

**OPERATIONAL AND METHODOLOGICAL LESSONS LEARNED FROM THE  
2003 JOINT CANADA/UNITED STATES SURVEY OF HEALTH**

Catherine Simile, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, Maryland 20782  
Ed Rama, Statistics Canada, Tunney's Pasture, Ottawa, Ontario, Canada K1A 0T6

**Introduction<sup>1</sup>**

This paper focuses on the operational and methodological lessons learned from the 2003 Joint Canada/United States Survey of Health, a survey jointly conducted by Statistics Canada and the U.S. National Center for Health Statistics (NCHS), a center of the Centers for Disease Control and Prevention (CDC). Specifically, it will focus on the lessons learned in three areas: cognitive testing and translation, sampling and data collection and data release and analysis. Specifically, the Joint Canada/United States Survey of Health (JCUSH) afforded new ways of doing cognitive testing, and provided the first opportunity to implement new U.S. national translation guidelines. Working collaboratively on the project revealed differences in each nation's staff's understanding of the concept "clean sample" and approaches to resolving sample. Cooperativeness of the U.S. and Canadian populations in responding was also different. Although legal requirements regarding the protection of human subjects were virtually the same in the two countries, differences in implementing these legal requirements impacted the day-to-day data collection efforts. Finally, legal restrictions on access to data made it difficult to collaborate on analysis.

**Cognitive Testing and Translation**

***Cognitive Testing***

One of the purposes of collaboration is cross-fertilization. The cognitive testing and translation phases of the JCUSH afforded several windows of opportunity to do business differently. The normal practices of each nation differed. In the U.S., for example, the usual practice is for the cognitive staff to conduct individual on-site one-on-one interviews, while the usual practice in Canada is to conduct focus groups in English in Ottawa and in French in Montreal. For the JCUSH, the staffs jointly decided to combine the approaches. They

conducted one-on-one interviews in both agency laboratories and in other locations within each country, and they conducted focus groups. One of the most important lessons learned from doing testing this way was that staff was forced to think differently about comparability. In fact, more differences in understanding the questionnaire and the survey process itself were found between the samples within each country than between the countries. Specifically, there were more differences between the predominantly poor rural Mississippi sample and the predominantly middle class sample in the Washington, D.C. area and more similarities between the Washington, D.C. sample and the middle class Ottawa sample. Discovering these differences has led the U.S. cognitive staff to a more regular program of testing questionnaires with diverse populations in locations other than the Washington, D.C. area.<sup>2</sup>

***Translation***

The survey was conducted in three languages: English (both countries), French (Canada only, required by law) and Spanish (U.S. only, customary). As with all translation projects, the goal of this was to ensure that questions meant the same thing in all three languages. In 2001, a U.S. Census Bureau expert panel proposed guidelines for preparing translations for Federally conducted surveys. Specifically, they suggested 10 steps.<sup>3</sup> The

<sup>2</sup> For more discussion of the cognitive testing results and the implications for comparability, please see Kristen S. Miller, 2003, Conducting Cognitive Interviews to Understand Question-Response Limitations among Poorer and Less Educated Respondents. *American Journal of Health Behavior*. 27(S3), Supplement 3.

<sup>3</sup> For a fuller analysis on how well the JCUSH performed on each of the 10 steps, please see Deborah Rose and Manuel de la Puente, 2003, Translating the Joint Canada/U.S. Health Survey into Spanish. Presented at the *Federal Conference on Survey Methodology*, Washington, D.C., November 18-19. Paper can be accessed at [www.fcsm.gov](http://www.fcsm.gov).

<sup>1</sup> The authors wish to thank Lorna Bailie of Statistics Canada, and Stephen Blumberg, Jane Gentleman and Jennifer Madans of the National Center for Health Statistics for their contributions to this presentation.

