Questioning one of life's certainties:

A comparison of three mortality data sets.

July 1999

Central East Health Information Partnership

Questioning one of life's certainties:

A comparison of three mortality data sets.

G.L. Woodward[1], T. Delmore[2], G. Fehringer[3], F. Goettler[4], B. Guarda[5], J.J. McGurran[6], and M. Pietrusiak[7]

1) Central East Health Information Partnership, Newmarket, Ontario
2) Halton Regional Health Department, Oakville, Ontario
3) Cancer Care Ontario, Toronto, Ontario
4) City of Toronto Department of Public Health, Toronto, Ontario
5) Simcoe County District Health Unit, Barrie, Ontario
6) Department of Public Health Sciences, University of Alberta, Edmonton, Alberta
7) Durham Region Health Department, Whitby, Ontario

July 1999

17250 Yonge Street, Box 234, Newmarket, Ontario L3Y 4X1 Tel: (905) 895-1231 Ext.1210 Fax: (905) 895-0848 URL: http://www.cehip.org

Executive Summary

Public health units and district health councils are increasing their use of population health data, such as vital statistics data, for local research and health planning purposes. It is important to assure that the data they are using are of the highest possible quality.

To examine the quality of mortality data, we have compared deaths counts derived from three data sources, Ontario Ministry of Health Public Health Branch (HELPS data), the Ontario Ministry of Health Provincial Health Planning Database (PHPD data), and Cancer Care Ontario (CCO data). Although the different data sets yielded similar results, they were not identical.

Based on our findings we have made three recommendations. First, there should be one set of criteria for inclusion/exclusion of cases and for coding/labeling that all users agree to and apply consistently. Second, there should be one original vital statistics data file which all researchers and planners have access to. Third, there should be regular communication between the users of the vital statistics data and the Registrar General, the collector and owner of these data.

Collaboration between the owners and users of vital statistics data will result in the development of a more reliable data file and will enhance the value of the research and health planning projects that make use of the data.

Introduction

The Central East Health Information Partnership (CEHIP) is a consortium of District Health Councils, Boards of Health and Universities in Central East Ontario. Our goal is to improve the partners' capabilities and effectiveness in health planning, research and education. As part of addressing this goal we examine the quality of the health data available to CEHIP partners.

In July of 1997, the Central East Health Information Partnership (CEHIP) published a report reviewing the Vital Statistics files provided to us by the HELPS initiative of the Public Health Branch of the Ontario Ministry of Health (McGurran 1997). This overview of the mortality, live birth, and stillbirth data files noted a number of problems and recommended that further examination of the files be carried out by the Public Health Branch in its role as data provider to the local health units. A subsequent analysis of live birth data, with respect to low birth weight rates in Ontario (McGurran et al 1998), revealed additional quality problems with this data set.

Currently, mortality statistics can be obtained from a number of sources, including:

- the Health Planning System (HELPS), Ontario Ministry of Health Public Health Branch;
- the Ontario Ministry of Health Provincial Health Planning Database (PHPD, Ontario Ministry of Health Information, Planning, and Evaluation Branch), and;
- Cancer Care Ontario (CCO).

Although the point of origin of the data for these three sources is the Office of the Registrar General of Ontario, only CCO receive the data directly from this agency. The Ontario Ministry of Health does not receive the data directly from the Registrar General; HELPS and PHPD data sets are derived from a mortality file that has been edited by Statistics Canada.

The purpose of this report is to explore further the quality of the vital statistics mortality data. Specifically, this report compares mortality data from three sources: Ontario Ministry of Health Public Health Branch (HELPS data), the Ontario Ministry of Health Provincial Health Planning Database (PHPD data), and Cancer Care Ontario (CCO data).

Methods

Durham Region Health Department, Halton Region Health Department, City of Toronto Public Health Department, Simcoe County District Health Unit, and CCO were invited and agreed to participate in the project.

All public health units were asked to provide CEHIP with the number of male and female deaths per year (1986-95) for residents of their health unit area. Three relatively common causes of death were selected based on the nature of the codes according to the International Classification of Diseases, Ninth Revision (ICD-9, WHO 1977):

1. female breast cancer (ICD-9 175), which was chosen because it is sex specific and falls under one three digit ICD-9 code;
2. ischemic heart disease (ICD-9 410-414), which was chosen because it spans more than one ICD-9 code;
3. motor vehicle traffic accidents (ICD-9 E810-E819), which was chosen because it involves ICD-9 External Causes of Injury and Poisoning (E) codes.

Using their own mortality data files, CCO were asked to provide the number of male and female deaths due to the selected causes for all four of the participating health unit areas. CEHIP also calculated the number of deaths for the four participating health unit areas using HELPS mortality data and mortality data from the PHPD.

