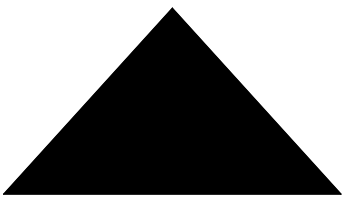
Central East
Health Information Partnership

Data Quality Report:

Effect of Residence Code Errors on Fertility Rates

July 2000



CEHIP

Data Quality Report:
Effect of Residence Code Errors on Fertility Rates

G.L. Woodward and S. Ardal

July 2000

17250 Yonge Street, Box 234, Newmarket, Ontario L3Y 4X1 Tel: (905) 764-6346 Fax: (905) 895-0848 URL: http://www.cehip.org

Executive Summary

The use of administrative data for health planning purposes requires that the data be reliable and valid. In other words, it must consistently capture what it purports to capture. Currently, two province-wide administrative systems, vital statistics and hospitalization, capture data for birth events occurring in Ontario. Birth events from these databases are used extensively to plan public health programs and estimate future health service need and population growth. Key to the use of these data is accurate residence coding at the county and municipal level.

Our analysis of fertility rates within Ontario found that coding for county or Census Division of residence appears accurate in both the vital statistics and hospitalization data. For a large number of births in the vital statistics database however, the municipal code (CSD) assigned by Statistics Canada is not in agreement with the residential postal code nor is it in agreement with hospitalization records. Radical shifts in the annual fertility rate (increase or decrease of 50% or more) can be seen at the municipality or Census Subdivision level. Although this may be expected in CSDs with few births annually, these shifts occur in some heavily populated CSDs such as Toronto and Vaughan.

Inaccurate residence coding can have serious implications for health planning and public health programs. Municipal based neonatal clinics could be inappropriately staffed and estimates of future health service need and population growth could be grossly over- or under-estimated. Methods used to assign residence information to birth events, particularly those captured by vital statistics, need to be examined to ensure that residence information is accurate.

Introduction

As the movement towards evidence based planning grows, the use of administrative data in public health planning and research has increased dramatically. Two of the data sets most commonly used by Ontario epidemiologists and health planners are vital statistics and hospital discharge abstracts. These data are used to monitor and predict health status and health service utilization. These analyses often require that the data be examined at relatively fine levels of geography, such as the municipality of residence. Thus, the geocoding of data at this level should be accurate and consistent within and across health related data sets.

One of the mandates of the Central East Health Information Partnership (CEHIP) is to investigate and improve the quality of the data used by its partner agencies, the Public Health Units, District Health Councils, and academic research centers in our service area. A major concern of the Partners is the quality of the vital statistics data they receive. Many Partners have questioned the accuracy of the residential geocoding assigned to birth events occurring in Ontario.

The purpose of this investigation is to address the reliability and validity of residential geocoding in Ontario birth data.

Methods

Data Sources

Two sources of data were used in our analyses:

(1) Vital Statistics Birth Data

Electronic records for these data are generated by the Ontario Office of the Registrar General. The Registrar's Office captures the mother's residence information (street address, city, province, postal code) and the place of occurrence (city, province). These data are used to create Statistics Canada's Canadian Vital Statistics File, which in turn is used to create the Ontario Live Birth Database, a product of the Ontario Ministry of Health and Long Term Care (MOHLTC), Public Health Branch. Included in the Ontario Live Birth Database are the Standard Geographical Codes for the mother's Census Division (CD) and Census Sub-division (CSD) of residence. These codes are assigned by Statistics Canada based on the mother's municipality and county of residence, as identified by the Registrar General's Office.

(2) Inpatient Hospitalization Data

These records are generated by each hospital in Ontario when a patient is discharged from the hospital. The records are submitted to the Canadian Institute for Health Information where they become part of the Discharge Abstract Database (DAD). We obtained DAD records for events occurring in Ontario hospitals from the MOHLTC Provincial Health Planning Database. Each DAD record contains the patient's residential postal code and the MOHLTC Residence Code indicating the county and municipality of residence.

Newborn records were selected from the DAD based on the patient's entry code (=newborn). All cases identified as being transferred from another hospital were removed from the resulting dataset to avoid counting each birth more than once.

Data Analysis

Data analyses focussed on the fertility rates at two geographic levels within Ontario, CD and CSD. The fertility rate is a measure of the number of births associated with an area relative to the number of women who are of childbearing age (15-49 years). Fertility rate was chosen as our measure because it takes into account changes in the female population of child bearing age (number of live births / number women aged 15-49 years x 1000).

Data analysis included only births to Ontario residents occurring in Ontario and proceeded in three stages:

1) **CD Fertility Rates, 1986-96 - Annual fertility rates of Ontario's CDs from 1986-1996 were charted and examined. This analysis used the existing residence code within the vital statistics file, as assigned by Statistics Canada, to identify the mother's CD of residence.**
2) **CSD Fertility Rates, 1986-96 - Annual fertility rates for the CSDs within two selected CDs, Metropolitan Toronto and Regional Municipality of York. These CDs were chosen because they fall within our area of service, were of concern to a number of the partner agencies, and contain a relatively large number of births per CD. Thus, large annual fluctuations in the fertility rate would not be expected.**
3) **1996 Birth Count and Fertility Rate Comparison - For the Metropolitan Toronto and York Region CSDs, 1996 birth counts and fertility rates were computed using two data sources and two different residence fields in each dataset:**

a) **Vital statistics and DAD data were used to count the number of births and compute the fertility rate in each CSD.**

b) **CSD of residence was assigned using two geographic fields from each data source. From vital statistics we used the CSD of residence assigned by Statistics Canada and residential postal code. From DAD data we used the MOHLTC Residence Code and residential postal code. Municipal boundaries associated with the MOHLTC Residence Codes within Metro Toronto and York Region directly correspond to the respective CSD codes. Postal codes in both files were converted into CSD codes using the Single Link Indicator[1] of the Postal Code Conversion File (Statistics Canada 1999).**

The intent of the 1996 CSD level analysis was not to determine the absolute number of births within the various CSDs. Rather, our intent was to examine the relative differences in counts and fertility rates of these CSDs based on the different data sources and different residence identifiers. No effort was made to allocate births with incomplete or invalid postal codes as it was assumed that errors within this field would be random and not systematically biased towards specific areas of geography. If the first three characters of an incomplete postal code were within Ontario, the record was assigned an Ontario residence code with no specific CD or CSD.

