



Statistics
Canada

Statistique
Canada

Canada



Statistics Canada
www.statcan.gc.ca

2011 National Household Survey (NHS): design and quality

Margaret Michalowski

2014 National Conference

Canadian Research Data Center Network (CRDCN)

Winnipeg, Manitoba, October 29-31, 2014



Outline of the presentation

- Overview of the National Household Survey (NHS)
- Key data quality indicators



NHS – the context

- Statistics Canada conducted the NHS in 2011, at the same time as the Census
- It was the largest voluntary household survey ever conducted by Statistics Canada



NHS Collection: who did we survey?

Phase 1 of sample selection (the initial sample)

Objective: To ensure representative results at low geographic levels

- A large, random sample of 4.5 million dwellings was used, corresponding to approximately one-third of all dwellings in Canada
- Canvasser in remote and northern areas (100%)

NHS Collection: how did we reduce bias?

Phase 2 of sample selection (the sub-sample)

Objective: To control for potential bias and, to some extent, sampling variation

- A random sub-sample was selected from the remaining non-responding households (as of July 14, 2011)
- The allocation of sub-samples varied by geography and was a function of the level of non-response and the degree of population homogeneity.



NHS Collection: how did we do?

Response rate for the NHS at the Canada level

- 68.6% for occupied private dwellings (unweighted) and 77.2% (weighted)
- 44.7% completed the questionnaire on-line



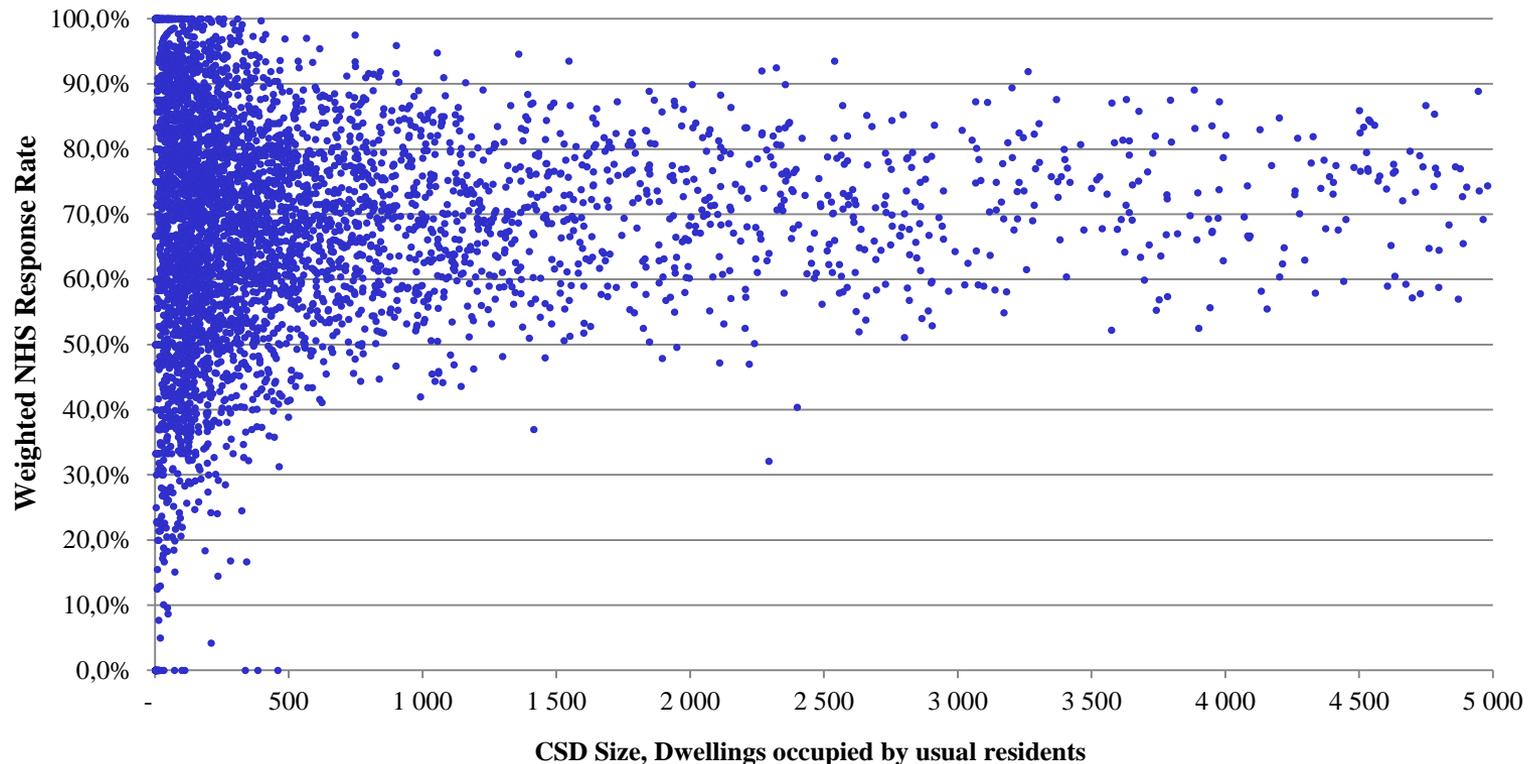
NHS collection response rate by province and territory, showing the unweighted and weighted rate

