Balancing Data Quality with a More Efficient Use of Collection Resources−Experiences from Statistics Canada CATI Surveys

Amélie Lévesque, Ann Lim, Christian Nadeau, Owen Phillips
Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6

Abstract

In an effort to reduce costs and perceived respondent burden, Statistics Canada adopted in 2007 a maximum of 25 call attempts for social surveys that collect data through Computer Assisted Telephone Interview. In the time since its implementation, a number of surveys have successfully petitioned for temporary exemptions from the call limit, often presenting as argument the impact such a restriction would have on response rates. In the summer of 2008, a coordinated effort was made among a number of longitudinal and cross-sectional surveys, including the Survey of Labour and Income Dynamics and the Youth in Transition Survey, to evaluate the impact that a cap of 25 call attempts would have on the quality of the resulting estimates. Non-response was simulated in the existing samples assuming a cap of 25, survey weights were recalculated and a large number of key estimates were compared. The study found that in general, there was a non-negligible impact to the quality of estimates and recommended an indefinite limit of 40 call attempts. The present document will examine in greater detail the results of the study and provide additional evidence in support of the higher cap.

Key Words: Call attempt, Call limit, Response rate, Accuracy, Reliability

1. Introduction

In recent years, much attention has been given to the effective management of collection resources at Statistics Canada. Despite new initiatives introduced to evaluate the collection process in real-time and implement corrective actions (Hunter and Carbonneau, 2005; Laflamme *et al*., 2008) social surveys have seen a general decline in response rates (Dufour, 2009).

In an effort to reduce costs and perceived respondent burden, Statistics Canada adopted in 2007 a maximum of 25 call attempts for social surveys that collect data through Computer Assisted Telephone Interview (CATI). Since its implementation, a number of surveys have successfully petitioned for temporary exemptions from the call limit. Much of the evidence presented examined only the impact such a restriction would have on already dwindling response rates, showing some evidence of potential bias in estimates produced for young singles and immigrants who that are typically harder to contact (Lévesque and Poulin, 2007; Lévesque, 2008). In order to sustain these exemptions, surveys were asked to show the impact on the quality of estimates (in terms of potential bias and precision).

In the summer of 2008, a coordinated effort was made among a number of surveys, including the Survey of Labour and Income Dynamics (SLID) and the Youth in Transition Survey (YITS), to evaluate the impact that a cap of 25 call attempts would have on the accuracy and precision of the resulting estimates. This document will present the methodology behind the evaluations done for

both surveys and examine in some detail the results of the study, providing additional evidence in support of a higher cap. More information is provided in Lévesque et al. (2010).

1.1 About the Surveys

1.1.1 Survey of Labour and Income Dynamics (SLID)

The SLID is a longitudinal survey used to measure changes in the economic well-being of Canadians and the factors that may influence these changes. The survey is also used to produce cross-sectional estimates of family and personal income, income inequality and low income measures.

The survey follows longitudinally a panel of approximately 17,000 respondent households from the Labour Force Survey (LFS). All members of the LFS household become members of the longitudinal panel. The panel is followed for six years, with a new panel introduced every three years. For the purpose of cross-sectional estimation, non-longitudinal cohabitants are interviewed.

The SLID annually collects data about work, income, family circumstances as well as socio-demographic information. Individuals 16 years of age and older in a selected household are asked questions on labour and income. All questions refer to activity in the calendar year (the reference year) preceding data collection. Proxy response is allowed. To reduce the response burden, respondents can authorize Statistics Canada to use their income tax return in lieu of responding to the income questions.

1.1.2 Youth in Transition Survey (YITS)

The YITS is a longitudinal survey, used to examine transitions to adulthood, including education, training and entry to the labour market. The survey follows two groups: Cohort A and Cohort B. The target population for Cohort A consists of people born in 1984 and who were attending any form of educational establishment in the ten provinces in 1999. For Cohort B the target population consists of people born between 1979 and 1981. Cohort A had an initial sample size of approximately 39,000 students while Cohort B, had approximately 29,000 people.

The first data collection was in 2000 with subsequent collections occurring every two years. The 6th cycle of data collection for Cohort A is currently in the field. Cohort B was followed for only 5 cycles. Because only the Cohort A sample was subject to follow-up at the time of this study, the impact of the 25-cap is considered only for this group.