Female population counts for the years 1986 through 1996 are based on adjusted counts of the 1986, 1991, and 1996 Census. Population data were obtained from the MOHLTC Provincial Health Planning Database.

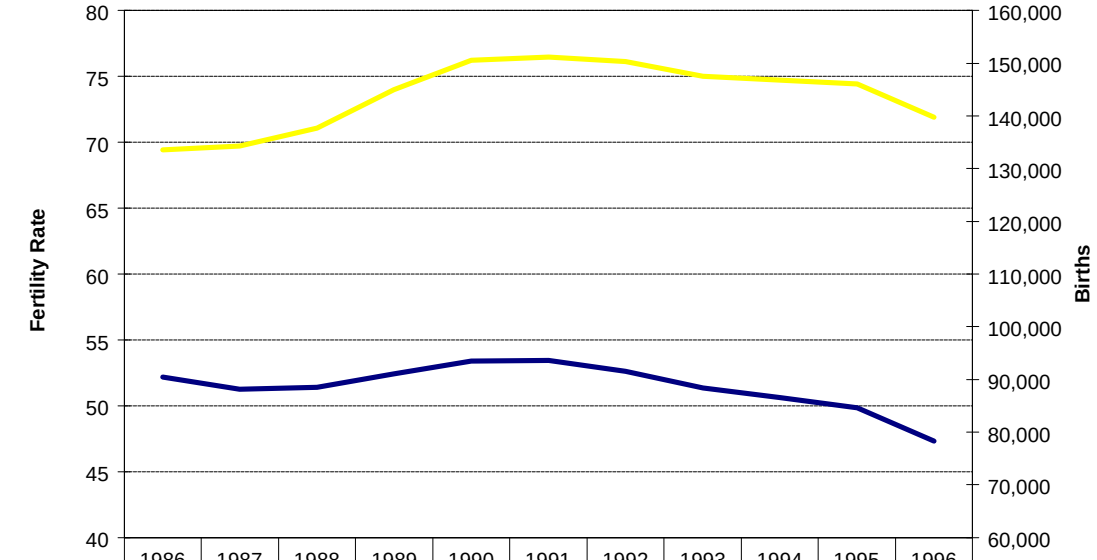
[1] When a postal code falls within more than one enumeration area, the Single Link Indicator allocates the postal code to a single enumeration area based on address (dwelling) counts. As an enumeration area is assigned to a CSD and CD, the postal code assigned to that enumeration area also becomes assigned to the associated CSDs and CDs.

Results

1) CD Fertility Rates, 1986-96

Vital statistics data for Ontario show an increasing number of births in the late 1980's and early 1990's, but the number and accompanying fertility rate has declined slightly over the past five years (Figure 1). A similar trend is evident for most of Ontario's Census Divisions (Appendix 1).

Figure 1. Ontario Births and Fertility Rate
(Source: Vital Statitics)

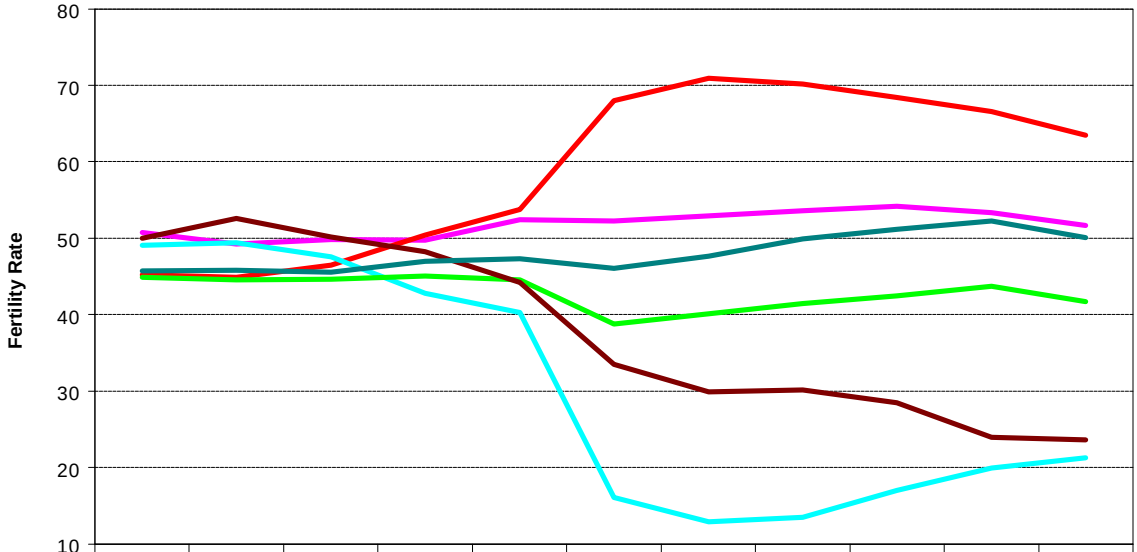


	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Fertility	52.2	51.3	51.4	52.4	53.4	53.4	52.6	51.4	50.6	49.8	47.3
Births	133,558	134,266	137,754	145,021	150,555	151,159	150,317	147,544	146,839	146,036	139,781

2) CSD Fertility Rates, 1986-96

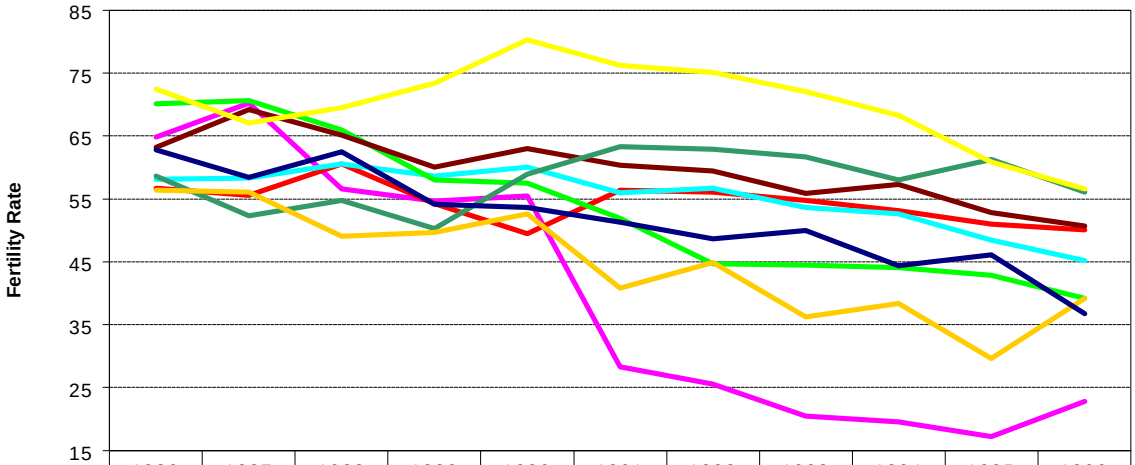
Fertility rates for the CSDs within Metropolitan Toronto and York Region show similar rates within each CD until 1990-1991 at which time the rates for some CSDs begin to diverge substantially (Figures 2 & 3). In Metropolitan Toronto, East York's rate decreases over 50% from 1990 to 1991 and York also shows a substantial decrease. During this time, the City of Toronto's fertility rate increases dramatically. In York Region, the greatest change occurs in Vaughan whose fertility rate decreased over 40% from 1990 to 1991.