Province and territory	Unweighted Response rate (%)	Weighted Response rate (%)
Canada	68.6	77.2
Newfoundland & Labrador	63.3	72.5
Prince Edward Island	60.4	70.0
Nova Scotia	65.0	74.8
New Brunswick	63.9	74.2
Quebec	71.9	80.7
Ontario	67.6	76.3
Manitoba	69.1	76.3
Saskatchewan	63.8	73.1
Alberta	67.3	75.4
British Columbia	69.5	77.1
Yukon	64.9	72.7
Northwest Territories	83.9	83.8
Nunavut	76.3	76.3



NHS collection response rate by census subdivision, showing the population size

Distribution of Weighted NHS Response Rate by CSD Size, for CSDs with fewer than 5,000 dwellings occupied by usual residents





NHS Collection: the results

- **Within provinces and territories:** Response was good in all large urban population centres; varied by the size of municipalities (census subdivisions)
- **For returned questionnaires:**
 - Completeness of responses to questions related to respondents' demographic, socio-cultural, language, mobility and education characteristics are comparable to the 2006 Census long-form responses
 - Non-response was higher than in 2006 for the remainder (the last part of the questionnaire)

Response quality indicator: imputation rate for NHS estimates

The imputation rate is the proportion of respondents who did not answer a given question or whose response is deemed invalid and for which a value was imputed

NHS Question	Imputation rates %
#09 - Place of birth	2.0
#10 - Citizenship	2.3
#11 - Landed immigrant status	1.3
#12 - Year of immigration	12.5
#13 - Knowledge of official languages	0.7
#15 - Home Language	0.6
#16 - Mother Tongue	0.6
#17 - Ethnic origin	5.8
#18 - Aboriginal group	3.7
#19 - Population group	3.9
#20 - Registered or Treaty Indian status	4.7
#21 - Membership in a First Nation or Indian Band	3.8
#22 - Religion	4.4
#23 - Mobility 1 year ago	4.8
#24 - Mobility 5 years ago	6.9
#25A - Place of birth – Father	6.0
#25B - Place of mother – Mother	5.7
#27 - Secondary School diploma or equivalent	4.6
#28 - Registered Apprenticeship or other trades certificate or diploma	5.5
#29 - College, CEGEP or other non-university certificate or diploma	5.5

NHS Question	Imputation rates %
#30 - University certificate, diploma or degree	4.7
#31 - Major field of study	14.2
#32 - Location of study	12.1
#33 - Attendance at school	6.1
#34 - Hours Worked	6.7
#35 - On lay off or absent	10.5
#36 - New job to start	8.0
#37 - Look for work	7.8
#38 - Reason unavailable for work	10.3
#39 - When last worked	8.7
#40 - Industry	13.6
#41 - Place of work status	11.3
#42 - Occupation	13.6
#44 - Class of Worker	12.2
#45 - Incorporation status	8.1
#46 - Place of work location	13.0
#47A - Mode of Transportation	12.1
#47B - Vehicle occupancy	13.7
#48A - Time leaving	15.5
#48B - Duration of trip	14.8

NHS Question	Imputation Rates %
#49 - Language of work	12.9
#50 - Weeks worked in 2010	15.1
#51 - Full-time or part-time work	14.6
#E1 - Household maintainer	11.8
#E2 - Tenure	10.7
#E3 - Condominium Status	9.4
#E4A - Rooms	12.8
#E4B - Bedrooms	11.4
#E5 - Period of construction	13.5
#E6 - Condition of dwelling	10.7
#E8A - Electricity payment	19.5
#E8B - Fuel payment	19.0
#E8C - Water, other service payment	19.5
#E9A - Rent	14.3
#E9B - Subsidy status	13.9
#E10A - Mortgage payment	18.0
#E10B - Property taxes included in mortgage	17.2
#E10C - Property taxes	20.8
#E10D - Value of dwelling	21.2
#E10E - Condominium fee	22.8



