
Good	20.3	59 (22.3)	27 (22.5)	32 (22.1)
Fair or poor	11.6	13 (4.9)	7 (5.8)	6 (4.1)
<b>Perceived weight</b>		n = 247	n = 112	n = 135
Overweight	43.2	80 (32.4)	42 (37.5)	38 (28.1)
Underweight	4.6	10 (4.0)	3 (2.7)	7 (5.2)
Just about right	52.2	157 (63.6)	67 (59.8)	90 (66.7)
<b>Has regular doctor</b>		n = 265	n = 120	n = 145
Yes	83.4	195 (73.6)	100 (83.3)	95 (65.5)
No	16.6	70 (26.4)	20 (16.7)	50 (34.5)

Note: CCHS = Canadian Community Health Survey, RDS = respondent-driven sampling.

\*Percentages for the CCHS (gold standard) are weighted; therefore, raw data are not supplied.

†Advertising plus RDS.

generated by advertising and respondent-driven sampling were similar to a large extent, but the differences from the CCHS subsample persisted even when the 2 study samples were pooled. Relative to the CCHS, the advertising strategy resulted in oversampling of women and participants with postsecondary education. Respondent-driven sampling resulted in oversampling of women, participants with postsecondary education, immigrants, and people with missing data for personal income. Over-representation of women and immigrants in the respondent-driven sampling likely relates to the fact that most of the active seeds were female immigrants.

Use of combined strategies to sample hard-to-reach populations has been documented in other studies not involving francophones. In Canada, Southern et al.<sup>11</sup> found that use of a media conference, supplemented by a nationwide media release and paid advertisements in newspapers, produced a large enough sample of analyzable responses from Americans living in Canada. In the United States, use of direct mail and media releases improved the participation of people from ethnic minorities in a clinical trial.<sup>15</sup> However, these studies did not examine the representativeness of their samples. Our comparison of the sample of participants recruited by advertising with the random subsample from the CCHS revealed differences that raise questions about the impact of selection bias on findings from studies that used advertising as the sole recruitment strategy.

The use of respondent-driven sampling as an effective strategy to recruit a representative sample of a hard-to-reach population was well documented in a study involving drug users in New York City.<sup>18</sup> Starting with 8 seeds, the investigators recruited 618 drug users during 6 waves of recruitment over 13 weeks. Their study sample may have been representative of the overall population, as sample characteristics were similar to the characteristics of drug users recruited for other studies conducted in the same city. Our use of respondent-driven sampling did not achieve the same results. This inconsistency may relate to differences in study populations and recruitment processes. In the previous study, a specific population with low socio-economic potential was surveyed in New York City,<sup>18</sup> and monetary incentives may have influenced participation. For the current study, the target population comprised francophone immigrants and interprovincial migrants, with differing economic potential and socio-demographic characteristics; here, monetary incentives may not have had the same impact. The oversampling of immigrants

that we observed was likely due to the tendency of immigrants to cluster in certain geographic areas and to have strong social ties. Thus, seeds who were immigrants were more likely than non-immigrants to recruit participants within their communities and the same language group. The benefits of respondent-driven sampling might be optimized in a future study by ensuring that active seeds are retained through appropriate types and amounts of incentives and by selecting diverse seeds to minimize oversampling of participants with similar traits.

We encountered a series of challenges during this study. Advertising was scheduled for 3 months, with a 2-week washout period before initiation of respondent-driven sampling. However, advertising had a residual effect that extended into the period for respondent-driven sampling. Survey respondents during the respondent-driven sampling phase reported learning about the study from at least one of several advertising sources. It was not possible for us to recall all forms of advertising about the study during the respondent-driven sampling period (e.g., posters, flyers, newspapers, and information distributed by Internet and email). However, use of media (newspapers and radio), posters and flyers yielded the fewest recruits relative to other forms of advertising.

The December 2010 earthquake in Haiti might have been a major distraction from the advertising campaign. During that time, media attention was focused on the devastation and on relief efforts for victims. Haitians form a considerable proportion of the francophone community in Calgary, and these potential participants might have been more concerned about the safety of family members and loved ones than participating in a study. Preparation for the launch of respondent-driven sampling began late in January 2010, but intensified at washout, with the distribution of coupons to initial seeds. The first participants from respondent-driven sampling completed the survey earlier than the scheduled launch date of 1 March 2010, which suggests that the initial seeds had been making referrals before the official start of the respondent-driven component of the study. Despite early referrals for respondent-driven sampling, recruitment was initially slow because some seeds were inactive. We addressed this problem by identifying new seeds, a strategy that substantially improved the final study sample.

