

National Household Survey versus Taxfiler (T1 Family File): A local comparison of income data in Waterloo Region, Ontario

Jessica Deming, MSc

Epidemiologist, Region of Waterloo Public Health

Community Data Program Webinar

June 18, 2014

Background

- **Region of Waterloo previously relied on long-form census data for numerous activities:**
 - To plan and evaluate programs and services
 - To make decisions and provide recommendations
 - To share local population data with community partners and the general public
- **Now we need to understand as much as possible about 2011 NHS, to decide how to proceed without long-form census data**

Objectives

- 1. To compare and contrast the NHS & Taxfiler data sources & variables**
- 2. To compare estimates of income & low income at the local level**

Data source characteristics

	2011 NHS	2010 Taxfiler
Population universe	All persons who usually live in Canada	All persons who filed an income tax return in the reference year (as well as their families – imputed)
Definitely excluded	Persons living in institutions; Canadians living in other countries; persons living in work camps, hotels, and student residences	Persons who did not file an income tax return
Possibly excluded		Non-permanent residents
Bias	Non-response, over sampling	Partial non-response (missing data from non-filers)
Income time frame	2010 tax year	2010 tax year
Measures of income	Total, median, average, LICO (pending), LIM, MBM (pending)	Total, median, LIM
Source of income	Market, government payments	Market, government payments
Level of geography	PT, CMA,CA,CT,CD,CSD,DA, Neighbourhood (Pending), Planning District (Pending) (custom geographies)	FSA, rural postal, postal city, CMA,CA,FED,CT,CD
Age groupings	5 year (<15,75+)	Single year, 5 year (0-24,65+)
Unit of measurement	Individual, household, census/economic family	Individual, census family

Apples to apples?

Overall assessment (objective 1):

- Reasonable agreement in population universe
- Both subject to some bias (more so for NHS)
- Several comparable measures of income (total, median, low income measure (LIM))
- Several options for comparisons at sub-provincial levels of geography (CMA, CA, CD, CT)
- Can compare individual-level data or family level

How does NHS income data compare to Taxfiler data in Waterloo Region?

NHS vs Taxfiler in Waterloo Region

Table 1. Population counts for Waterloo Census Division, by population sub-group, 2011 NHS and 2010 Taxfiler

Waterloo Census Division, 2010	NHS	Taxfiler	Percent Difference
Estimated total	499,615	497,940	+0.3%
Males	247,305	243,493	+1.6%
Females	252,310	254,447	-0.8%
Couple families	116,930	123,440	-5.3%
Lone-parent families	19,170	19,980	-4.1%
Persons not in families	64,155	71,700	-10.5%

NHS vs Taxfiler in Waterloo Region

Table 2. Median total income values for Waterloo Census Division, by population sub-group, 2011 NHS and 2010 Taxfiler

Waterloo Census Division, 2010	NHS	Taxfiler	Percent Difference
Number of persons with income	387,445	378,750	+2.3%
Overall	\$32,780	\$33,050	-0.8%
Males	\$40,266	\$41,590	-3.2%
Females	\$26,713	\$26,570	+0.5%
Couple families	\$83,928	\$87,150	-3.7%
Lone-parent families	\$47,237	\$40,120	+17.7%
Persons not in families	\$30,364	\$28,450	+6.7%

NHS vs Taxfiler in Waterloo Region

Table 3. Median after-tax income values for Waterloo Census Division, by population sub-group, 2011 NHS and 2010 Taxfiler

Waterloo Census Division, 2010	NHS	Taxfiler	Percent Difference
Number of persons with income	387,390	378,820	+2.3%
Overall	\$29,878	\$30,110	-0.8%
Males	\$35,940	\$36,840	-2.4%
Females	\$24,873	\$24,800	+0.3%
Couple families	\$73,163	\$75,220	-2.7%
Lone-parent families	\$44,280	\$38,200	+15.9%
Persons not in families	\$27,647	\$26,370	+4.8%

NHS vs Taxfiler in Waterloo Region

Table 4. Low income measure (LIM) values for Waterloo Census Division, 2011 NHS and 2010 Taxfiler

Waterloo Census Division, 2010	NHS	Taxfiler	Percent Difference
Persons with low income (LIM)	59,225	58,160	+1.8%
% of population	11.9%	11.7%	+1.5%

How do these differences in NHS to Taxfiler income data compare to the differences previously observed between Census and Taxfiler income data?