Evaluation of the quality of NHS estimates

- Evaluation of quality involved comparisons with data from previous censuses, and other sources:
 - Comparisons were done mainly at higher levels of geography: Canada, provinces and territories
- For validation purposes, the NHS was linked to other data sources, for example to identify areas of potential non-response bias:
 - To the 2011 and 2006 Censuses
 - To some administrative files , e.g. Longitudinal Immigration Database

Indicator of the quality of NHS estimates: Global non-response (GNR)

- GNR combines complete non-response (household) and partial non-response (question) into a single rate. It is calculated for different geographic areas
- Why is the GNR used as indicator of quality?
- Similar approach as used for past long-form censuses
 - Assumption: Non-response level is correlated with non-response bias; hence, the GNR provides an indicator of potential non-response bias
 - A smaller GNR indicates a lower risk of non-response bias and, as a result, lower risk of inaccuracy
- NHS estimates are released for areas with a GNR of less than 50%
- Coefficient of variation is available to users to indicate estimates' variability



Global non-response rate (GNR) by province and territory

Global non-response rate of the 2011 National Household Survey, Canada, provinces and territories

Provinces and territories	Global non-response rate (%)
Canada	26.1
Newfoundland and Labrador	31.5
Prince Edward Island	33.5
Nova Scotia	28.3
New Brunswick	28.6
Quebec	22.4
Ontario	27.1
Manitoba	26.2
Saskatchewan	29.4
Alberta	27.5
British Columbia	26.1
Yukon	29.9
Northwest Territories	16.1
Nunavut	25.2