Figure 2. Fertility Rate by Census Subdivision for Metropolitan Toronto
(Source: Vital Statistics)



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Scarborough	50.74	49.21	49.80	49.73	52.42	52.31	52.91	53.55	54.18	53.35	51.67
Toronto	45.24	44.89	46.51	50.44	53.80	68.02	70.96	70.21	68.45	66.57	63.50
East York	49.08	49.41	47.60	42.78	40.27	16.08	12.94	13.50	16.97	19.89	21.32
North York	44.92	44.52	44.60	45.09	44.57	38.82	40.12	41.44	42.49	43.67	41.71
York	50.01	52.62	50.18	48.22	44.18	33.49	29.92	30.12	28.50	23.97	23.67
Etobicoke	45.73	45.79	45.59	46.96	47.29	46.08	47.62	49.89	51.18	52.23	50.10

Figure 3. Fertility Rate by Census Subdivision for York Region
(Source: Vital Statitics)



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Vaughan	64.84	70.22	56.62	54.61	55.45	28.34	25.58	20.42	19.49	17.18	22.75
Markham	56.63	55.61	60.51	54.24	49.44	56.39	56.02	54.72	53.09	50.94	50.07
Richmond Hill	58.08	58.30	60.51	58.59	60.06	55.94	56.63	53.68	52.63	48.46	45.12
Whitchurch-Stouffville	70.07	70.59	65.97	58.03	57.53	51.85	44.63	44.49	44.07	42.84	39.19
Aurora	63.20	69.23	65.17	60.04	62.99	60.33	59.49	55.90	57.29	52.82	50.67
Newmarket	58.62	52.36	54.72	50.27	58.95	63.30	62.91	61.70	58.00	61.24	56.03
King	56.38	56.03	49.07	49.62	52.65	40.78	44.91	36.26	38.34	29.64	39.21
East Gwillimbury	62.80	58.45	62.47	54.14	53.60	51.29	48.62	49.96	44.32	46.04	36.68
Georgina	72.46	67.09	69.56	73.41	80.29	76.17	75.10	72.08	68.30	60.90	56.62

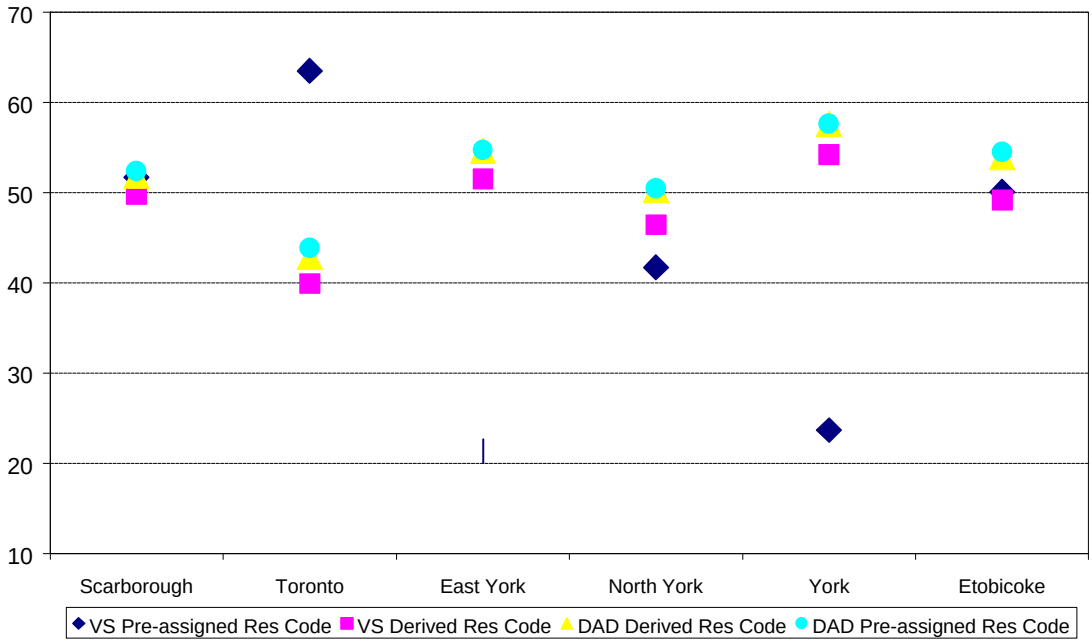
3) 1996 Birth Count and Fertility Rate Comparison

Table 1 shows a comparison of the birth counts based on the pre-assigned and derived residence codes from the vital statistics and DAD data. At the Census Division level the relative differences are not large. For both data sets, the derived counts are slightly lower than the pre-assigned counts due to births with incomplete or invalid Ontario postal codes.

Table 1. Vital Statistics and DAD births counts for York Region and Metropolitan Toronto by pre-assigned and postal code derived CSD of residence.