The survey uses a monotonic design, i.e. in a given cycle, only the respondents from the previous cycle are contacted for follow-up.

1.2 Definition of a Call Attempt

This study uses the concept of a call attempt. Each time an interviewer accesses a case to make contact with a respondent, the attempt counter is incremented. Note that tracing activity is not considered a call attempt, but all other call outcomes are, including answering machine, ring no answer and busy dial. The counter is used to determine whether or not a case has reached the maximum number of call attempts allowed. The information regarding the number of call attempts is taken from the Blaise Transaction History (BTH) file generated by the CATI application.

1.3 Response Rates and Call Attempts

New initiatives and collection procedures are frequently introduced to make a more effective use of collection resources. For example, many surveys have used *time slices* and *time slice sets*[1] in an attempt to optimize the probability of contacting a particular household. Despite these improvements, a decline in response rates and a larger proportion of high-effort (i.e. more than 25 call attempts) respondents has been observed in CATI social surveys.

For example, for the SLID (Tables 1 and 2) two trends seem to emerge. First, the response rates among high-effort SLID households, while good, are considerably lower than for households that required fewer call attempts. Second, the percentage of high-effort households has increased over the last four collections.

Table 1: Percentage of Eligible Households Classified as Respondent at Collection
SLID 2005 to 2008

Collection year	Number of attempts		
	≤ 25	**26 - 40**	**> 40**
2005	92.4	82.4	73.9
2006	86.6	67.6	52.3
2007*	86.0	67.6	–
2008*	87.0	61.6	–

* The number of call attempts was capped at 40.

Table 2: Distribution (%) of Respondent Households by Number of Call Attempts
SLID 2005 to 2008

Number of attempts	Collection year			
	2005	**2006**	**2007***	**2008***
≤ 25	94.2	93.0	92.5	91.3
26 - 40	3.5	4.5	7.5	8.7
> 40	2.2	2.5	–	–

* The number of call attempts was capped at 40.

For the YITS, similar trends emerge (Table 3). The percentage of responding high-effort households is even lower than for SLID, perhaps owing to the younger age group targeted by YITS. The percentage of respondents requiring between 26 and 40 attempts has increased from Cycle 2 to Cycle 5, having more than tripled in that time (Table 4).

Table 3: Percentage of Eligible Households Classified as Respondent at Collection
YITS Cycles 2 to 5

Cycle (collection year)	Number of attempts		
	≤ 25	**26 – 40**	**> 40**
Cycle 2 (2002)	91.7	72.7	56.0
Cycle 3 (2004)	88.5	48.1	29.3
Cycle 4 (2006)	86.8	69.2	41.8
Cycle 5 (2008)*	85.8	38.3	–

* The number of call attempts was capped at 40.

[1] *Time slices* represent a partitioning of the week based on time of day and the day of week. Contact attempts are allocated to time slices in order to optimize the probability of contacting a particular household. *Time slice sets* are defined based on the demographic characteristics of a household. They are used to determine the allocation of call attempts to the time slices.

Table 4: Distribution (%) of Respondent Households by Number of Call Attempts
YITS Cycles 2 to 5

Number of attempts	Cycle (collection year) Cycle 2 (2002)	Cycle 3 (2004)	Cycle 4 (2006)	Cycle 5* (2008)
Previous deaths	0.0	0.1	0.3	0.6
≤ 25	96.8	94.8	91.2	91.7
26 – 40	2.2	4.6	5.7	7.8
> 40	1.0	0.6	2.8	0.0

* The number of call attempts was capped at 40.

The figures presented in this section give a picture of the changes in respondent behaviour over time, but give little indication of how capping the number of call attempts might affect the quality of the estimates. The next sections describe the methods used to assess the impact the cap would have and present the results of the study.

2. Methods

Nonresponse was simulated in the existing samples assuming a limit of 25 call attempts and response codes were derived. Based on the new response codes, the survey weights were recalculated. In SLID, specific actions are taken in an effort to reduce respondent burden and make the most effective use of collection resources. For example, households coded as hard refusal in two consecutive years will not be contacted in future cycles. This procedure was also applied for the purpose of the simulation. Response codes were recalculated for collection years 2005, 2006 and 2007. In doing so, we were able to examine the impact on the cross-sectional and longitudinal response rates. The cross-sectional weights[2] for 2007 collection year were then recalculated, as the majority of the standard outputs are cross-sectional in nature. For the YITS, response codes were rederived for cycles 2, 3 and 4 of Cohort A, and the response rates and longitudinal weights for Cycles 2 through 4 were recalculated.