Census vs Taxfiler in Waterloo Region

Table 5. Population counts for Waterloo Census Division, by population sub-group, 2006 Census and 2005 Taxfiler

Waterloo Census Division, 2005	Census	Taxfiler	Percent Difference	NHS/Tax Percent Difference (2010)
Estimated total	478,121	465,260	+2.8%	+0.3%
Males	235,545	223,325	+5.5%	+1.6%
Females	242,570	241,935	+0.3%	-0.8%
Couple families	114,385	114,840	-0.4%	-5.3%
Lone-parent families	19,340	19,760	-2.1%	-4.1%
Persons not in families	64,190	63,770	+0.7%	-10.5%

Census vs Taxfiler in Waterloo Region

Table 6. Median total income values for Waterloo Census Division, by population sub-group, 2006 Census and 2005 Taxfiler

Waterloo Census Division, 2005	Census	Taxfiler	Percent Difference	NHS/Tax Percent Difference (2010)
Number of persons with income	365,105	350,200	+4.3%	+2.3%
Overall	\$29,449	\$29,900	-1.5%	-0.8%
Males	\$38,436	\$39,100	-1.7%	-3.2%
Females	\$22,517	\$22,900	-1.7%	+0.5%
Couple families	\$81,906	\$78,200	+4.7%	-3.7%
Lone-parent families	\$43,614	\$34,100	+27.9%	+17.7%
Persons not in families	\$28,785	\$26,000	+10.7%	+6.7%

Census vs Taxfiler in Waterloo Region

Table 7. Median after-tax income values for Waterloo Census Division, by population sub-group, 2006 Census and 2005 Taxfiler

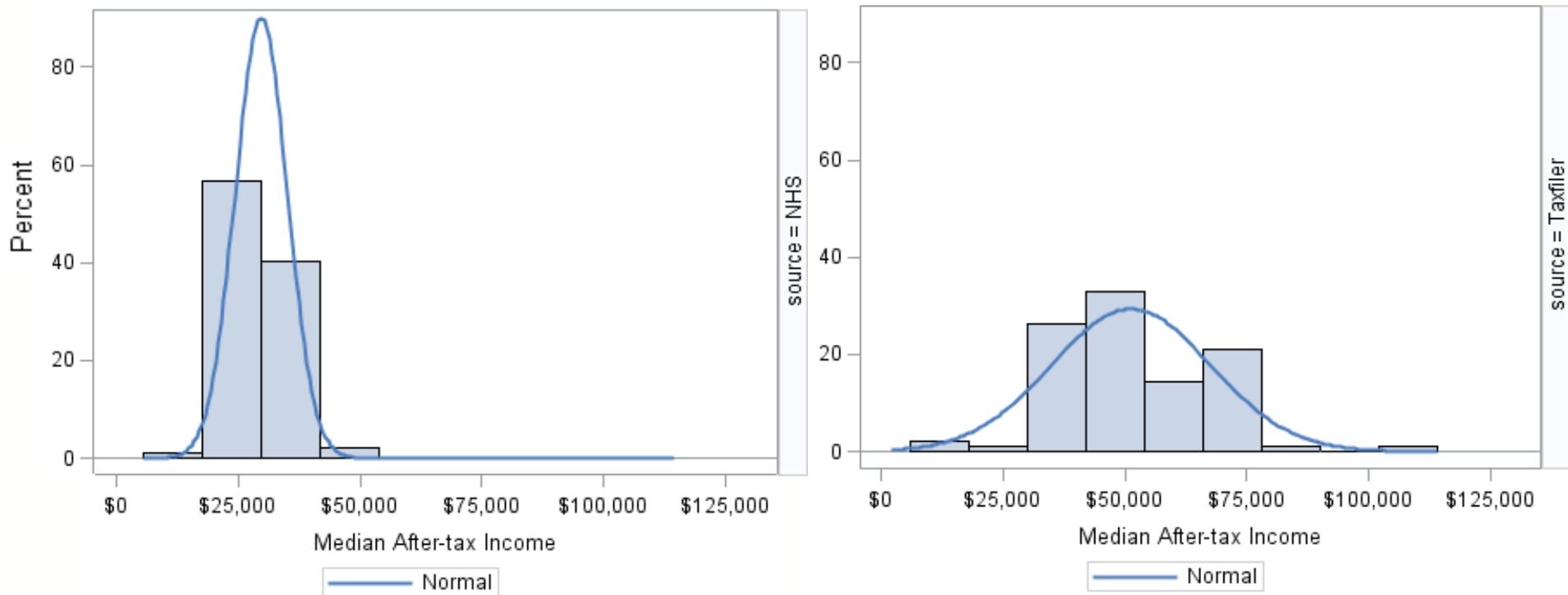
Waterloo Census Division, 2005	Census	Taxfiler	Percent Difference	NHS/Tax Percent Difference (2010)
Number of persons with income	365,005	350,350	+4.2%	+2.3%
Overall	\$26,187	\$26,500	-1.2%	-0.8%
Males	\$32,958	\$33,400	-1.3%	-2.4%
Females	\$20,748	\$21,100	-1.7%	+0.3%
Couple families	\$69,076	\$65,800	+5.0%	-2.7%
Lone-parent families	\$39,454	\$31,600	+24.9%	+15.9%
Persons not in families	\$25,690	\$23,500	+9.3%	+4.8%

Income estimates in 2010 are generally similar at the CD level, but:

Do the estimates hold at a more granular level?