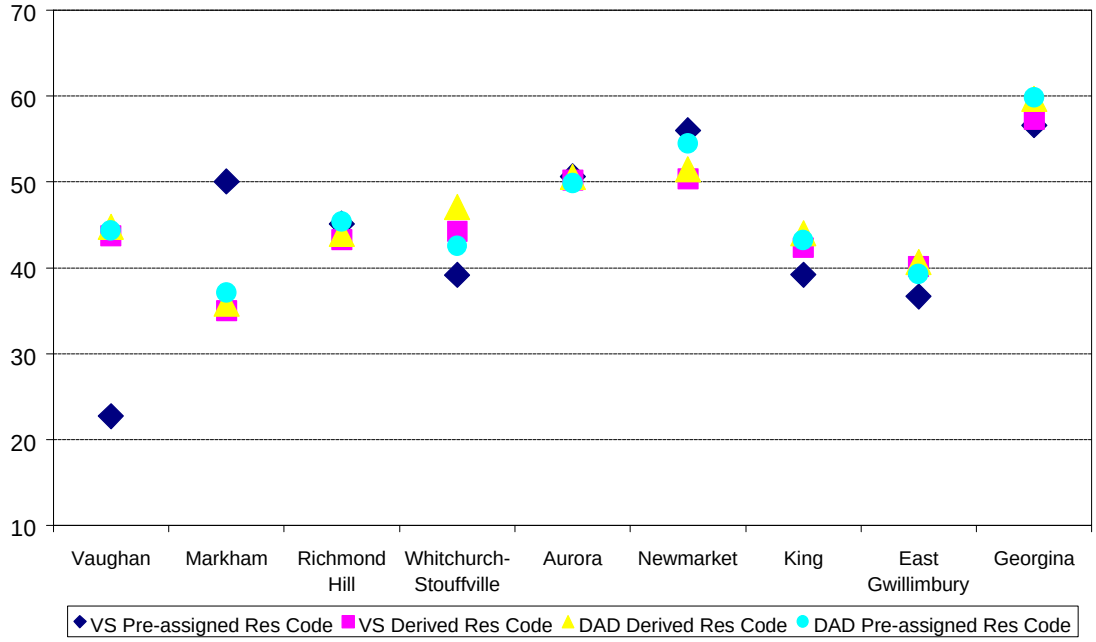
	Vital Statistics		DAD	
	Pre-assigned CSD	Postal Code Derived CSD	Pre-assigned CSD	Postal Code Derived CSD
YORK REGION	**7397**	**7359**	**7632**	**7535**
Vaughan	893	1717	1741	1757
Markham	2565	1793	1899	1837
Richmond Hill	1367	1311	1375	1332
Whitchurch-Stouffville	208	235	226	250
Aurora	501	497	493	500
Newmarket	941	846	915	865
King	187	202	206	210
East Gwillimbury	200	219	214	222
Georgina	533	539	563	562
First Nations Reserve	2	0	0	0
METROPOLITAN TORONTO	**33766**	**31628**	**34175**	**33818**
Scarborough	8166	7871	8275	8194
Toronto	12806	8049	8838	8639
East York	657	1587	1687	1684
North York	6747	7513	8158	8134
York	1007	2306	2450	2451
Etobicoke	4383	4302	4767	4716
Unknown	1	0	0	0
ONTARIO	**139781**	**137772**	**141373**	**141000**

Figure 4. Metropolitan Toronto Fertility Rate by CSD, 1996



Within York Region and Metropolitan Toronto however, large differences appear among some of the municipalities when the Vital Statistics (VS) pre-assigned CSD birth counts are compared to the other CSD based counts. These differences are very evident when the fertility rates for the four types of CSD counts are compared (Figures 4 & 5). In York

Figure 5. York Region Fertility Rate by CSD, 1996



Region, the VS pre-assigned CSD fertility rate for Vaughan is 50% less than the other rates and the Markham rate is 50% higher. Similar differences appear in Metropolitan Toronto when Toronto, York, and East York are examined.

Discussion

Data quality is always an issue when using administrative databases to answer health related questions. This stems from the fact that the databases were not designed to capture information for health services research. However, these data can provide valuable, and often easily accessible, information not available from other sources. This coupled with the demand for more evidence-based decision-making has resulted in a dramatic increase in the use of administrative data for health research and planning over the past decade. The increased use has raised data quality issues that in the past may have been unimportant or may have gone unnoticed. In order for these data to be used effectively and accurately, the quality must be monitored and improved whenever possible.

Our findings indicate that the number of live births in Ontario has fluctuated around 135,000 to 150,000 over the past ten years. However, the increasing population has resulted in an overall decrease in the fertility rate over this same time period. Similar trends are seen for most of the Census Divisions in Ontario.

Fertility rates for the CSD/municipalities within Metropolitan Toronto and York Region also appear to be stable or decreasing from 1986 to 1990. However, around 1990-91 dramatic changes in the fertility rate are evident for a number of these CSDs due to a large, sudden increase or decrease in the number of births. This time period corresponds with the ORG's relocation from Metropolitan Toronto to Thunder Bay - a move that did not include a majority of the staff. It is suspected that the methodology used by the ORG in assigning the mother's municipality changed following this move.

It appears that the data are adequate for public health programming and planning at the Census Division level but not at the municipal or Census Subdivision level. For a large number of births in the vital statistics database, the municipal code (CSD) assigned by Statistics Canada is not in agreement with the residential postal code nor is it in agreement with hospitalization records. This can have serious implications for health planning and public health programs. Municipal based neonatal clinics could be inappropriately staffed

and estimates of future health service need and population growth could be grossly over- or under-estimated.

Our results also show a greater number of births occurring in Ontario hospitals according to the DAD than occurring in Ontario overall based on vital statistics data. Reasons for this discrepancy are not clear. Differences may be due to differences in the assignment of residence in the two systems; some of the births assigned to Metropolitan Toronto and York Region in the DAD may have been assigned to other jurisdictions in the vital statistics data. If this is true, the totals from the two databases should be similar for Ontario residents giving birth in Ontario, with the VS database showing a slightly greater number of births due to the fact that it captures births occurring in and out of hospital. Another reason for the discrepancy may be the fact that a number births each year are not captured by the ORG database due to incomplete documentation. In 1996 this represented about 1% of all births.

Conclusions and Recommendations

The accuracy of the mother's municipal residence code within the vital statistics database, subsequently recoded to CSD by Statistics Canada, needs to be examined. The Ontario Vital Statistics Improvement Project is reviewing the ORG database. Much of this review appears to be focussed on the development of a new data capture and retrieval system. However, data quality of previous years, as well as within the new data system, needs to be addressed. Included in this should be an examination of how the residence code is assigned, development of a system that cross validates this code with the postal code and street address, and coordination with Ontario hospitals to ensure that codes are assigned in a consistent fashion.