Public health unit, CCO, and CEHIP results were compiled and charted by CEHIP to compare the number of deaths reported by the various sources. The maximum and minimum percent difference per year per chart was then calculated. These differences represent the greatest and smallest difference between two of the three results sets per year divided by the greatest number of deaths for the same year and were expressed as a percentage (multiplied by 100). The greatest maximum difference and smallest minimum difference for each chart then were tabulated. For example, the maximum percent difference in Chart 2 occurs in 1990 and is 12.5% [(56-49)/56 x 100]. The minimum percent difference for this chart is 0%. All results then were shared with the participants for comment and discussion.

Results

Mortality results were received from all participants. HELPS results from the public health units and CEHIP were identical except for breast cancer in Toronto, 1993 (see below). Because of this, HELPS results were represented by only one line in each chart.

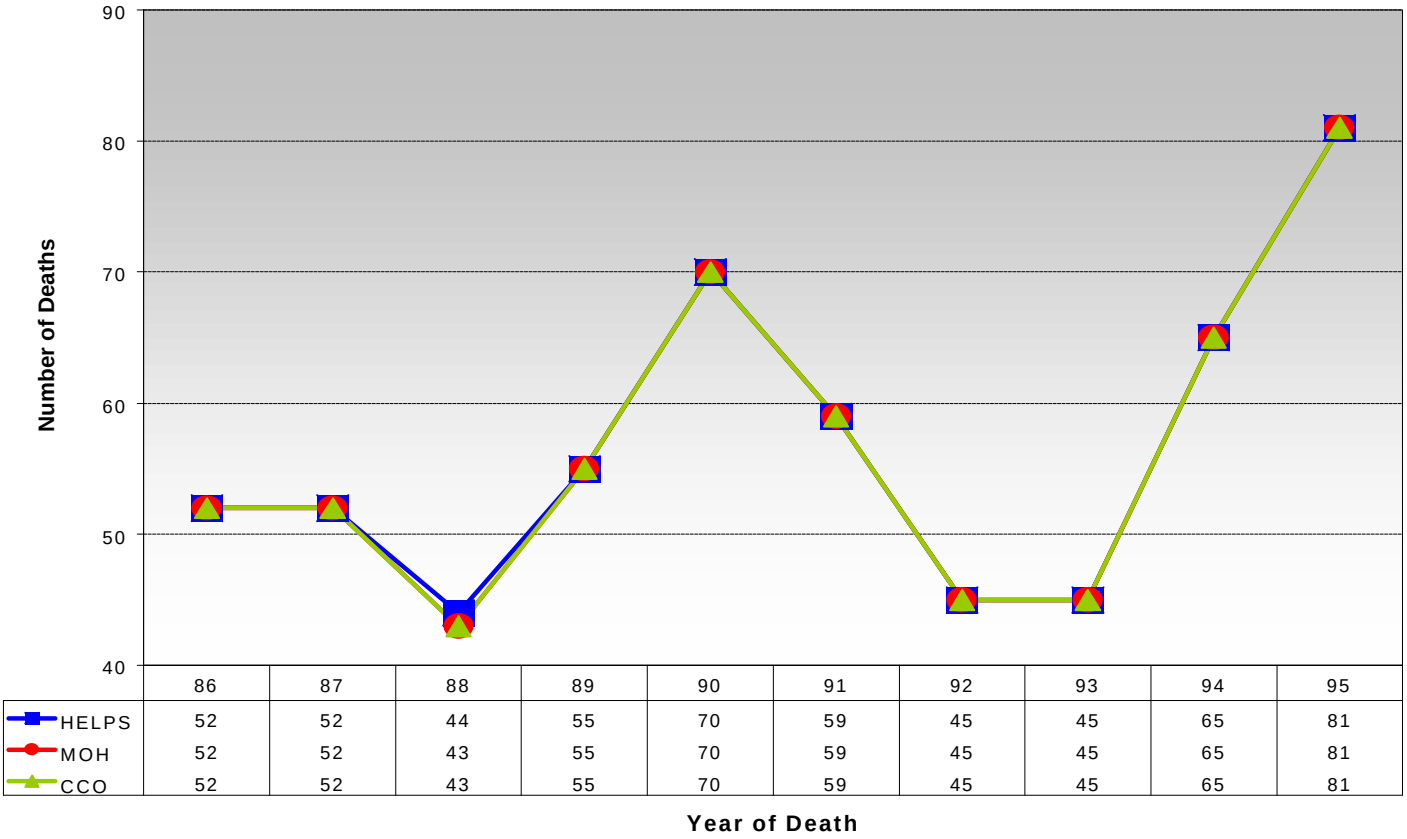
A slight discrepancy did exist between the 1993 HELPS breast cancer results from CEHIP and those from the City of Toronto. CEHIP results for 1993 identified two deaths of males due to female breast cancer that were not identified in the Toronto results. However, CEHIP's results

showed two fewer female deaths in 1993 (n=418) than Toronto (n=420). The two male deaths were assumed to be incorrectly coded female deaths and were added to the female total for 1993.