Following the reweighting exercise, a large number of key estimates were compared for both surveys. There were two dimensions to this evaluation. The first step was to compare the estimates produced based on the production weights with the estimates produced using the weight calculated under the assumption of the 25-cap on call attempts. The assumption here is that any observed (significant) differences are indicative of a potential bias in the resulting estimates.[3] This involved first testing the hypothesis $H_0 : \theta_{prod} - \theta_{cap25}$ versus the alternative that the two estimates are different. Variances $\hat{v}_B\left(\hat{\theta}_{prod} - \hat{\theta}_{cap25}\right)$ were estimated using the Rao-Wu bootstrap method. In the case of SLID, where the differences were significant, the relative differences (RD) $\dfrac{\hat{\theta}_{prod} - \hat{\theta}_{cap25}}{\hat{\theta}_{prod}}$ were considered.

[2] A detailed description of the SLID weighting methodology can be found in LaRoche (2007).
[3] This assumption could be false, in that for some estimates, the loss of sample due to the 25-cap could improve the representativity of the responding sample. However, in general, we believe this assumption to be valid.

The second part of the evaluation involved looking at the impact on the reliability of the estimates. For each survey, publication guidelines exist for releasing estimates. The reliability of an estimate is based on two factors: the size of the respondent sample in a particular domain, and the coefficient of variation (CV) of the estimate $\frac{\sqrt{\hat{v}_B(\hat{\theta})}}{\hat{\theta}}$. Domain sample sizes and CVs were recalculated for all estimates and changes in the reliability of the estimates with respect to the production estimates were noted.

For the SLID, a total of 127,193 estimates included in the publication *Income Trends in Canada*[4] were calculated using the production and 25-cap cross-sectional weights. These included means, medians and Gini coefficients for a number of income measures and domains. For the YITS, 3,684 estimates, using the variables and domains of greatest interest to researchers, were calculated using the production and 25-cap longitudinal weights. These included frequencies for a number of categorical variables and means for scales.

3. Results

Results are presented separately for both surveys.

3.1 SLID

3.1.1 Response rate

Tables 5 and 6 present the household- and person-level cross-sectional response rates, and the longitudinal response rates, based on the final response codes used in production, and the rederived response codes assuming a cap of 25 call attempts.

Table 5: SLID Response Rates, Based on Final Response Codes, 2005 to 2007 (%)

Response rate	Collection year		
	2005	2006	2007*
Cross-sectional, household	74.7	76.1	74.9
Cross-sectional, person	73.2	73.3	72.5
Longitudinal, person			
Panel 4 (began 2003)	78.3	75.0	71.6
Panel 5 (began 2006)	…	78.8	80.6

* The number of call attempts was capped at 40.

Table 6: SLID Response Rates, Based on 25-cap Response Codes, 2005 to 2007 (%)

Response rate	Collection year		
	2005	2006	2007*
Cross-sectional, household	70.4	70.6	69.0
Cross-sectional, person	69.3	68.5	67.2
Longitudinal, person			
Panel 4 (began 2003)	73.7	69.2	65.6
Panel 5 (began 2006)	…	73.2	74.4

* The number of call attempts was capped at 40.

[4] These include only the estimates that are deemed publishable under the SLID guidelines for publication. To consult the *Income Trends in Canada* tables, please visit http://www.statcan.gc.ca/pub/13f0022x/2006000/5213032-eng.htm.

In general, the cross-sectional rates drop between 4 and 6%, with both the household and person level response showing increasing differences over the three years under study. A similar trend is observed for the longitudinal response rates. While these differences in response rates seem important, more analysis is needed to assess the impact of reduced response rates on the accuracy and reliability of the estimates. The next sections will examine these issues.

3.1.2 Accuracy of the Estimates

For the set of estimates considered, the proportion of estimates found to be significantly different at the 5% level when comparing the two scenarios is 9.8%. However, there are some important differences by variable. Table 7 presents descriptive statistics for the five variables most affected by the call limit.