References

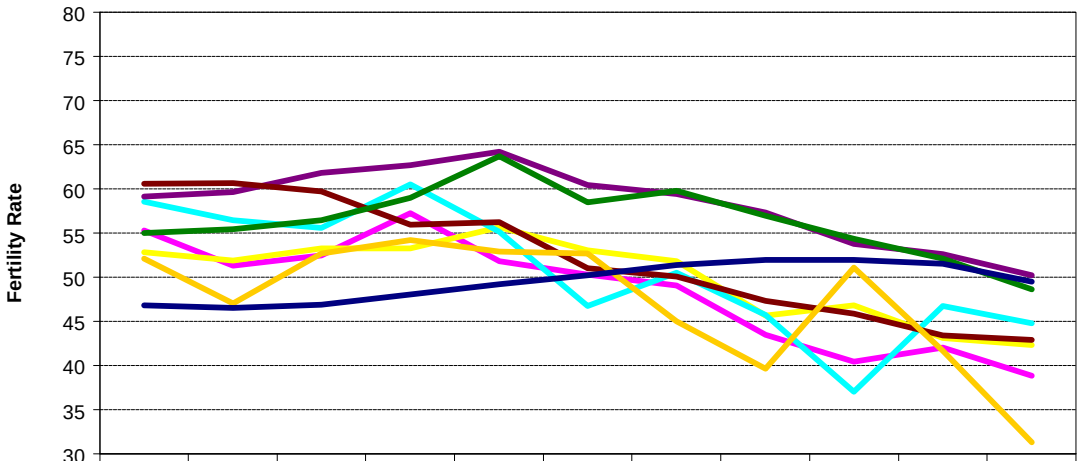
Statistics Canada. 1999. Postal Code Conversion File. May 1999 Postal Codes. References Guide. Ministry of Industry, Ottawa, Canada.

Appendix 1

Fertility Rates for Ontario Census Divisions

(Data Source: Vital Statistics)

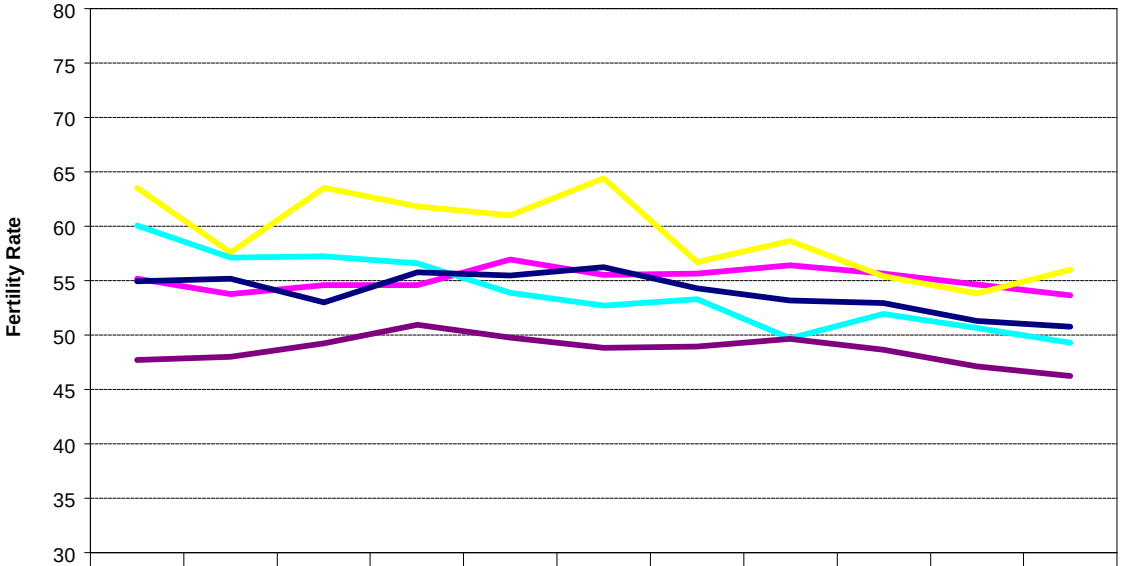
Fertility Rate by Census Division, Central East Ontario and Toronto



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(14) Northumberland County	55.3	51.3	52.5	57.3	51.9	50.3	49.0	43.5	40.5	42.1	38.9
(15) Peterborough County	52.8	51.9	53.2	53.3	55.6	53.0	51.8	45.7	46.8	43.1	42.4
(16) Victoria County	58.5	56.5	55.6	60.5	55.2	46.7	50.5	45.8	37.1	46.7	44.8
(18) Durham Region	59.1	59.6	61.8	62.7	64.2	60.4	59.4	57.3	53.8	52.6	50.2
(19) York Region	60.6	60.7	59.7	56.0	56.2	51.0	50.1	47.4	45.9	43.4	42.9
(43) Simcoe County	55.0	55.4	56.5	59.0	63.7	58.5	59.8	57.0	54.4	52.1	48.7
(46) Haliburton County	52.1	47.0	52.7	54.3	52.9	52.7	45.0	39.7	51.1	41.7	31.3
(20) Metropolitan Toronto	46.8	46.5	46.9	48.0	49.2	50.2	51.4	51.9	52.0	51.5	49.5

Year

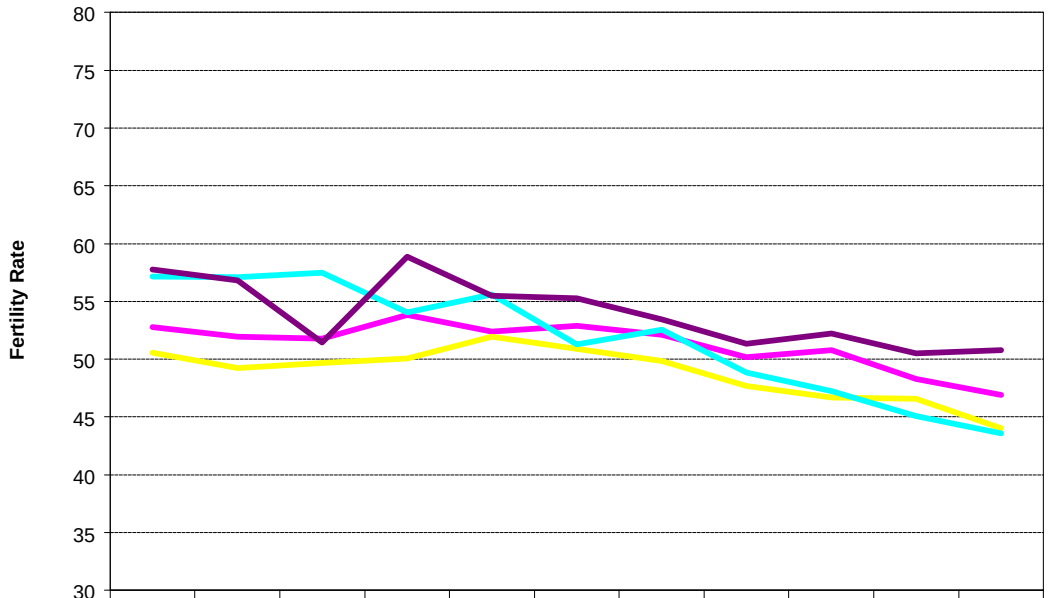
Fertility Rate by Census Division, Central West Ontario