Deaths due to female breast cancer are summarised by public health unit area in Charts 1 through 4. Results from all three data sources are very similar (Table 1). Most results are identical or within one death of each other. In only one area and year, Halton 1990, is there a notable difference (12.5%), with CCO reporting 49 deaths and the other two data sources reporting 56.

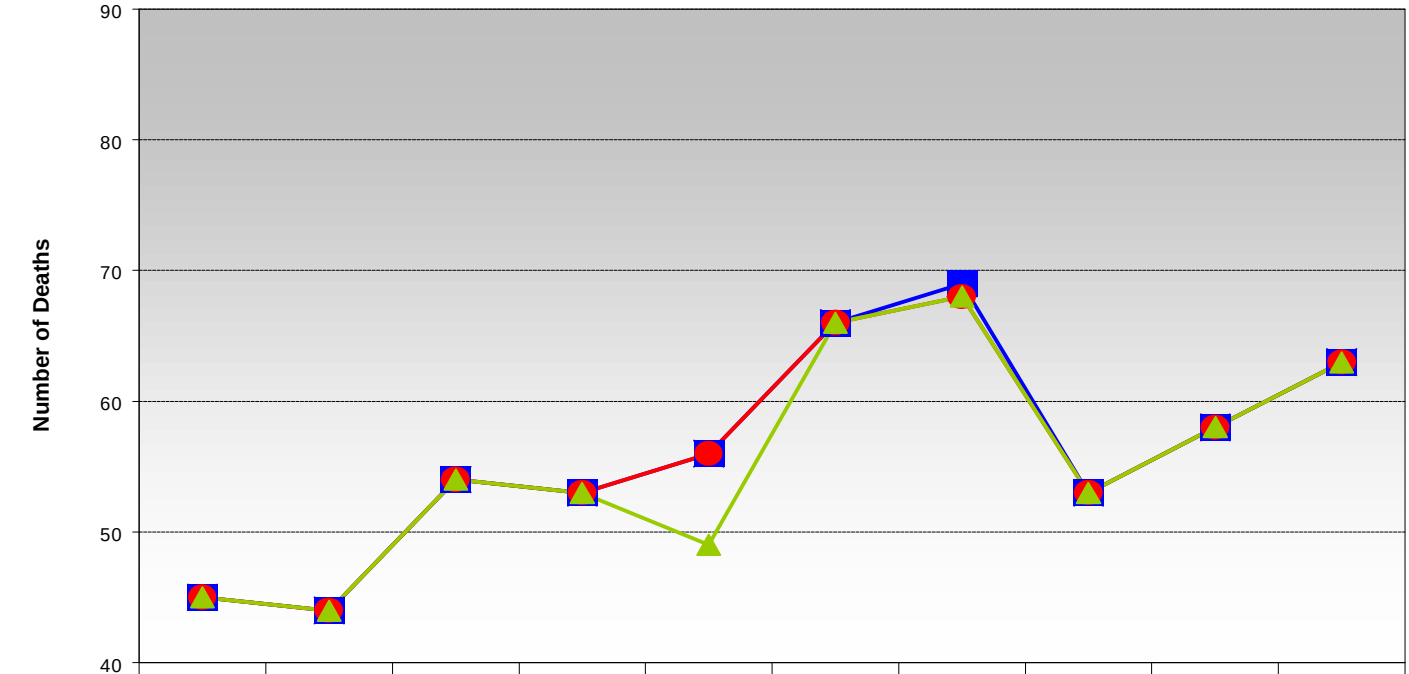
Charts 5 through 8 show the findings for ischemic heart disease. As for breast cancer, the number of deaths recorded by each data set is very similar, although the level of agreement is not as high (Table 1). Complete agreement among the data sets is rare but the difference among the sets is relatively small. HELPS data regularly shows more deaths that the CCO and PHPD in all health unit areas from 1986 up to and including 1992.

Motor vehicle collision (MVC) deaths are summarised in Charts 9-12. Of the three causes of death included in this report, MVCs account for the fewest number of deaths in all health unit areas. Typically, MVCs account for 30-40 deaths of Durham, Halton, and Simcoe residents, and 100-150 in Toronto but the number varies quite unpredictably between years. The results also vary among the three data sources more than for breast cancer and ischemic heart disease (Table 1). Although the absolute difference between the sources is not very large, the proportional difference is 5% or more in a number of instances due to the low total number of events.