Table 7: SLID Variables Most Affected by the Call Limit, Reference Year 2006

Variable[5]	Test result	Number of estimates	Percentage	Absolute relative difference (%)	Minimum relative difference (%)	Maximum relative difference (%)
One	Significant	450	23.4%	9.9	-41.7	17.8
	Non significant	1,474	76.6%	2.6	-17.6	18.2
earng42	Significant	3,629	16.6%	12.7	-71.7	107.7
	Non significant	18,214	83.4%	3.9	-97.9	409.4
chben27	Significant	172	15.6%	8.0	-27.6	24.1
	Non significant	928	84.4%	4.4	-34.1	87.9
fmsz46	Significant	108	13.7%	6.2	-25.2	17.5
	Non significant	680	86.3%	2.6	-19.0	15.6
Licofb	Significant	581	13.7%	11.1	-59.1	35.9
	Non significant	3,671	86.3%	7.2	-100.0	91.3

The variable representing earnings (*earng42*) is highlighted as an example. Of all the estimates considered (based on all measures and domains), 3,629 (16.6%) were significantly different at the 5% level. The average of the absolute relative differences was 12.7%. Considering this and the magnitudes of the minimum (-71.7%) and maximum (107.7%) relative differences, it becomes apparent that the 25-cap would have an important impact on some estimates.

To add strength to the previous statement, we consider the impact on two important measures in the SLID. Table 8 presents the distribution of absolute relative differences (ARD) for the low income gap (*licoag27x*) and the indicator (*licofa*) used to identify family after-tax incomes below the low income cut-off (LICO). The low income gap is the difference between family income and the LICO. The LICO represents an income threshold where a family is likely to spend 20% more of its income on food, shelter and clothing than the average family, leaving less income available for other expenses such as health, education, transportation and recreation. Here we see, for example, that for the incidence of low after-tax income, of the 470 estimates where a significant difference was observed, 75% had an ARD greater than 5.9%; 50% had an ARD greater than 9.5%; 25% had an ARD greater than 16.0%. In other words, a quarter of the 470 estimates of incidence of low after-tax income would increase or decrease by 16% under the 25-cap. Thus, with a cap of 25 call attempts, very different conclusions would be made about the incidence of low after-tax income. Similar results are observed for the low income gap.

[5] *one* – indicator variable equal to 1 for all sample members, used to estimate size measures for domains; *earng42* – earnings, including salaries, wages; *chben27* – child benefits for economic family; *fmsz46* – number of census family members; *licofb* – indicator variable used to identify family before-tax incomes below the low income cut-off.

Table 8: Distribution of Absolute Relative Differences for Two Important SLID Variables Reference Year 2006

Variable	Test result	Mean	Med.	P25	P75	P90	P95	Min	Max
licoag27x	Significant	11.8	11.0	6.3	16.0	20.6	25.0	2.5	31.1
	Non significant	5.3	4.0	1.7	6.9	11.7	15.6	0.0	36.8
Licofa	Significant	11.6	9.5	5.9	16.0	21.6	24.9	2.2	55.3
	Non significant	8.3	4.8	2.1	9.8	17.4	26.7	0.0	100.0

3.1.3 Reliability of the Estimates

For the SLID, guidelines are provided based on the estimated coefficient of variation (CV) to qualify the reliability on an estimate. For example, estimates with CVs less than 16% are considered to be of acceptable quality (qualifiers A, B, C, D) and can be used for general publication; those with CVs greater than, or equal to, 16% are qualified as 'use with caution' (qualifier E). Estimates based on small sample sizes[6] are deemed unsuitable for publication and are suppressed (qualifier F).

Table 9: Changes in the Reliability of SLID Estimates, Reference Year 2006

Change in reliability	Number of estimates	Percentage
From (A, B, C or D) to E	4,482	3.5
From (A, B, C or D) to F	528	0.4
From E to F	2,936	2.3
Improvement	1,475	1.2
Stable	117,772	92.6
Total	127,193	100.0

From Table 9, almost 93 % of the estimates stay within the same category under the cap of 25 call attempts. Where substantial changes occurred, 3.5% of the estimates moved from 'acceptable' to 'use with caution', and 2.3% became too unreliable for publication. A small portion (1.2%) of estimates saw their reliability improve under the 25-cap, due to a more homogeneous sample of respondents and/or increases in the estimates themselves.