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(21) Peel Region	55.2	53.8	54.6	54.6	56.9	55.5	55.6	56.4	55.7	54.7	53.6
(22) Dufferin County	63.6	57.6	63.5	61.8	61.0	64.4	56.7	58.7	55.4	53.8	56.0
(23) Wellington County	60.1	57.1	57.2	56.6	53.9	52.7	53.3	49.7	51.9	50.6	49.3
(24) Halton Region	47.7	48.0	49.2	50.9	49.8	48.8	48.9	49.6	48.6	47.1	46.2
(30) Waterloo Region	54.9	55.2	53.0	55.8	55.5	56.3	54.3	53.2	53.0	51.3	50.8

Year

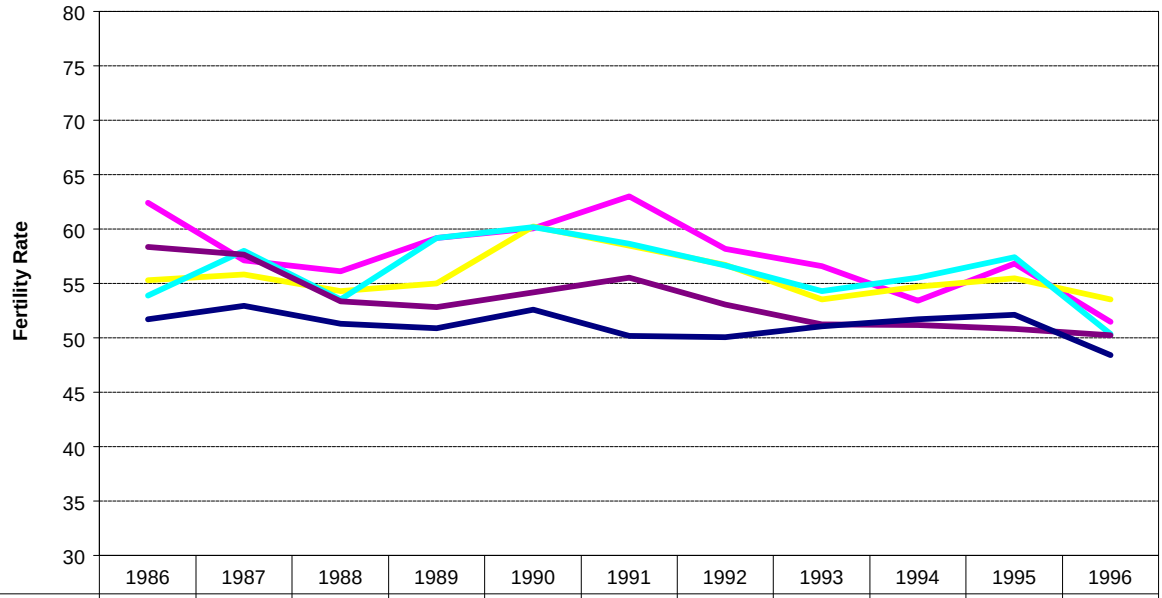
Fertility Rate by Census Division, Central South Ontario



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(25) Hamilton-Wentworth Region	52.8	51.9	51.8	53.8	52.4	52.9	52.1	50.2	50.7	48.3	46.9
(26) Niagara Region	50.6	49.2	49.6	50.1	51.9	50.9	49.9	47.6	46.7	46.6	44.0
(28) Haldimand-Norfolk Region	57.2	57.1	57.5	54.0	55.6	51.3	52.6	48.9	47.2	45.1	43.6
(29) Brant County	57.8	56.8	51.5	58.9	55.5	55.3	53.4	51.3	52.2	50.5	50.8

Year

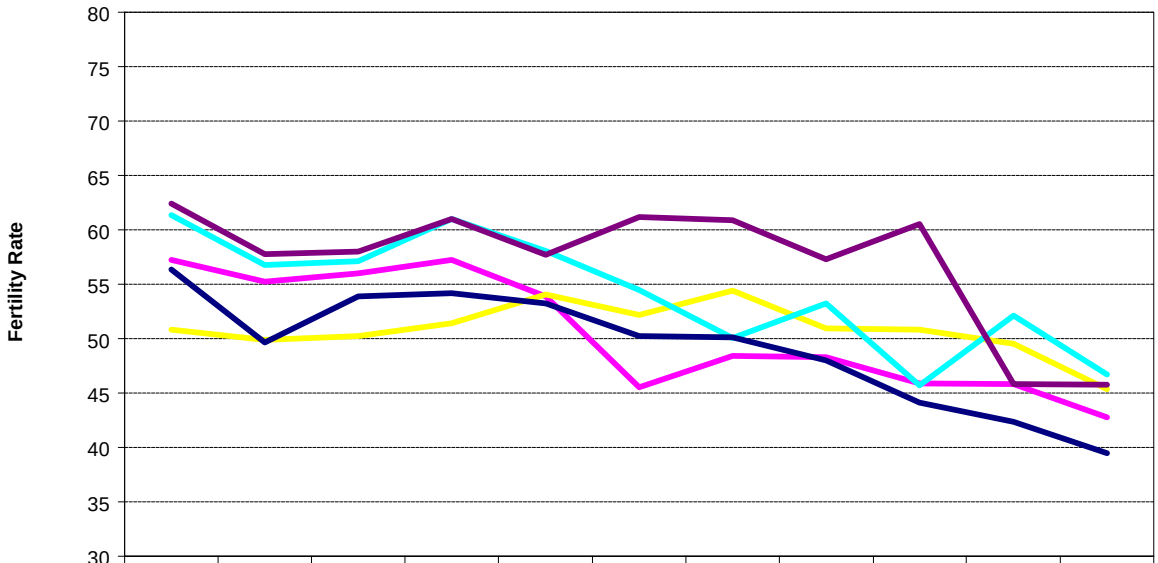
Fertility Rate by Census Division, Southwestern Ontario, Part 1