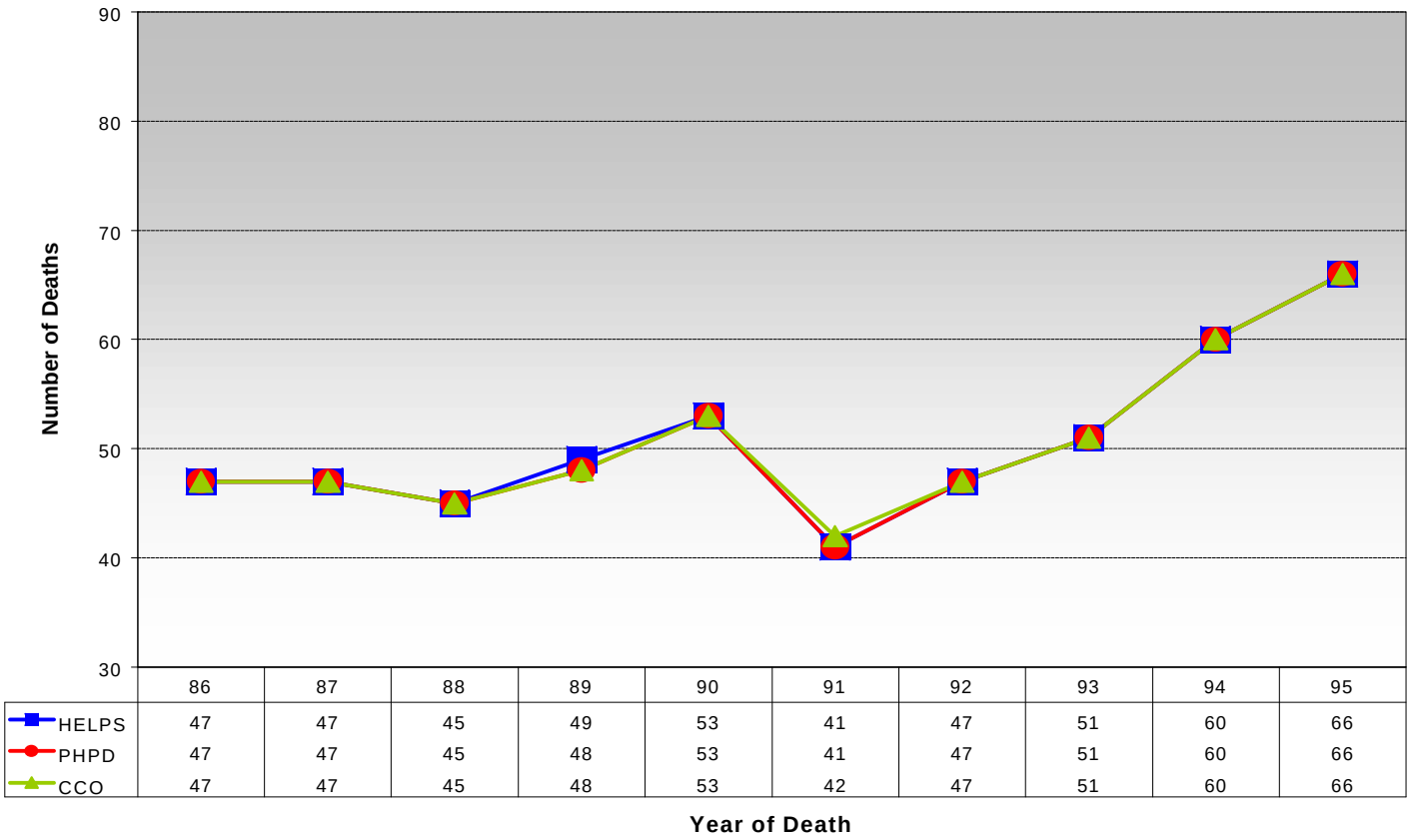
	86	87	88	89	90	91	92	93	94	95
HELPS	52	52	44	55	70	59	45	45	65	81
MOH	52	52	43	55	70	59	45	45	65	81
CCO	52	52	43	55	70	59	45	45	65	81

Chart 1. Deaths of Durham Region residents due to female breast cancer.



