



Munich Personal RePEc Archive

Time to Face the Fact that the National Household Survey Is Just the Compulsory Long-Form Census Made Voluntary

Patrick, Grady

Global Economics

6 February 2015

Online at <https://mpra.ub.uni-muenchen.de/63188/>

MPRA Paper No. 63188, posted 24 Mar 2015 13:29 UTC

Global Economics



**Time to Face the Fact that the National Household Survey
Is Just the Compulsory Long-Form Census Made Voluntary**

By
Patrick Grady

Global Economics Working Paper 2015-2

February 6, 2015

Abstract:

A highly-charged debate that has raged now for four years on the replacement of the long-form census with the voluntary National Household Survey (NHS) has become more political and symbolic than substantive at this point. This article makes the case that the NHS has produced useable high quality data, which is comparable in coverage to that produced by the long-form, at the national, provincial, territorial and census metropolitan level. It also contends that going back to a mandatory questionnaire may not produce much better results given the increased awareness resulting from the controversy that the stiff penalties under the Statistics Act are rarely, if ever, imposed.

JEL Classification Codes: C83 Survey Methods • Sampling Methods; J1 Demographic Economics

Keywords: National Household Survey, long-form census, Statistics Canada

Introduction

The 2011 National Household Survey (Statistics Canada, 2014a) is a voluntary survey that replaced the compulsory long-form census questionnaire. Indeed it might be more accurate to say that it is nothing less than the long-form census transformed into a voluntary survey. The changes were motivated by the Government's desire to make the questions asked less intrusive by giving people the option of not answering without facing potential legal sanctions. But the change raised concerns among analysts that the elimination of penalties would reduce response rates and render the survey results unreliable.

While changes in statistical programs are usually boring affairs only noted by statisticians and economists, this replacement has turned out to be anything but. In fact, it is the only government-mandated change in a statistical program ever to precipitate the resignation of the Chief Statistician. Moreover, it ignited a ferocious controversy that is still raging more than four years later.

A bill (C-626) that was recently introduced by Liberal MP Ted Hsu to reinstate the mandatory long-form questionnaire was hotly debated in the House before being voted down by Government members. Op eds denouncing the NHS have appeared in the *Globe and Mail* (Jacobsen, 2014; Maioni, 2015), accompanied by editorials urging the reinstatement of the long-form 2014, and 2015). And most recently, two professors with the Martin Prosperity Institute at the Rotman School of Management at the University of Toronto wrote a piece making the obviously exaggerated claim that "in scrapping the long-form census, the Harper administration has threatened the country's long-term economic prosperity." (Florida and Martin, 2014) There is even a Facebook page "Keep the Canada Census Long Form".

Looking at the Survey

Having, like most economists and commentators, been against to the replacement of the long-form census, I had fairly low expectations when the new National Household Survey was released. I thought that it would probably turn out to be fairly useless as alleged by many because it would suffer from low response rates often associated with voluntary surveys and because it would be difficult to make comparisons of data across time given the change in methodology. I certainly recognized that the long-form had the advantage of consistency of approach across time and that this facilitated intertemporal comparisons of data. I also expected that the amount of detailed information collected in the NHS would be less given that it is voluntary.

But when I actually looked into the 2011 NHS as packaged for public distribution in a Public Use Microdata File (PUMF), I was actually quite pleasantly surprised. The first thing I noted when I ordered the PUMF was that it was free. This was an especially welcome development as the comparable 2006 Census PUMF cost me \$1,150 plus 5 per cent GST to acquire. It cannot be emphasized too much that Statistics Canada's new

policy of making all its data available free of charge to all Canadians is a great improvement over its previous policy of restricting access through high user fees. In my view, this should be quite helpful in encouraging researchers to actually use the data. The previous policy of spending vast sums of money to collect data that were then sold for prices too expensive for many potential users did not make much sense. Once the data are collected, they are in effect a public good available at zero marginal cost, which could and should be provided to all for free.

Statistics Canada (2014a, p.5) notes in the user guide that the content of the 2011 PUMF is largely the same as that of the 2006 PUMF. It cautions, however, that there are various changes, resulting from content changes in the 2011 NHS, as well as the creation of new variables from existing questions or the use of updated classifications on existing questions. These include: the addition of 20 new variables; the removal of 13 old variables; and a change of universe to Mobility, Generation status and Place of birth of parents variables. Nevertheless, the basic structure of the long-form questionnaire had been retained, including the inclusion of the key question that allowed respondents to link their responses to their tax forms as had been done in the census long-form. This resulted in the preservation of most of the data series collected in the census and in a database that even uses the same names for most of the 124 variables included. The sample size included in the PUMF is 887,012 which is also comparable, representing 2.7 per cent of the Canadian population.