NHS vs Taxfiler in Waterloo Region

Figure 1. Distribution of median after-tax income by census tract, Kitchener-Cambridge-Waterloo Census Metropolitan Area, 2010



LIM-AT Proportion Quartiles First Quartile Second Quartile Third Quartile Fourth Quartile

2010 Taxfiler

2011 National Household Survey

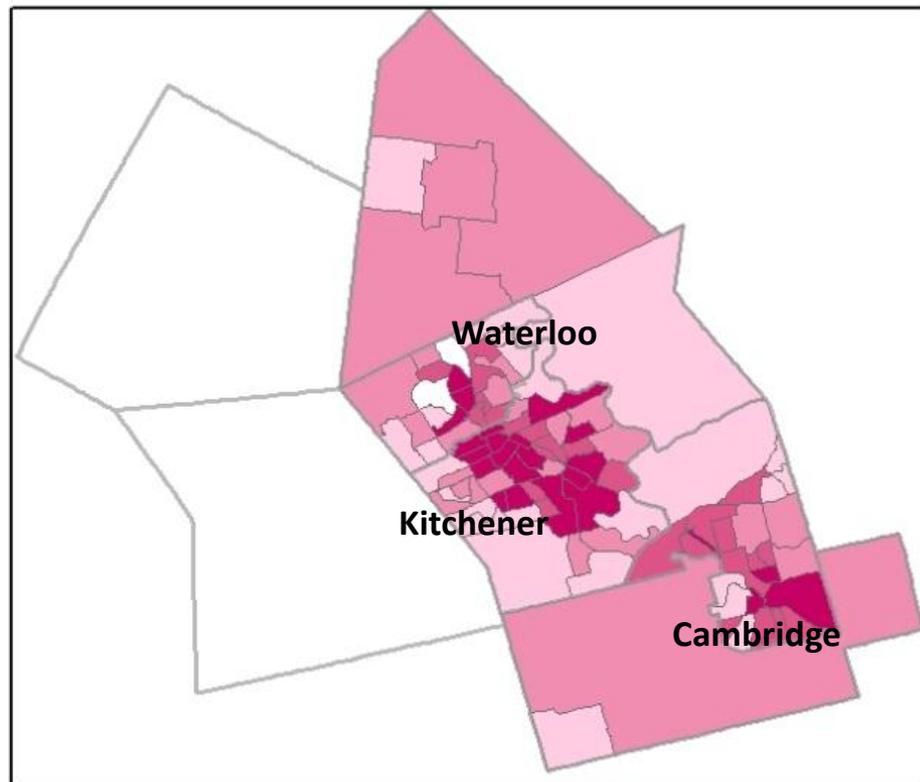
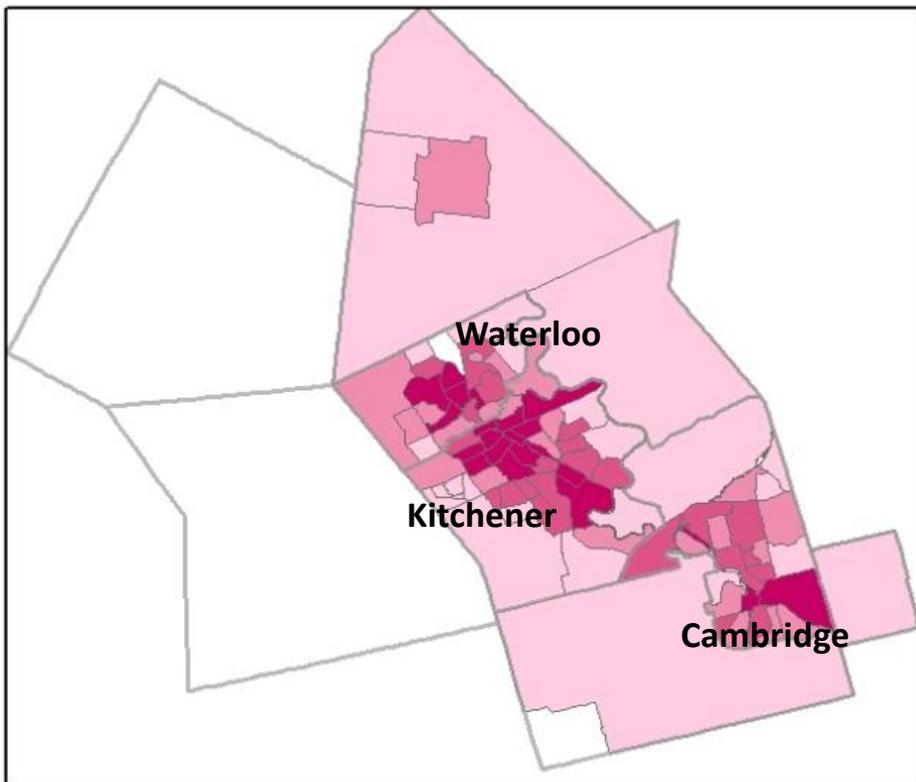