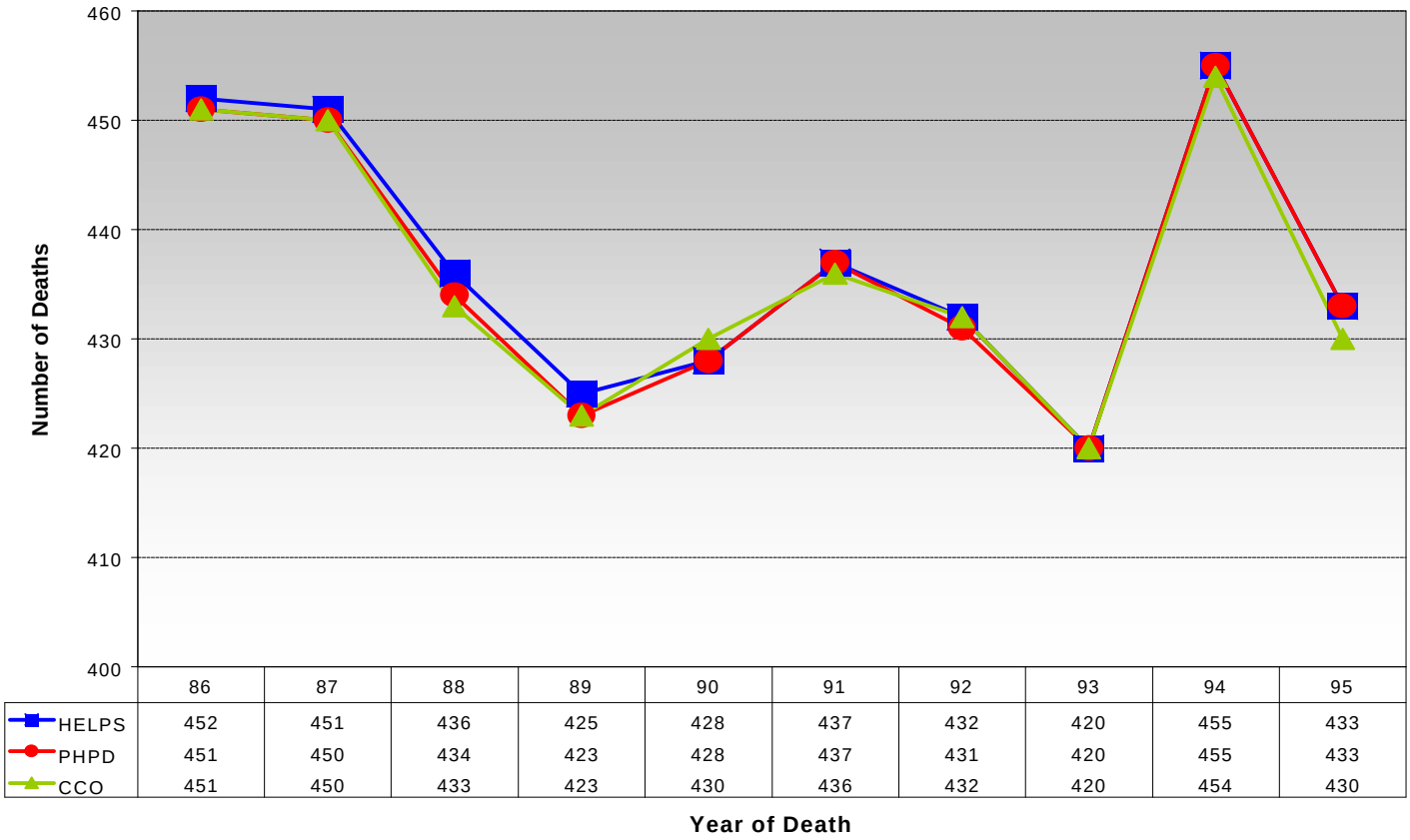
	86	87	88	89	90	91	92	93	94	95
HELPS	45	44	54	53	56	66	69	53	58	63
PHPD	45	44	54	53	56	66	68	53	58	63
CCO	45	44	54	53	49	66	68	53	58	63

Chart 2. Deaths of Halton Region residents due to female breast cancer.