3.2 YITS

3.2.1 Response Rates

Tables 10 and 11 present the cross-sectional (i.e. cycle-specific) and longitudinal response rates, based on the final response codes used in production, and the response codes assuming a cap of 25 call attempts. Much like for the SLID, limiting the number of call attempts to 25 would have lead to a considerable decline in the response rates for the YITS Cohort A sample, with the gap increasing with each successive cycle. Given the nature of the survey, particular attention is given to the longitudinal response rates, which would have dropped between 2.5% and 7.3% under the 25-cap.

Table 10: YITS Cohort A Response Rates, Based on Final Response Codes, Cycles 1 to 4 (%)

Response Rate	Cycle 1 (2000)	Cycle 2 (2002)	Cycle 3 (2004)	Cycle 4 (2006)
Cross-sectional	86.6	90.5	84.4	83.1
Longitudinal	86.6	78.4	66.2	55.0

[6] In general, estimates based on a sample size of less than 25 (unweighted) respondents are suppressed, although special rules exist for percentages, ratios and quantiles.

Table 11: YITS Cohort A Response Rates, Based on 25-cap Response Codes, Cycles 1 to 4 (%)

Response Rate	Cycle 1* (2000)	Cycle 2 (2002)	Cycle 3 (2004)	Cycle 4 (2006)
Cross-sectional	86.6	87.6	80.9	77.7
Longitudinal	86.6	75.9	61.4	47.7

* Cycle 1 data were collected using a self-administered paper questionnaire and thus not subject to call attempts.

3.2.2 Accuracy of the Estimates

Tables 12 and 13 present summaries of tests for differences between the estimates obtained using the production weights and those obtained using the 25-cap weights. The categorical variables were used to define domains of interest and the proportion of the population (of the major domain) belonging to a particular domain was estimated. For the scale measures, mean scores were estimated. The tables present the percentage (number of differences/total number of tests) of estimates for which the results were significantly different at the 5% level.

Table 12: Summary of Significant Test Results for Estimates Based on Categorical Variables
YITS Cohort A, Cycles 2 to 4

Major domain	Cycle 2 (2002)	Cycle 3 (2004)	Cycle 4 (2006)	Total
Canada	18.6% (13/70)	14.3% (10/70)	4.2% (3/71)	12.3% (26/211)
Province	15.3% (104/680)	9.4% (66/700)	6.9% (49/710)	10.5% (219/2,090)
Urban/rural	18.0% (25/139)	12.1% (17/140)	7.0% (10/142)	12.4% (52/421)
Sex	16.4% (23/140)	10.7% (15/140)	7.0% (10/142)	11.4% (48/422)
Total	16.0% (165/1,029)	10.3% (108/1,050)	6.8% (72/1,065)	11.0% (345/3,144)

For the categorical variables, between 10.5% and 12.4% of the differences were significant across the major domains. Looking at differences by cycle, the differences seem to decrease with time, going from 16.0% in Cycle 2 to 6.8% in Cycle 4. This trend is observed for all major domains, and is perhaps attributable to a loss of precision in the estimates over time.

Table 13: Summary of Significant Test Results for Estimates Based on Scale Measures
YITS Cohort A, Cycles 2 to 4

Major domain	Cycle 2 (2002)	Cycle 3 (2004)	Cycle 4 (2006)	Total
Canada	16.7% (2/12)	25.0% (3/12)	25.0% (3/12)	22.2% (8/36)
Province	16.7% (20/120)	15.8% (19/120)	11.7% (14/120)	14.7% (53/360)
Urban/rural	4.2% (1/24)	25.0% (6/24)	20.8% (5/24)	16.7% (12/72)
Sex	16.7% (4/24)	29.2% (7/24)	25.0% (6/24)	23.6% (17/72)
Total	15.0% (27/180)	19.4% (35/180)	15.6% (28/180)	16.7% (90/540)

For the scale measures, the proportions of estimates that are significantly different under the 25-cap are greater than those observed for the categorical variables, with proportions ranging from

14.7% to 23.6% across the major domains. However, unlike the categorical variables, the proportions of significant differences do not decrease with successive cycles.