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(31) Perth County	62.4	57.1	56.1	59.2	60.1	63.0	58.2	56.6	53.4	56.8	51.5
(32) Oxford County	55.3	55.9	54.3	55.0	60.2	58.5	56.7	53.6	54.7	55.4	53.5
(34) Elgin County	53.9	58.0	53.5	59.2	60.2	58.6	56.7	54.3	55.5	57.4	50.4
(36) Kent County	58.4	57.7	53.3	52.8	54.2	55.5	53.1	51.2	51.2	50.8	50.2
(37) Essex County	51.7	52.9	51.3	50.9	52.6	50.2	50.1	51.0	51.7	52.1	48.4

Year

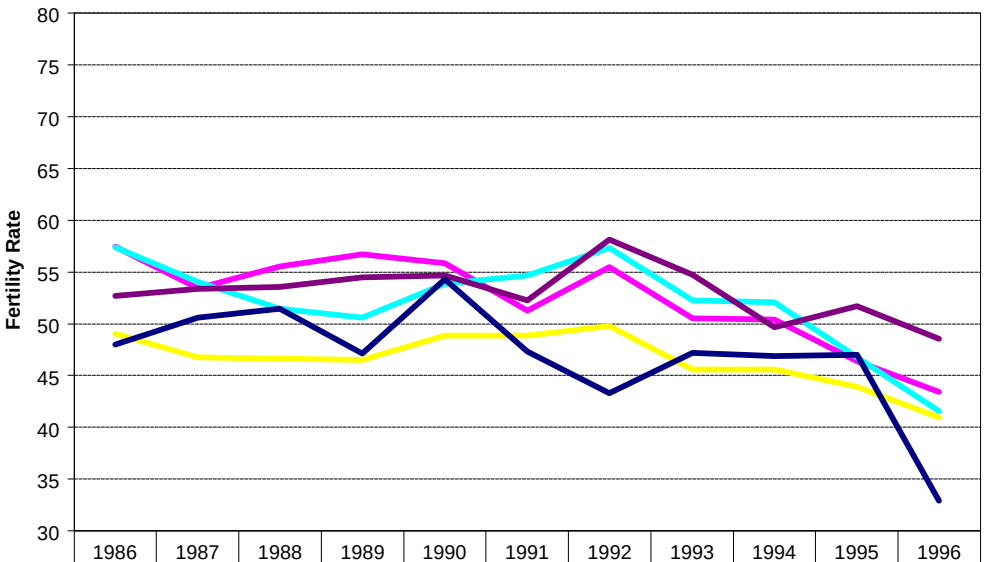
Fertility Rate by Census Division, Southwestern Ontario, Part 2



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(38) Lambton County	57.2	55.2	56.0	57.3	53.9	45.6	48.4	48.3	45.9	45.9	42.8
(39) Middlesex County	50.8	49.9	50.2	51.4	54.1	52.2	54.4	51.0	50.8	49.5	45.4
(40) Huron County	61.3	56.8	57.1	61.0	58.1	54.5	50.1	53.2	45.7	52.1	46.7
(41) Bruce County	62.4	57.8	58.0	61.0	57.7	61.2	60.9	57.3	60.5	45.8	45.8
(42) Grey County	56.3	49.6	53.9	54.2	53.3	50.3	50.1	48.0	44.1	42.3	39.5

Year

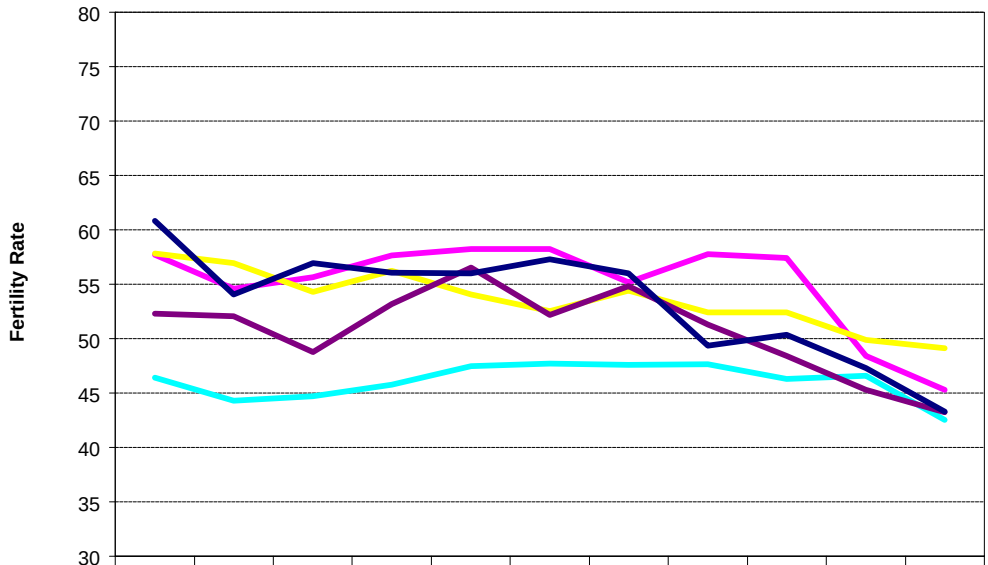
Fertility Rate by Census Division, Eastern Ontario, Part 1



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(1) Stormont, Dundas and Glengarry United Counties	57.5	53.5	55.5	56.7	55.8	51.3	55.5	50.5	50.4	46.4	43.4
(10) Frontenac County	49.0	46.8	46.7	46.5	48.8	48.9	49.8	45.6	45.6	43.9	41.0
(11) Lennox and Addington County	57.4	54.0	51.4	50.6	53.8	54.6	57.3	52.2	52.1	46.7	41.5
(12) Hastings County	52.7	53.3	53.6	54.5	54.7	52.3	58.1	54.7	49.7	51.7	48.6
(13) Prince Edward County	48.0	50.6	51.4	47.1	54.3	47.3	43.3	47.2	46.9	47.0	32.9