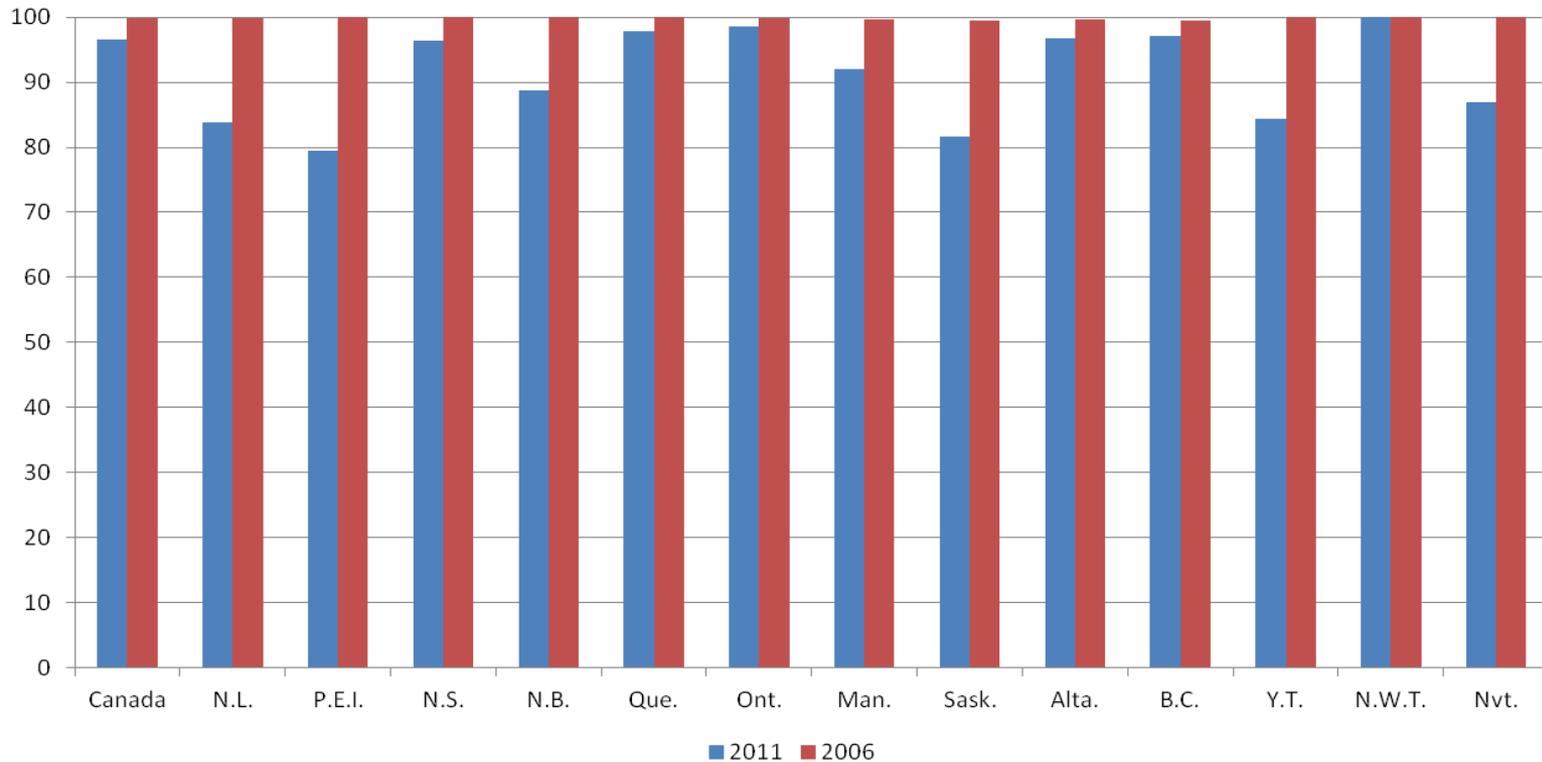


Dissemination of the NHS estimates

- Relative to 2006, there is a reduced level of data released for smaller geographic areas:
 - NHS estimates were released for **3,439 Census Subdivisions** (e.g., municipalities), down by about 1,100 CSDs from 2006
 - Proportionally, this represents **75% of all CSDs** in 2011 (97% in 2006)
 - NHS estimates were released for **97%** of the total population (99.8% in 2006)



Proportion of provincial/territorial population in CSDs for which NHS estimates are available (vs. 2006 Census)





Comparability to 2011 Census counts: an indicator of NHS estimates' quality

- For the same target population, a discrepancy vs. 2011 Census counts is an indication of the quality of the NHS estimates
- Calibration was used as a method to reduce (or eliminate) differences between the 2011 Census counts and the NHS estimates for common topics
- Final weight adjustment based on calibrated areas (some made up of several small municipalities) may lead to discrepancies between the NHS estimates and the census counts for small municipalities



Are NHS estimates affected by non-response bias?

- Non-response bias is a potential source of error for all surveys, including the NHS.
 - It increases as the response rate declines.
 - It arises when the characteristics of those who choose to participate in a survey are different than those who refuse.
- It is impossible to definitively determine how much the NHS may be affected by non-response bias. However, based on information from other data sources, evidence of non-response bias does exist for certain populations and for certain geographic areas.
- Generally, the risk of error increases for lower levels of geography and for smaller populations.



Examples of possible non-response bias

- **Aboriginal**: Estimates and trends from other data sources suggest that the Inuit population living outside of Inuit Nunangat is overestimated at the national level.
- **Ethnic origin**: The population born in the Philippines is overestimated at the national level (comparisons to administrative data from Citizenship and Immigration Canada).
- **Education**: Comparisons with other data sources suggest that the category 'university certificate or diploma below bachelor level' was over-reported in the NHS.
- **Income**: Low-income estimates from the 2011 NHS compared to previous censuses show markedly different trends than those derived from other surveys (such as the Survey of Labour and Income Dynamics) and administrative data (the T1 Family File).

Examples of possible non-response bias (continue)

- **Mobility**: The estimation of internal migrants from the 2011 National Household Survey are, in general, lower than the ones derived from administrative data.
- **Language**: Evidence suggests that the population who reported having Malayo-Polynesian languages as their mother tongue is overestimated at the national level. The population reported having Romance languages as their mother tongue appears to be underestimated.
- **Housing**: Comparison of the homeownership rate in Alberta to that from the 2010 Survey of Labour Income Dynamics (SLID) showed the NHS rate was higher with a statistical significance. The national rate and the rate for other provinces were not statistically different.
- **HOWEVER, the numerous data evaluations conducted support the general reliability of the data at the national, provincial and territorial levels.**



Comparability to the 2006 Census long form

- **Two different target populations:** the 2006 Census long form includes usual residents in collective dwellings and persons living overseas whereas the NHS excludes them
- **Two different methodologies:** the NHS estimates are derived from a voluntary survey and are therefore subject to potentially higher non-response error than those collected in the 2006 Census long form.



Information on Data Quality

- A data quality indicator (GNR) is provided with each geographic area published
- *National Household Survey User Guide*
- Reference Guides for each NHS topic (e.g. Ethnic Origin)
- Standard notes about quality attached to products