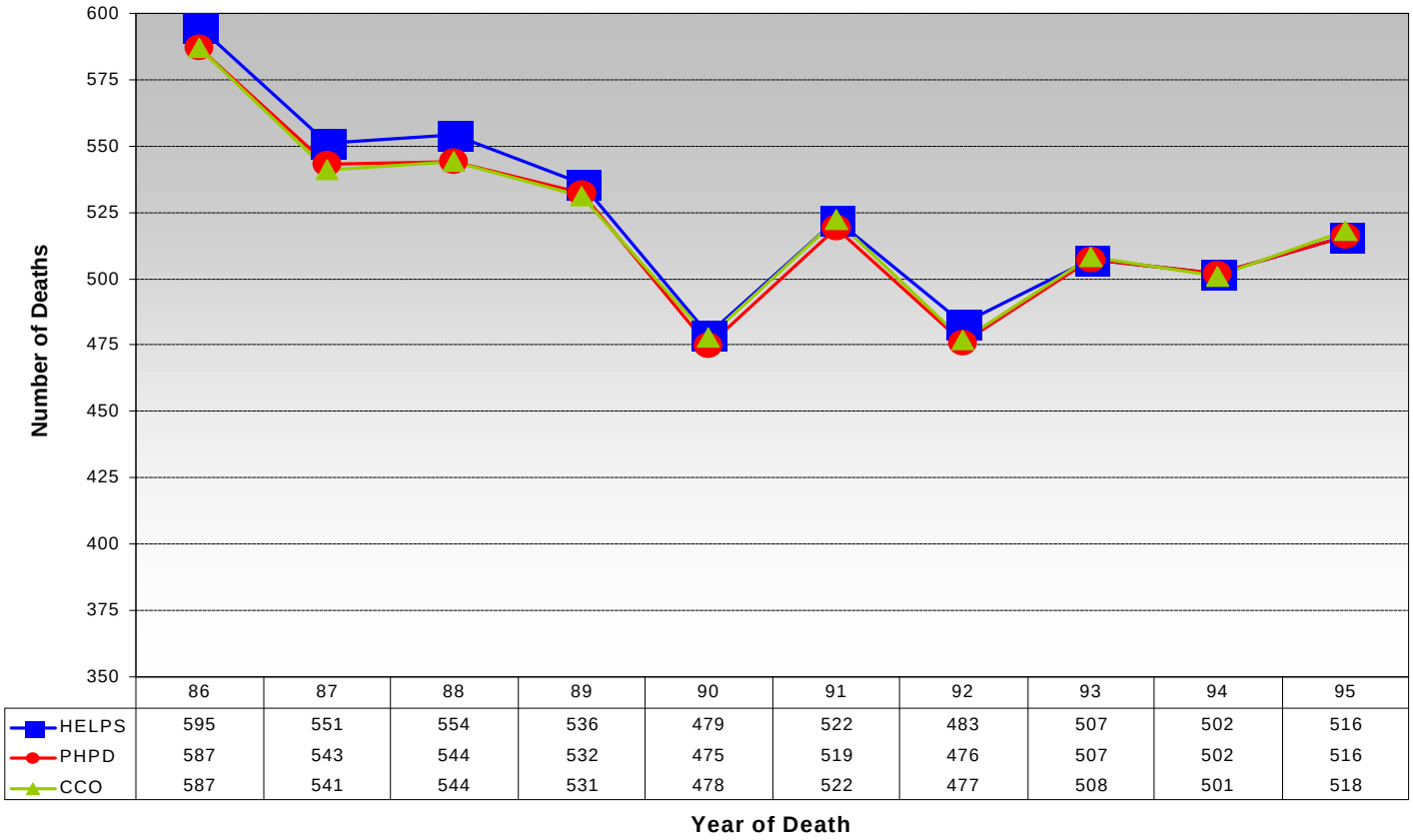
	86	87	88	89	90	91	92	93	94	95
HELPS	47	47	45	49	53	41	47	51	60	66
PHPD	47	47	45	48	53	41	47	51	60	66
CCO	47	47	45	48	53	42	47	51	60	66

Chart 3. Deaths of Simcoe County residents due to female breast cancer.



	86	87	88	89	90	91	92	93	94	95
HELPS	452	451	436	425	428	437	432	420	455	433
PHPD	451	450	434	423	428	437	431	420	455	433
CCO	451	450	433	423	430	436	432	420	454	430

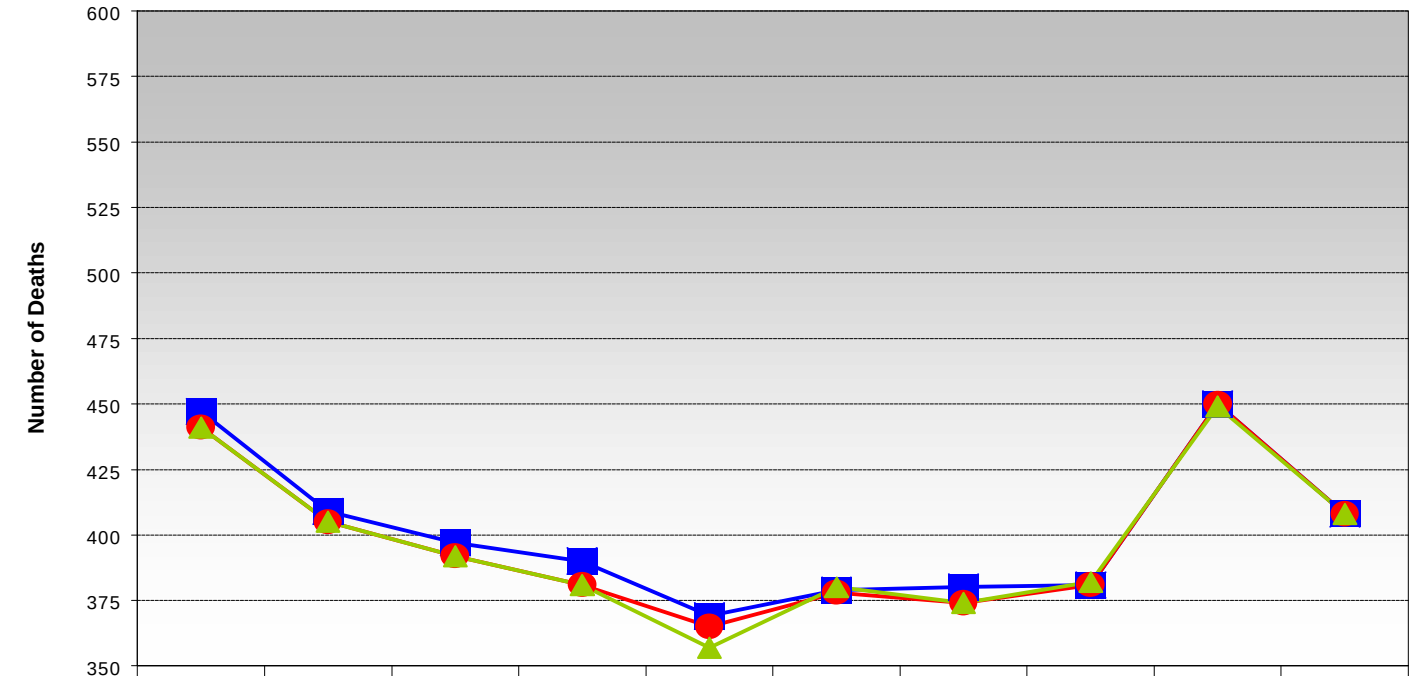
Chart 4. Deaths of Toronto residents due to female breast cancer.