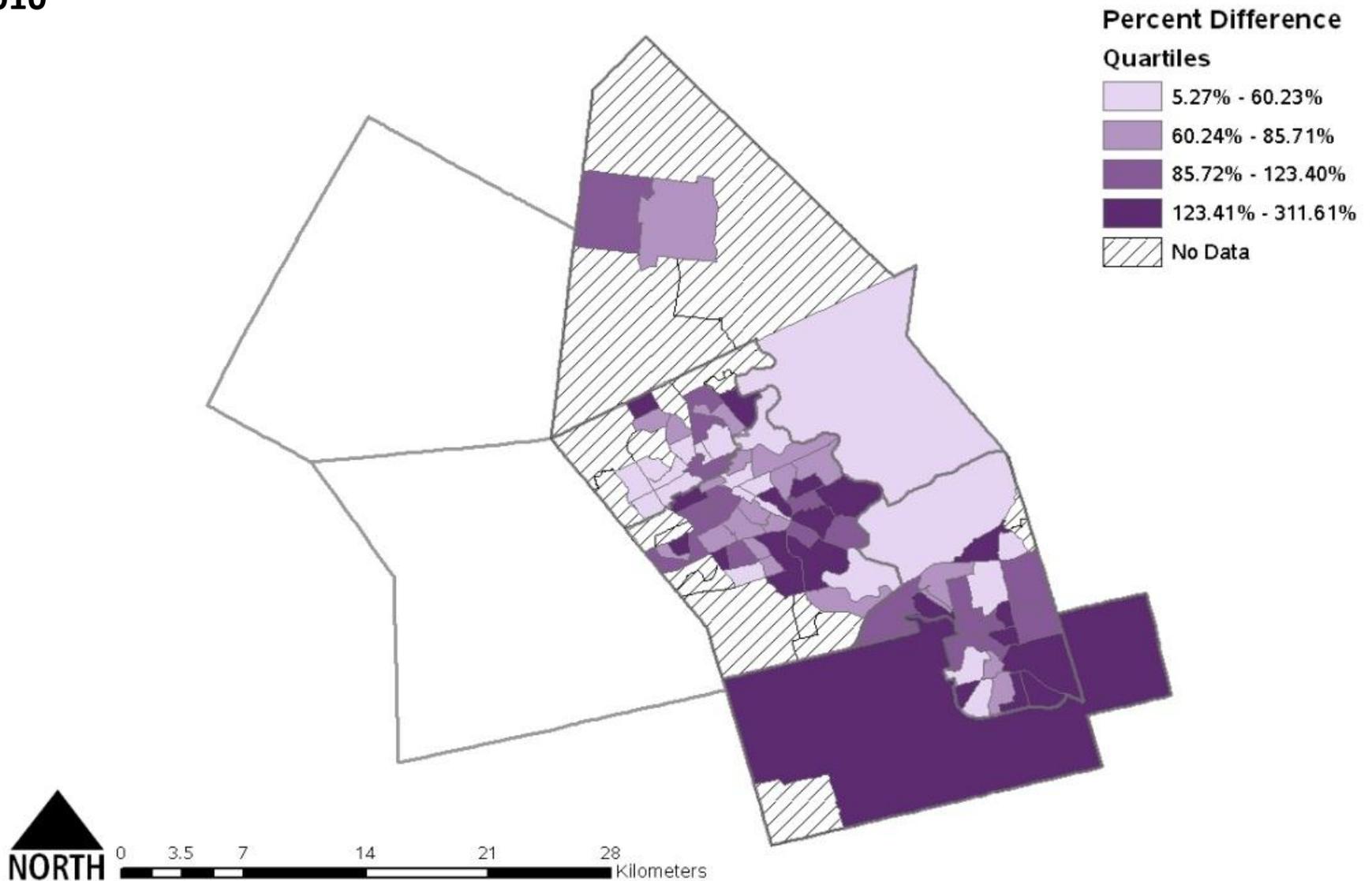
Figure 1. Quartiles of proportion of population with low income (LIM-AT), by Census Tracts, Kitchener-Cambridge-Waterloo CMA, NHS and Taxfiler, 2010