A concern of most analysts is that the integrity of the NHS data would be compromised by a low response rate. It had been predicted to fall to as low as 50 per cent by at least one statistician at the time the NHS was announced. This concern, happily, was alleviated by the sense of civic duty of individual Canadians combined with the rigorous reliability checking carried out by Statistics Canada and reported in the dictionary and PUMF user guide. A very sophisticated statistical methodology was utilized for sampling and weighting to ensure that the sample was representative of the population. This is the approach that Statistics Canada has pioneered and used with great success with all its voluntary surveys to produce reliable information. Statistics Canada's statisticians are definitely professionals capable of getting the most information out of voluntary surveys and they have put their skills to the test on the NHS, which is the largest voluntary survey ever conducted.

With a sampling rate of about 3 in 10 and an overall response rate of 68.6 per cent, Statistics Canada (2014b, p.12) estimated that about 21 per cent of the Canadian population participated in the NHS (only a fraction of which are included in the PUMF). Not coincidentally, this is comparable to the population participating in the long-form census questionnaire in 2006, which was provided to one in five households. Thus it is impossible to make a convincing case that Statistics Canada did not obtain information from a large enough sample of people (the representativeness of the sample is, of course, another issue about which there can still be legitimate concerns).

While the response rates to the individual questions were lower than in the 2006 Census as was expected with a voluntary survey (for instance, 59.3 per cent versus 67.4 per cent

for income; and 57.6 per cent versus 76.6 per cent for income tax paid), the data was judged to be of publishable quality by Statistics Canada after performing its usual rigorous editing and consistency checks and comparisons with the Survey of Labour and Income Dynamics (SLID) and the Annual Estimates for Census Families and Individuals or T1 Family File (T1FF), an income tax data file prepared for the Canada Revenue Agency (CRA).

As an aside, it is worth noting that even for the mandatory long-form questionnaire the response rates are significantly less than 100 per cent (93.5 per cent in 2006), which means that, even if the respondents filled out the questionnaire as required by law, they often did not answer all the questions. And while the response rates of the NHS are lower than the mandatory long-form census questionnaire, there is a question of whether the quality of the responses provided might actually be better. Statistics Canada can by law make people fill out the long-form, but they cannot make them answer accurately if the respondents consider the questions to be too intrusive or even too time-consuming to answer. Thus, it is quite possible that the responses of people voluntarily agreeing to answer questions might be more accurate than the responses of people who are compelled. However, this is a question that properly should be left for the statisticians to resolve.

According to Statistics Canada (2014c, pp.15-16), the estimates of the number of income recipients from the NHS estimates of 2010 income are between the estimates from the 2010 SLID (3.2 per cent lower) and the 2010 T1FF (2.7 per cent higher). But the NHS estimate of median total income was 4.0 per cent greater than that from the 2010 SLID and 2.3 more than in the 2010 T1FF. Estimates of income tax from the NHS are also between the SLID and the T1FF. The relative closeness of these comparisons is one indication of the quality of the data and should reassure potential users.

Overall, Statistics Canada characterizes the quality of the NHS data at the national, provincial, territorial and census metropolitan area levels as high. But it also cautions that in some smaller areas and for some smaller populations, the response rate may be insufficient to provide a valid statistical picture. In these cases, users have to use the data at higher geographic levels or for larger subgroups of the population. This is why city planners have been so prominent in lamenting the passing of the long-form (Grant and Church, 2015). However, most users do not need to make use of the data at this level of granularity and most public policy analysis is done at the national, provincial, territorial and census metropolitan area level.

Another observation can be made on the replacement of the long-form survey. Most economists and statisticians are preoccupied with data quality and not so worried about response burden because they benefit from the former and are not so burdened by the latter. It can't be denied that mandatory surveys backed by legal sanctions and penalties certainly have higher response rates than voluntary surveys. But politicians are concerned about using the force of law to require people (i.e. voters) to respond to surveys, such as the long-form, which at 40 some pages was by a wide margin the granddaddy of them all. That is probably why there are so few mandatory surveys of households in Canada (in

contrast to business or agricultural surveys, which are mandatory under the Statistics Act). Other than the Census of Population, which has 8 pages of questions, there is only the Labour Force Survey, which has 21 pages of questions, which is still mandatory. And these two surveys are much shorter than the National Household Survey, which is 40 pages long and about the same as the long-form.