	86	87	88	89	90	91	92	93	94	95
HELPS	595	551	554	536	479	522	483	507	502	516
PHPD	587	543	544	532	475	519	476	507	502	516
CCO	587	541	544	531	478	522	477	508	501	518

Year of Death

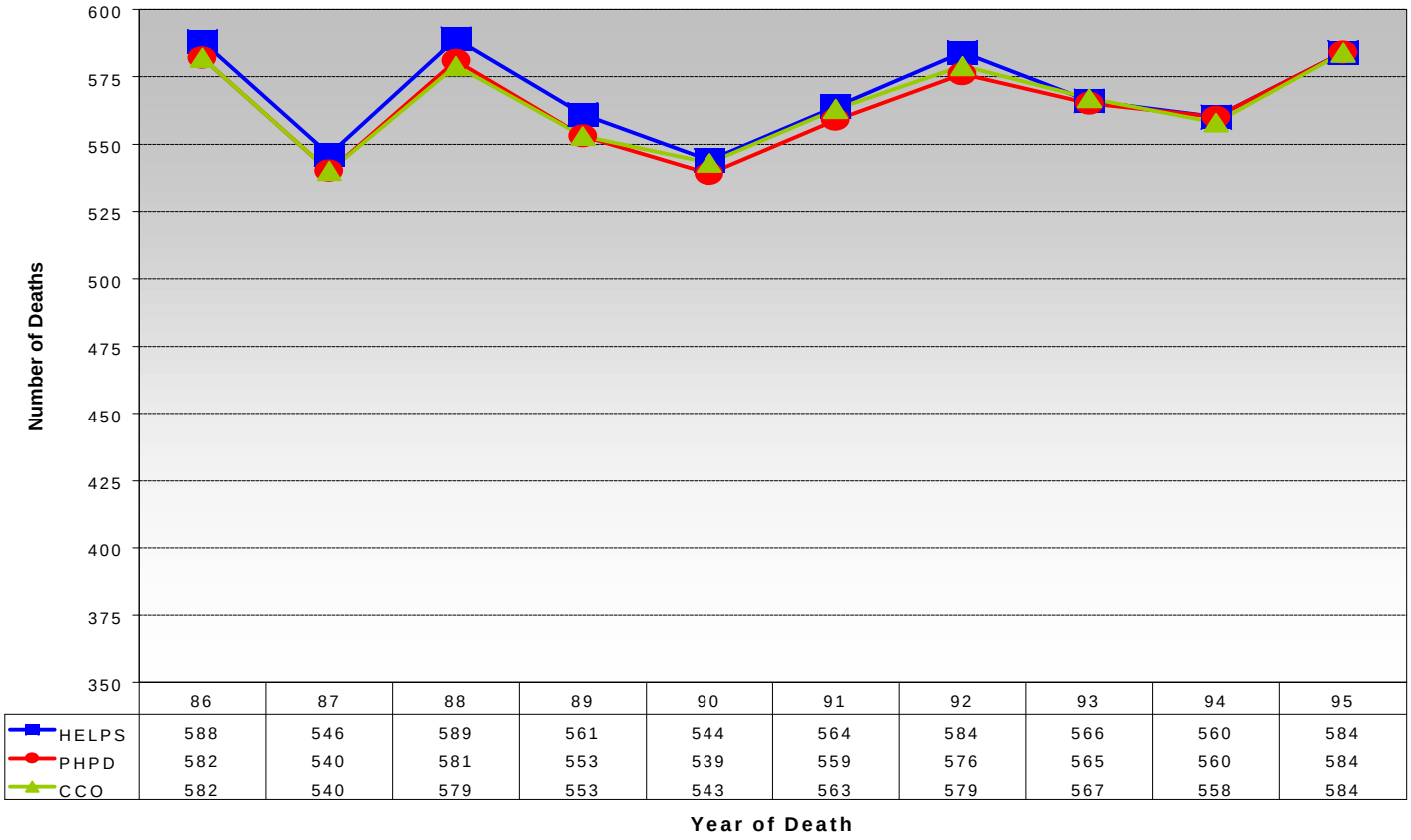
Chart 5. Deaths of Durham Region residents due to ischemic heart disease.