3.2.3 Reliability of the estimates

Table 14 presents the number and percent of estimates for which, under the limit of 25 call attempts, the reliability deteriorated, improved or remained stable. For the YITS, estimates based on an unweighted sample of size 25 or greater are: deemed 'acceptable' if the estimated CV is 16.5% or smaller; 'mediocre' if the estimated CV lies between 16.5% and 33.3%; and, 'unacceptable' if the estimated CV is greater than 33.3%. Estimates based on a sample of size less than 25 are also deemed unacceptable.

Table 14: Changes in the Reliability of YITS Cohort A Estimates, Cycles 2 to 4

	Categorical variables		**Scale measures**	
	Number of estimates	**Percentage**	**Number of estimates**	**Percentage**
From acceptable to mediocre	56	1.8	0	0.0
From acceptable to unacceptable	0	0.0	0	0.0
From mediocre to unacceptable	29	0.9	0	0.0
Improvement	9	0.3	0	0.0
Stable	3,050	97.0	540	100.0
Total	3,144	100.0	540	100.0

Ninety seven percent of the estimates for the categorical variables, and 100% for the scale measures considered, saw no change in their reliability. Where changes occurred, nearly 2% saw their reliability fall from acceptable to mediocre, and less than 1% became unacceptable.

4. Conclusions and Future Work

Limiting the number of call attempts to 25 contributed to a reduction of 5% in the response rates for the two surveys (the impact on longitudinal response was even greater), However, careful examination of the survey results is necessary to asses the effects of a reduced response rate brought by the 25-cap on the accuracy and reliability of the survey estimates. Evidence has been presented from both the SLID and the YITS that shows that such a policy would introduce significant changes for an important number of estimates. And while this is not necessarily evidence of bias, it does raise strong concerns about the impact of the 25-cap might have on the accuracy of the survey data and inconsistencies it would create with the estimates from previous cycles.

Also considered was the impact the cap would have on the reliability of the resulting estimates. The results show that limiting the number of call attempts might have a greater impact on the CVs for estimates from the SLID than those from the YITS, as, under the simulated cap, a larger percentage of the SLID estimates saw their reliability deteriorate to the point where they were deemed no longer fit for publication.

Based on the evidence presented here, a 40-cap was recommended for both surveys, with the understanding that new initiatives to make the most efficient use of collection resources would be investigated. To this end, for the SLID 2010 (i.e. reference year 2009) data collection, a responsive design strategy was implemented (Tabuchi *et al.*, 2009; Laflamme and Karaganis, 2010). The goal of this strategy is to maintain adequate response rates and ensure the representativity of the sample, while introducing efficiencies in the management of the collection process. The results of this initiative are forthcoming.

Acknowledgements

The authors would like to thank Hélène Bérard and Mark Armstrong for their helpful comments.

References

Dufour, S. 2009. A Review of Response Rates for Statistics Canada Surveys. Internal presentation, Statistics Canada.

Hunter, L. and Carbonneau, J.-F. 2005. An Active Management Approach to Collection Management. Proceedings of Statistics Canada Symposium 2005.

Laflamme, F. and Karaganis, M. 2010. Development and Implementation of Responsive Design for CATI Surveys at Statistics Canada. To be presented at the European Quality Conference, Helsinki, Finland.

Laflamme, F., Maydan, M. and Miller, A. 2008. Using Paradata to Actively Manage Data Collection. 2008 American Statistical Association, Proceedings of the Section on Survey Research Methods.

LaRoche, S. 2007. Longitudinal and Cross-sectional Weighting of the Survey of Labour and Income Dynamics. *Income Research Paper Series*, Statistics Canada Catalogue no. 75F0002MIE — No. 007.

Lévesque, A. 2008. Survey of Labour and Income Dynamics Collection Activities Management: 2008 Evaluations. Internal report, Statistics Canada.

Lévesque, A., Lim, A., Wang, Q. 2010. Limite du nombre d'appels pour les enquêtes longitudinales. Methodology Branch Working Paper DMEM-2010-002F, Statistics Canada.

Lévesque, A. and Poulin, J. 2007. Collection Activities Management for the Survey of Labour and Income Dynamics: 2005 to 2007 Evaluations. Internal report, Statistics Canada.

Tabuchi, T., Laflamme, F., Phillips, O., Karaganis, M. and Villeneuve, A. 2009. Responsive Design for the Survey of Labour and Income Dynamics. Proceedings of Statistics Canada Symposium 2009.