**Limitations.** This study had some limitations. First, our choice of the CCHS as the gold standard constituted a



major limitation. The CCHS is a household survey administered nationally to English- and French-speaking Canadian households in all provinces and territories. However, for our study, we selected a subsample of French speakers restricted to one geographic area, whose response rate in the CCHS is unknown. Additionally, the CCHS data were 3 years older than the data obtained through advertising and respondent-driven sampling. As a consequence, generalizability of our findings is limited. Second, we did not weight the respondent-driven sample in our analysis. Such an analysis would require integrating network size to estimate the probability of recruitment from a seed. This probability would then be used to generate mathematical weights to adjust for sampling deficiencies, because the respondent-driven sample depends on the design and the weights. One reason for not weighting the sample was the substantial missing data for the francophones in each network. For the few seeds that did generate networks the range was wide (5–40 individuals), which rendered the quality of reported network size questionable for weighting. Third, because different statistical methods were used to analyze data from different sources, it was not possible to combine datasets and perform multivariate comparisons to assess variability in estimates and thus determine whether the samples were significantly different. Without showing variability around the estimates, it is likely that the differences in magnitude of the findings between the combined sample and the individual samples (advertising or respondent-driven sampling) may have been due simply to selection size. Merging the CCHS data with study survey data was not feasible given the data access and reporting requirements of Statistics Canada. Finally, the findings are based on only one sample each from advertising and respondent-driven sampling, and therefore the observed biases may have arisen by chance.

**Conclusion.** We examined the yield and adequacy of advertising and respondent-driven sampling to recruit minority francophones in the absence of a sampling frame. We determined that samples generated through these methods differed from the gold standard (CCHS cycle 3.1) in terms of basic demographic characteristics such as age and sex and were thus not representative of the general francophone population. These biased samples could have generated misleading conclusions for variables such as having a regular doctor. Although not examined in the current study, the growing popularity of online social networks such as Facebook and

Twitter offers opportunities for future research to test the applicability of social media in sampling, especially for seed selection in respondent-driven sampling. However, employing social media for research will require that issues of consent, confidentiality, and data security be addressed.

**Contributors:** All authors (Emmanuel Ngwakongnwi, Kathryn King-Shier, Brenda Hemmelgarn, Richard Musto, and Hude Quan) participated equally in the conception and design of the study and interpretation of data. Emmanuel Ngwakongnwi and Hude Quan performed the data analysis and drafted the article. All authors revised, read, and approved the final manuscript before submission. Emmanuel Ngwakongnwi is the guarantor for the manuscript, and all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## References

1. Forques É, Landry R, Boudreau J. Identifying francophones: an analysis of definitions based on census variables. Moncton (NB): Canadian Institute for Research on Linguistic Minorities; 2009. Available from: [www.icrml.ca/en/site\\_content/92-rodrique-landry/883-qui-sont-les-francophones-analyse-de-dnitions-selon-les-variables-du-recensement](http://www.icrml.ca/en/site_content/92-rodrique-landry/883-qui-sont-les-francophones-analyse-de-dnitions-selon-les-variables-du-recensement) (accessed 2011 Jun 10).
2. Table A-5. Population of French mother tongue, Canada, provinces, territories and Canada less Quebec, 1996 to 2006. Ottawa (ON): Statistics Canada; 2009. Available from: [www12.statcan.ca/census-recensement/2006/as-sa/97-555/table/A5-eng.cfm](http://www12.statcan.ca/census-recensement/2006/as-sa/97-555/table/A5-eng.cfm) (accessed 2011 Jun 10).
3. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS* 2005;19 Suppl 2:S67–S72.
4. Johnston LG, Sabin K, Hien MT, Huong PT. Assessment of respondent driven sampling for recruiting female sex workers in two Vietnamese cities: reaching the unseen sex worker. *J Urban Health* 2006;83 Suppl 1:16–28.
5. Wang J, Falck RS, Li L, Rahman A, Carlson RG. Respondent-driven sampling in the recruitment of illicit stimulant drug users in a rural setting: findings and technical issues. *Addictive Behav* 2007;32(5):924–937.
6. Duncan DF, White JB, Nicholson T. Using Internet-based surveys to reach hidden populations: case of nonabusive illicit drug users. *Am J Health Behav* 2003;27(3):208–218.
7. Quan H, Wong A, Johnson D, Ghali WA. The public endorses collection of ethnicity information in hospital: implications for routine data capture in Canadian health systems. *Healthc Policy* 2006;1(3):55–64.
8. Penrod J, Preston DB, Cain RE, Starks MT. A discussion of chain referral as a method of sampling hard-to-reach populations. *J Transcult Nurs* 2003;14(2):100–107.
9. Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Soc Probl* 2002;49(1):11–34.
10. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl* 1997;44(2):174–199.
11. Southern DA, Lewis S, Maxwell CJ, Dunn DR, Noseworthy TW, Corbett G, et al. Sampling 'hard-to-reach' populations in health research: yield from a study targeting Americans living in Canada. *BMC Med Res Methodol* 2008;8:57.