Around the world concern about compliance and collection costs is resulting in increased efforts to rely on administrative data for statistical purposes instead of more costly mandatory surveys. These efforts are most advanced in Europe. There is also concern about the timeliness of census data which is only collected every five or ten years and is only available after a one or two year processing lag. This is why the United States has implemented an annual mandatory American Community Survey (ACS) to replace their census long-form. With a sample of currently around 3.5 million housing units it is significantly smaller than the 4.5 million in the NHS both in absolute terms and especially as a percentage of the much smaller Canadian population.

Statistics Canada is not really in a position to make the trade-off between economic costs of lower quality data and the political costs of using coercion to collect the data. All it can do is advise on the implications for data quality and cost of the decisions facing the government, which would ultimately be made taking into account political as well as technical considerations. This is probably why, prior to a decision being made about what to do about the long-form, a voluntary survey was presented to the Industry Minister (who is the Minister responsible for Statistics Canada) as an option, even if it was not the preferred option of Statistics Canada (as determined by the Chief Statistician and officials) (Harris, 2010).

In the case of the replacement of the long-form census with the NHS, it seems to me that things have turned out much better than feared at the time the decision was announced. The cost in terms of lower data quality doesn't seem to be that high. This is due to Statistics Canada ability to in work around the political constraints imposed on it and to continue to produce high quality data (except in the limited situations of small areas). All of this, of course, has come at a cost (estimated to be around \$30 million [Yalnizyan, 2013] presumably to process the larger sample and to analyze the data), but this is evidently a cost that the government is willing to bear in return for the benefits of making the survey voluntary.

Conclusions

In conclusion, the NHS is far from the unmitigated statistical catastrophe still portrayed by many. Plans are already well advanced for next NHS, which is scheduled for 2016. There are grounds for optimism that the experience gained with the 2011 survey and the subsequent data analysis will enable Statistics Canada to improve the NHS and make the data even better. At this point, it doesn't make much sense to reverse course and reinstate the long-form census as advocated by the many vocal critics. This would just have the unfortunate consequence of making the data less comparable going forward and not necessarily produce anything more useful. It would really only mean transforming the

existing voluntary NHS back into a mandatory survey, not its replacement by something totally different and more appropriate. It is noteworthy that for all the talk of the need to reinstate the long-form census, there has not been much criticism of the specific questions or proposals about how they should be changed, probably because the questions in the NHS are so similar to those formerly asked in the long-form census. And it would be difficult, if not impossible, to get as high of a response rate for the so-called reinstated long-form as for the old version now that the long-running controversy has publicized the fact that the penalties under the Statistics Act are rarely, if ever, enforced (not to mention that the defeated Liberal bill and another Conservative bill would substitute a fine for imprisonment as the penalty for not responding).

The fight for a reinstatement of the long-form census has gone on for too long and become more political and symbolic than substantive at this point. The time has come to move on to more important things on the statistical front. A more constructive approach would be to see what can be done to improve the census and NHS moving forward. For instance, some of the more innovative techniques pursued in other countries such as greater use of administrative data and a smaller more frequent household survey could be explored. Reinstating the long-form is no panacea.

References

- Florida, Richard and Roger Martin (2015) "For Competitiveness' Sake, Restore the Canadian Census," *Huntington Post Canada*, January 29, 2015.
- Globe and Mail* (2014) "Ending mandatory long-form census has hurt Canada," November 6.
- Globe and Mail* (2015) The census: Little knowledge is a dangerous thing, February 3.
- Grant, Tavia and Elizabeth Church (2015) "Cities to weigh loss of long-form census for community planning," *Globe and Mail*, February 3.
- Harris, Megan (2010) "Why did top statistician take so long to resign over census?," *Globe and Mail*, July 21.
- Hsu, Ted (2014) "Bill C-626 – An Act to Amend the Statistics Act." <http://www.tedhsu.ca/C626/overview.html>
- Jacobson, Paul (2014) "Policy making suffering in Canada without the long-form census," *Globe and Mail*, November 5.
- Maioni, Antonia (2015) "We haven't forgotten the long-form census," *Globe and Mail*, February 6.
- Statistics Canada (2009) *2006 Census Public Use Microdata File (PUMF) Individuals File Documentation and User guide*. Catalogue no. 95M0028XVB.
- Statistics Canada (2014a) *2011 National Household Survey Public Use Microdata File (PUMF) Individuals File*. Catalogue no. 95M00081.
- Statistics Canada (2014b) *NHS User Guide*, Catalogue no. 99-001-X2011001.
- Statistics Canada (2014c) *Income Reference Guide: National Household Survey, 2011*, Catalogue no. 99-014-X2011006.
- Yalnizyan, Armine (2013) "National Household Survey provides blurred look at housing," *Globe and Mail Economy Lab*, September 12.