



Statistica  
Canada

Statistique  
Canada

Canada



Statistics Canada  
[www.statcan.gc.ca](http://www.statcan.gc.ca)



# The National Household Survey: Evaluating Data Quality in a Large Voluntary Survey

Sander Post  
Statistics Canada



# Outline

- Overview of the National Household Survey
- Record Linkage and Analysis
- User Reaction
- Conclusions



# Overview of the National Household Survey

- Canada conducts a Census every 5 years
- Up to 2006, 80% of dwellings got a short form questionnaire, the remaining 20% got a long form questionnaire – both were mandatory



# Overview of the National Household Survey

- In 2010, it was decided to cancel the mandatory long form Census for 2011 and replace it with a voluntary long form survey, sent to about 1/3<sup>rd</sup> of households
- The short lead time made it difficult to assess and estimate expected response rates for this new survey



# Overview of the National Household Survey

- After initial collection, we subsampled remaining nonrespondents, and achieved a weighted response rate of about 77%



# Overview of the National Household Survey – Major Questions

- Given the higher nonresponse, how reliable are the estimates at different geographic levels and in areas with different response rates?
- Will users follow our guidance?



# Linkage and Analysis

- One way to measure the impact of nonresponse is to link NHS respondents and NHS nonrespondents to the 2006 Census
- Given a  $1/5$  sample in 2006 and a  $1/3$  sample in 2011, we could expect to link up to  $1/15$  of the population, around 2 000 000 people



# Linkage and Analysis

- The 2011 Census and the 2011 NHS shared a common control database, the master control system (MCS)
- For NHS nonrespondents that were Census Respondents, we were able to easily obtain their names and their dates of birth





# Linkage and Analysis

- This linkage file allows us to analyse nonresponse bias in ways that are typically not possible
- Some data we expect to be constant, even for surveys conducted 5 years apart
  - Place of birth
  - Year of immigration



# Linkage and Analysis

- Some can change
  - Highest level of education
  - Labor force variables
  - Citizenship



# Linkage and Analysis

- Quality of linkage
  - Are 2006 linked and 2006 unlinked similar?
  - Of linked NHS respondents, are answers similar in 2006 and 2011?
  - Using information about 2011 NHS nonrespondents, can we generate indicators for nonresponse bias?



# Linkage and Analysis

## Comparing NHS & 2006 responses

Variable Name	Matching Percentage	Unmatched Percentage
Place of Birth	98.6%	1.4%
Place of Birth – Mother	96.5%	3.5%
Place of Birth – Father	91.9%	8.1%
Citizenship	93.5%	6.5%
Immigration Status	98.5%	1.5%
Visible Minority	95.7%	4.3%



# Linkage and Analysis

## Comparing NHS & 2006 responses

Variable Name	Matching Percentage	Unmatched Percentage
Aboriginal Status	98.1%	1.9%
Registered Indian	99.3%	0.7%
Highest Degree	66.1%	33.9%
Industry	47.4%	52.6%
Occupation	47.3%	52.6%
Language at work	86.5%	13.5%



# Linkage and Analysis

## Nonresponse bias indicators

$$\hat{B}_2 = \hat{\beta}_0 \left( \sum_{i \in S_R} (w_i - d_i) f_{Ri} y_i^{2006} - \sum_{i \in S_{\bar{R}}} d_i f_{\bar{R}i} y_i^{2006} \right)$$



# Linkage and Analysis

## Nonresponse bias indicators

Place of Birth	Bias Indicator	Total	Relative Bias Indicator
Canada	4 930 4	1 847 712 7	0.3%
China, SARs, and Taiwan	1 884	4 945 56	0.4%
United Kingdom & Republic of Ireland	-4 767	4 629 39	-1.0%
India	6 047	3 164 98	1.9%
Philippines	1 927 4	2 364 86	8.1%



# Overview of the National Household Survey – Major Questions

- Given the higher nonresponse, how reliable are the estimates at different geographic levels and in areas with different response rates?
  - For large areas, quite good. For areas of low response, we suppressed estimates based on nonresponse





# User Reaction

“Caution must be exercised when NHS estimates are compared with estimates produced from the 2006 Census long form, especially when the analysis involves small geographies.”

NHS User Guide



# User Reaction

“In 2006, for instance, the census found that two per cent of Ottawans routinely commute to work by bike; in 2011, figures released last week say, that was up to 2.2 per cent.

But is that a real change or are bike commuters just more likely to fill out the new household survey? What if they're less likely to fill it out and that's hiding a bigger jump in biking?”



# Overview of the National Household Survey – Major Questions

- Will users follow our recommendations to be cautious in making historical comparisons?
  - Users are using our data, but also being cautious in making historical comparisons.

# Conclusions

- Given the ability to link nonrespondents to historical responses, we can calculate nonresponse bias indicators
- When calculating nonresponse bias indicators, we must examine the data quality of individual variables. We cannot calculate reliable indicators for variables with other quality issues
- Data user are taking note of our recommendations in the use of 2011 NHS data



# Thanks to

Chantal Grondin

Dominic Grenier

Christian Olivier Nambeu

Laurent Roy

Scott McLeish