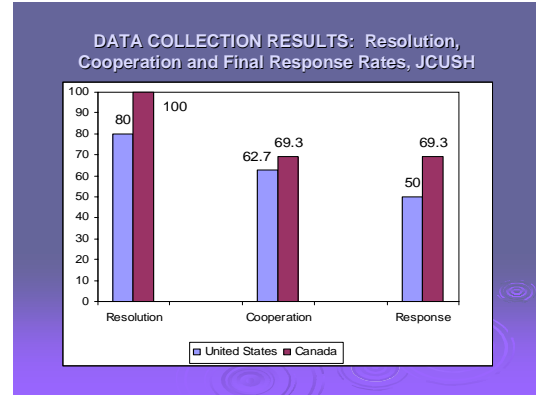
JCUSH staff was successfully able to implement six of the 10 steps. One of the most significant accomplishments was a bilingual review in which survey design and topic experts, bilingual interviewers (Spanish and English), translators (Spanish) and a French-speaking Canadian methodologist reviewed the three translations. The JCUSH experience with these guidelines was valuable because it demonstrated that following these guidelines is not easy, but that they can be very helpful in achieving comparable translations. With a little more time and money, all 10 steps could have been completed.

**Sampling and Data Collection**

One of the most crucial facets of any collaboration is the identification and communication of assumptions. In the best of circumstances this is difficult to accomplish, and in long-distance relationships, such as the one between the U.S. and Canadian staffs, it is even more difficult. For the most part, communication was successful, but in hindsight there were several areas where more care could have been given. Specifically, the definition of a “clean sample” was different in each country yet the differences were poorly communicated. There were also different U.S needs regarding the resolution of cases. In addition, while legal requirements are similar, differences in implementing them impacted the discretion and authority of the data collection staff.

**Data Collection Results**

The following graph can be used as a framework for discussing data collection problems. Three types of rates are shown. The one furthest to the left, the resolution rate, is the proportion of telephone numbers with a known eligibility, i.e., known to be out of scope or eligible for subsequent interview. In the case of JCUSH, an eligible line was a household with a working telephone landline. The final resolution rate for Canada was 100%, but 80% for the U.S. The middle bars, the cooperation rates, are the proportion of eligible cases that are completed. For the Canadian sample the cooperation rate was 69.3%, for the U.S. sample, 62.7%. The bars furthest to the right, the response rate, is calculated by multiplying the resolution rate by the cooperation rate. For the U.S., the final response rate was 50%, for Canada, 69.3%. Note that because Canada has a 100% resolution rate, response rates and cooperation rates are the same, 69.3%.



Difficulties in monitoring these rates during data collection might have resulted in lower than shown completion rates.

**Resolving Cases**

Sampling methodologists for each country determined the number of telephone lines necessary to reach intended sample sizes. Once that step had been accomplished, the sample was to be sent to Statistics Canada because data collection for both the U.S. and Canadian sample was to be conducted by Statistics Canada interviewers. The U.S. staff promised Statistics Canada a “clean sample,” but there was a significant misunderstanding as to what that meant. The following table summarizes the sample size determinations made by the sampling methodologists.

	# of lines selected	Targeted sample size
United States	32,009	5,000
Canada	10,334	3,500

Once the total number of telephone lines necessary was determined, in the United States, this sample was selected using the GENESYS sampling system, a commercial vendor commonly used by NCHS. When GENESYS draws a sample, they perform a screening process by which they are able to remove a proportion of business and non-working numbers from the sample. GENESYS removed 1/3 of the telephone numbers (11,476) from the initial sample after completing their screening process. The resulting 2/3 of the sample, 20,524 telephone numbers, was sent to Statistics Canada. The U.S. staff considered that the sample had been “cleaned.”

This process led the Statistics Canada staff to believe that because GENESYS had

determined that 1/3 of the initial sample was out of scope, the remaining 2/3 of the sample was made up only of working residential numbers. Because the field staff was expecting a sample comprised only of working residential numbers, they did not realize that a significant portion of the interviewing staff time would be spent further “cleaning” the sample, that is, determining whether a number was eligible or out of scope. While the sample was “clean” by U.S. standards and further cleaning is assumed to be needed, it was not “clean” by Canadian standards.