Year

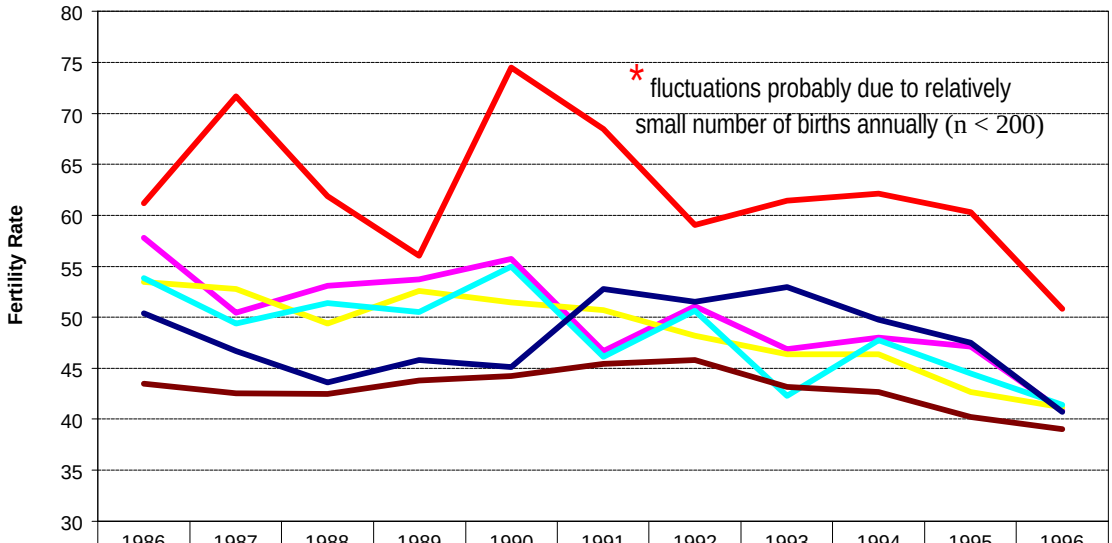
Fertility Rate by Census Division, Eastern Ontario, Part 2



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(2) Prescott and Russell United Counties	57.7	54.6	55.6	57.6	58.3	58.3	55.2	57.7	57.4	48.4	45.3
(47) Renfrew County	57.9	56.9	54.3	56.2	54.1	52.6	54.4	52.4	52.4	49.9	49.1
(6) Ottawa-Carleton Region	46.4	44.3	44.7	45.8	47.5	47.7	47.6	47.6	46.3	46.6	42.5
(7) Leeds and Grenville United Counties	52.3	52.1	48.8	53.2	56.5	52.2	54.9	51.3	48.4	45.3	43.2
(9) Lanark County	60.9	54.1	57.0	56.1	56.0	57.3	56.0	49.4	50.4	47.3	43.3

Year

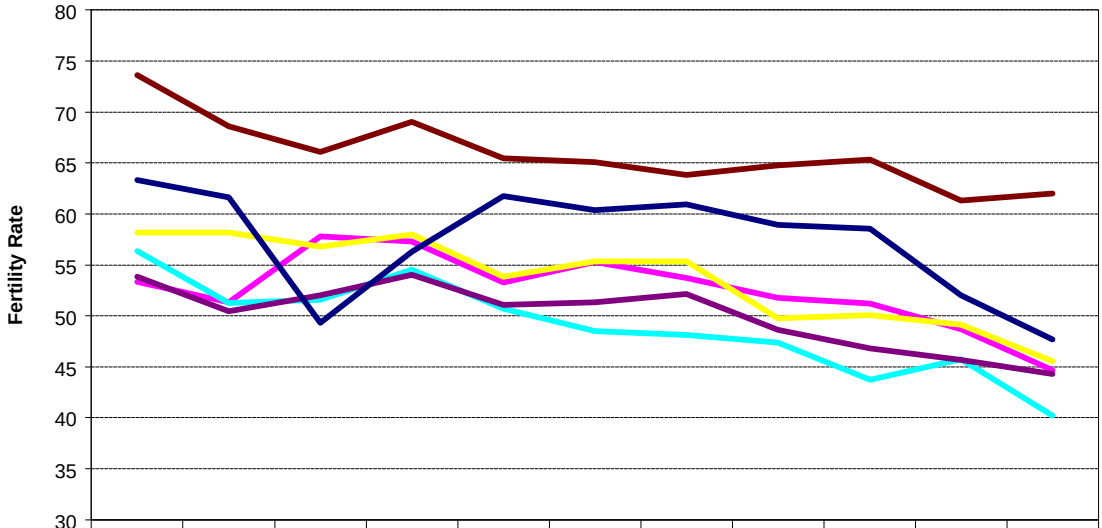
Fertility Rate by Census Division, Northern Ontario, Part 1



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(44) Muskoka District	57.8	50.4	53.1	53.7	55.7	46.7	51.1	46.9	48.0	47.1	40.8
(48) Nipissing District	53.5	52.8	49.4	52.6	51.5	50.7	48.2	46.4	46.3	42.6	41.2
(49) Parry Sound District	53.8	49.4	51.4	50.5	54.9	46.1	50.6	42.3	47.8	44.5	41.4
(51) Manitoulin District	61.2	71.7	61.9	56.0	74.5	68.5	59.1	61.5	62.1	60.3	50.8
(52) Sudbury District	50.4	46.6	43.6	45.8	45.1	52.8	51.5	52.9	49.7	47.5	40.7
(53) Sudbury Region	43.5	42.6	42.5	43.8	44.2	45.4	45.8	43.2	42.6	40.2	39.0

Year

Fertility Rate by Census Division, Northern Ontario, Part 2



	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(54) Timiskaming District	53.3	51.3	57.8	57.3	53.2	55.3	53.7	51.8	51.2	48.7	44.7
(56) Cochrane District	58.1	58.2	56.8	58.0	53.8	55.4	55.3	49.7	50.1	49.1	45.5
(57) Algoma District	56.4	51.3	51.6	54.5	50.7	48.5	48.1	47.4	43.7	45.7	40.2
(58) Thunder Bay District	53.9	50.5	52.0	54.0	51.1	51.3	52.1	48.6	46.8	45.7	44.3
(59) Rainy River District	63.3	61.6	49.3	56.3	61.7	60.3	60.9	58.9	58.6	52.0	47.7
(60) Kenora District	73.6	68.6	66.1	69.0	65.5	65.1	63.8	64.8	65.3	61.3	62.0

Year