12. Kral AH, Malekinejad M, Vaudrey J, Martinez AN, Lorvick J, McFarland W, et al. Comparing respondent-driven sampling and targeted sampling methods of recruiting injection drug users in San Francisco. *J Urban Health* 2010;87(5):839–850.
13. Platt L, Wall M, Rhodes T, Judd A, Hickman M, Johnston LG, et al. Methods to recruit hard-to-reach groups: comparing two chain referral sampling methods of recruiting injecting drug users across nine studies in Russia and Estonia. *J Urban Health* 2006;83 Suppl 1:39–53.
14. Andersen B, Østergaard L, Møller JK, Olesen F. Effectiveness of a mass media campaign to recruit young adults for testing of *Chlamydia trachomatis* by use of home obtained and mailed samples. *Sex Transm Infect* 2001;77(6):416–418.
15. Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. *J Med Internet Res* 2008;10(5):e50. Erratum in: *J Med Internet Res* 2009;11(1):e2.
16. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. *J Med Internet Res*. 2005;7(1):e6.
17. Kendall C, Kerr LR, Gondim RC, Werneck GL, Macena RH, Pontes MK, et al. An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. *AIDS Behav* 2008;12(4 Suppl):S97–S104.
18. Abdul-Quader AS, Heckathorn DD, McKnight C, Bramson H, Nemeth C, Sabin K, et al. Effectiveness of respondent-driven sampling for recruiting drug users in New York City: findings from a pilot study. *J Urban Health* 2006;83(3):459–476. Erratum in: *J Urban Health* 2008;85(1):148.
19. Reisner SL, Mimiaga MJ, Johnson CV, Bland S, Case P, Safren SA, et al. What makes a respondent-driven sampling “seed” productive? Example of finding at-risk Massachusetts men who have sex with men. *J Urban Health* 2010;87(3):467–479.
20. *Canadian Community Health Survey (CCHS) cycle 3.1 (2005). Public use microdata file (PUMF) user guide*. Ottawa (ON): Statistics Canada; 2006. Available from: [http://prod.library.utoronto.ca:8090/datalib/codebooks/cstdli/cchs/cycle3\\_1/Documentation/GUIDE\\_E.pdf](http://prod.library.utoronto.ca:8090/datalib/codebooks/cstdli/cchs/cycle3_1/Documentation/GUIDE_E.pdf) (accessed 2011 Jun 3).
21. Francophone community profile of Alberta. Ottawa (ON): Fédération des communautés francophones et acadienne du Canada; 2009. Available from: [www.acfa.ab.ca/Documents/alberta\\_en.pdf](http://www.acfa.ab.ca/Documents/alberta_en.pdf) (accessed 2014 Sep 29).
22. *Canadian Community Health Survey cycle 3.1. Final questionnaire*. Ottawa (ON): Statistics Canada; 2005. Available from: [www23.statcan.gc.ca/imdb-bmdi/pub/instrument/3226\\_Q1\\_V3-eng.pdf](http://www23.statcan.gc.ca/imdb-bmdi/pub/instrument/3226_Q1_V3-eng.pdf) (accessed 2011 Jun 3).
23. Morse JM. *Qualitative research methods for health professionals*. 2nd ed. Thousand Oaks (CA): Sage Publications; 1995.
24. Volz E, Wejnert C, Degani I, Heckathorn DD. *Respondent-driven sampling analysis tool (RDSAT)*. Version 5.6. Ithaca (NY): Cornell University; 2007.
25. *Canadian Community Health Survey cycle 3.1*. Ottawa (ON): Statistics Canada; 2007. Available from: [www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=1630&InstalD=22642](http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=1630&InstalD=22642) (accessed 2014 Sep 30).

---

**Published:** 14 October 2014

**Citation:** Ngwakongnwi E, King-Shier KM, Hemmelgarn BR, Musto R, Quan H. Comparison of sampling methods for hard-to-reach francophone populations: yield and adequacy of advertisement and respondent-driven sampling. *Open Med* 2014;8(4):e120–e129.

**Copyright:** Open Medicine applies the Creative Commons Attribution Share Alike License, which means that anyone is able to freely copy, download, reprint, reuse, distribute, display or perform this work and that authors retain copyright of their work. Any derivative use of this work must be distributed only under a license identical to this one and must be attributed to the authors. Any of these conditions can be waived with permission from the copyright holder. These conditions do not negate or supersede Fair Use laws in any country. For more information, please see <http://creativecommons.org/licenses/by-sa/2.5/ca/>.