Further exacerbating this misunderstanding is the fact that Canadian field staff normally expect to resolve 100% of the cases with means other than directly working the sample. In the U.S., the sample is never expected to be resolved 100% and it is expected that interviewing staff will spend a proportion of their time determining the eligibility of the sample. In Canada, numbers can be verified to be working and residential by calling the phone companies; in the U.S., working residential numbers cannot be verified in the same manner—there are too many phone companies. Determining the eligibility of a phone number is a significant interviewer task.

Because of the lack of clear communication between the two project staffs, the Canadian interviewing staff received a sample less clean than they had expected, did not know they would be unable to depend on their usual practice of verifying phone numbers, and had to spend considerable unexpected time “cleaning” the U.S. sample.

### ***Cooperation***

The cooperativeness of the Canadian and U.S. sample in responding also posed problems for data collection. Although the final cooperation rates are more similar than the staff expected given the difficulty working the U.S. sample (for the Canadian sample the rate was 69.3%, for the U.S., 62.7%), other evidence points to a relative lack of cooperation in the U.S. It can first be seen in the resolution rate. The proportion of unresolved telephone numbers in the U.S. was 20%. Further analysis indicates that of the unresolved numbers, 31% of those resulted from working lines not being answered. (This is often called “ring no answer”) Although there is no way to prove this, the hypothesis is that because of increased telemarketing in the U.S., residents are employing screening technologies such as caller

ID. There is no way to know whether these numbers are businesses or residences, so they had to be classified as unresolved. An additional 9% resulted from unresolved answering machine messages. Contact was made and a message explaining the survey was left, but staff was unable to determine from the outgoing message whether the answering machine belonged to a residence or a business. In either case, the refusal to return the call indicates a lack of cooperativeness of the U.S. population—whether residential or business. Another 8% likely resulted from persons hanging up during the introduction, another indication that the U.S. population does not want to cooperate in a survey.<sup>4</sup>

The relative lack of cooperation is also evident in the differing refusal and break-off rates. In the U.S., the percent of the eligible sample that refused to participate at all was 21%, and the percent that broke off the interview was 11%. In Canada, the percent who refused was 14%, and the percent that broke off once the interview started was less than .01%. These findings indicate that while the U.S. refusal rate was indeed much higher than the Canadian rate, the Canadian rate wasn’t exactly low. However, these findings clearly indicate that once a Canadian agreed to participate in the survey, the respondent almost always stayed with it until the end, whereas in the United States, respondents were less willing to complete the survey even if they agreed initially.

### ***Collection Monitoring***

Careful monitoring of unresolved telephone numbers, that is, being able to identify whether an unresolved number is a “ring no answer,” an “always busy” or a “hang-up during introduction” during data collection is extremely important in the U.S. for keeping track of progress. In Canada, because unresolved numbers are sent out for “research” to determine their eligibility, monitoring the exact nature of the problem is unnecessary. In the case where 100% of the cases can be resolved, it is not necessary to develop a monitoring system that records much detail on the progress of resolution. The monitoring system developed for Statistics Canada works well for a sample

---

<sup>4</sup> The term “likely” is used because the system used to monitor data collection and assign temporary outcome codes does not allow for this classification. This assessment was made post-collection.

where the emphasis is on cooperation of the eligible sample, and not as much on determining the eligibility of the sample. It didn't work so well for the U.S. situation where both resolution rates and cooperation rates need to be carefully monitored. These differences were not realized ahead of time and created tremendous frustration on both sides of the border. Daily reports were produced and discussed by project staff, but the staff soon realized that it was impossible to really keep track of progress on the U.S. sample without a coding scheme sufficient to understand the unresolved cases. Because the staff had difficulty identifying the nature of the problem, it was difficult to determine where to apply resources