Created by Epidemiology and Health Analytics. Created February 28, 2014

Source: Statistics Canada, 2013. 2011 National Household Survey. Released 11 September 2013.
Statistics Canada, Income Statistics Income Division, 2011, Annual Estimates for Census Families and Individuals.

Figure 2. Quartiles of percent difference in the proportion of population with low income (LIM-AT) between NHS and Taxfiler, by Census Tracts, Kitchener-Cambridge-Waterloo CMA, 2010



Source: Statistics Canada, 2013, 2011 National Household Survey. Released 11 September 2013.
Statistics Canada, Income Statistics Division, 2011. Annual Estimates for Census Families and Individuals.

Summary of findings

Overall assessment (objective 2):

- Good coverage of population for both data sources
- NHS data similar to Taxfiler, except lone-parent families and persons not in families
- Census data also appears similar to Taxfiler, except lone-parent families and persons not in families (larger differences than NHS)
- Income data at the Census Tract level is highly variable between the two sources

Which source do we have greater confidence in?

Which source is the "better choice" to report on income at small levels of geography?

Food for thought

1. NHS is a voluntary survey (albeit a very large one) while Taxfiler is an administrative database
2. NHS: "Risk of error increases for lower levels of geography and for smaller populations"
3. NHS is adjusted for undercoverage; Census and Taxfiler are not
4. Much of the Census and NHS income data is based on linkages to tax records (82.4% vs. 73.2%)
5. Income-related item response rate in NHS is lower than the Census (57.9-60.6% vs. 67.4-80.6%)
6. More incentive for Canadians to respond (and respond honestly) when filing taxes vs. voluntary survey or mandatory census

More food for thought

- **Statistics Canada comparisons of NHS to Taxfiler at the national level revealed:**
 - NHS showed more income recipients and earners, median total income, median employment income and median wages and salaries were all higher than Taxfiler data
 - NHS estimates more people earning \$100,000+ compared to Taxfiler, but estimates fewer people with total income of \$1,000,000+ (20% lower)
 - NHS estimates more family units with income under \$10,000 but fewer with income \$10,000-\$29,999 (reverse was true for 2005)
 - NHS estimates of Low-Income Cut-offs (LICOs) were less comparable to those calculated from other sources

Practical points to consider

- Taxfiler is available on an annual basis, whereas NHS will only be repeated every 5 years
- Taxfiler data is available historically (in CDP, back to 2006) while NHS currently has only been collected once
- Taxfiler has fewer measures of income (median, total, Low Income Measure) while NHS has more (e.g., Market Basket Measure, average income)
- NHS is the richer source of demographic data & other variables to allow for cross-tabulations

Notes on data suppression

- **Taxfiler data suppression:**
 - Areas compromised of less than 100 taxfilers
 - Cells representing less than 15 taxfilers
 - Cells that are dominated by a single taxfiler
 - Cells for median income based on a rounded count of less than 20 taxfilers
 - Where one category (e.g., income, gender or age, geographic area) is suppressed another is suppressed to avoid disclosure by subtraction

Notes on data suppression

- **NHS data suppression:**
 - Estimates are suppressed (to zero) if the number of records is less than 4
 - Statistics (e.g. median income) are suppressed if:
 - the number of actual records used in the calculation is less than 4,
 - the calculation cell contains an outlier,
 - the range of the data is below a certain threshold
 - the sum of the weights is less than 10
 - Income data are suppressed in areas where the population is less than 250 or where the number of private households is less than 40

Conclusions

Region of Waterloo's current approach to using NHS:

- Use the NHS data with caution, and advise others accordingly, including staff responses to requests from the media
- Not determine trends or change over time by comparing data from the 2011 NHS with data from previous Census periods
- Not publish data at sub-municipal levels of geography such as Census Tracts and Dissemination Areas, when made available

Our investigation validates that these are sound principles to approaching use of NHS income data

Thank you!

Questions?

**Email Jessica at
jdeming@regionofwaterloo.ca**