#### ***Discretion and authority of data collection staff***

The considerable difference between the discretion and authority given to interviewers in the respective countries was evident primarily in the deployment of the advance and follow-up letters. Canadian interviewers were frustrated by the delay in the mailing of advance introductory letters. Statistics Canada and the U.S. staff had agreed that the U.S. staff would control all correspondence to the U.S. sample. This allowed for greater cost savings and ensured that the postmark would be from the U.S. As a result of an unexpected problem associated with getting initial human subjects clearance there was a two week delay in sending out the advance letters. This contributed to problems at the start of collection with Canadian interviewers having difficulty in proving the legitimacy of the call to a U.S. sample. Furthermore, because correspondence was controlled by the U.S. staff, the Canadian interviewers could not immediately send out a letter to those who asked for one. Their only recourse was to have the respondent call a predetermined U.S. contact number.

This problem with control centralized in the U.S. again became a problem a little later in the data collection process. Once lack of cooperation—in the form of ring no answer and high refusal rates—was identified as a major issue, a decision was made to send out non-response follow-up letters. In Canada, the collection offices usually have the discretion and the authority to send these letters as part of normal collection operations. Often a letter is sent immediately by the senior collection staff in the case of non-response. Such discretion and authority are not usual practice in the U.S. Rather, the mailing of any letters must be specified as part of the initial protocol. Any

change requires appropriate review by the U.S. investigators and appropriate review panels. Because of these requirements, there were major delays in receiving approval to send out the U.S. letters. Because Canadian interviewers have found such letters to be useful in converting reluctant respondents in past surveys, these delays led to considerable frustration among the interviewers and other data collection staff. That the centralized control of letters would have such an impact was not clear to either staff at the beginning of the project.

#### **Data Release and Analysis**

The project staff agreed to follow Statistics Canada's usual procedure of both editing and reviewing the microdata for public release and producing an analytical report to be released at the same time. The microdata were required to undergo independent disclosure review in each country. In order to maximize comparability of the data, the staffs of the two agencies worked closely to make sure that the variables both maximized the usability of the collected data, and were constructed exactly the same for both samples. The project was very successful in this. The staffs also agreed to collaboratively release an analytic report. However this goal was hampered by the privacy laws governing the release of data associated with the Canadian sample. It became evident that U.S. analysts would be unable to work with Canadian data until they were publicly released, a restriction that restricted the U.S. staff's ability to truly collaborate. (The Canadian staff, because they collected the data, on the other hand, had access to data from both the U.S. and the Canadian sample) However, the Canadian staff shared aggregate data and findings and conducted analyses suggested by the U.S. staff, and discussions between the respective staffs were held regularly.

#### **Summary**

The JCUSH project taught the respective staffs some very valuable lessons in the conduct of an international survey. New ways of doing cognitive testing were employed and new guidelines on translation were tested. Further, conducting analysis at the same time that data are being edited provides significant opportunity to ensure that data are accurate when they are released. The JCUSH was similar to all projects in that advance communication and assessment of assumptions would be crucial to the smooth operation of a joint survey. Knowing

ahead of time the differences in the meaning of clean sample, the differences in the cooperativeness of the respective samples, and the need to create an applicable monitoring system would have contributed to smoother data collection operations. While the end product may not have been significantly different, a considerable amount of time and frustration could have been saved. Finally, while legal requirements may be similar, the differences in implementing these requirements are not insignificant in their impacts.

### References

Miller, Kristen S. 2003. Conducting Cognitive Interviews to Understand Question-Response Limitations among Poorer and Less Educated Respondents. *American Journal of Health Behavior*. 27(S3), Supplement 3.

Rose, Deborah and Manuel de la Puente. 2003. Translating the Joint Canada/U.S. Health Survey into Spanish. Presented at the Federal Conference on Survey Methodology, Washington, D.C., November 18-19.