	86	87	88	89	90	91	92	93	94	95
HELPS	447	409	397	390	369	379	380	381	450	408
PHPD	441	405	392	381	365	378	374	381	450	408
CCO	441	405	392	381	357	380	374	382	449	408

Year of Death

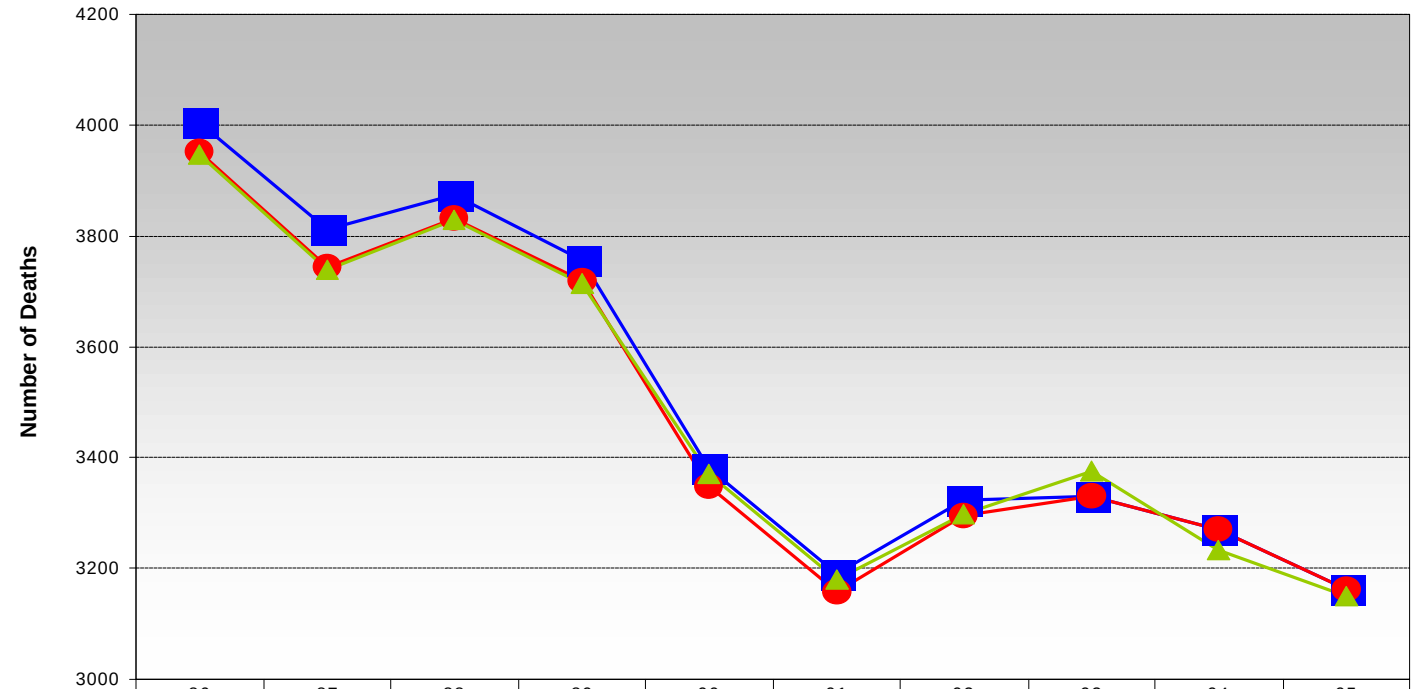
Chart 6. Death of Halton Region residents due to ischemic heart disease.



	86	87	88	89	90	91	92	93	94	95
HELPS	588	546	589	561	544	564	584	566	560	584
PHPD	582	540	581	553	539	559	576	565	560	584
CCO	582	540	579	553	543	563	579	567	558	584

Year of Death

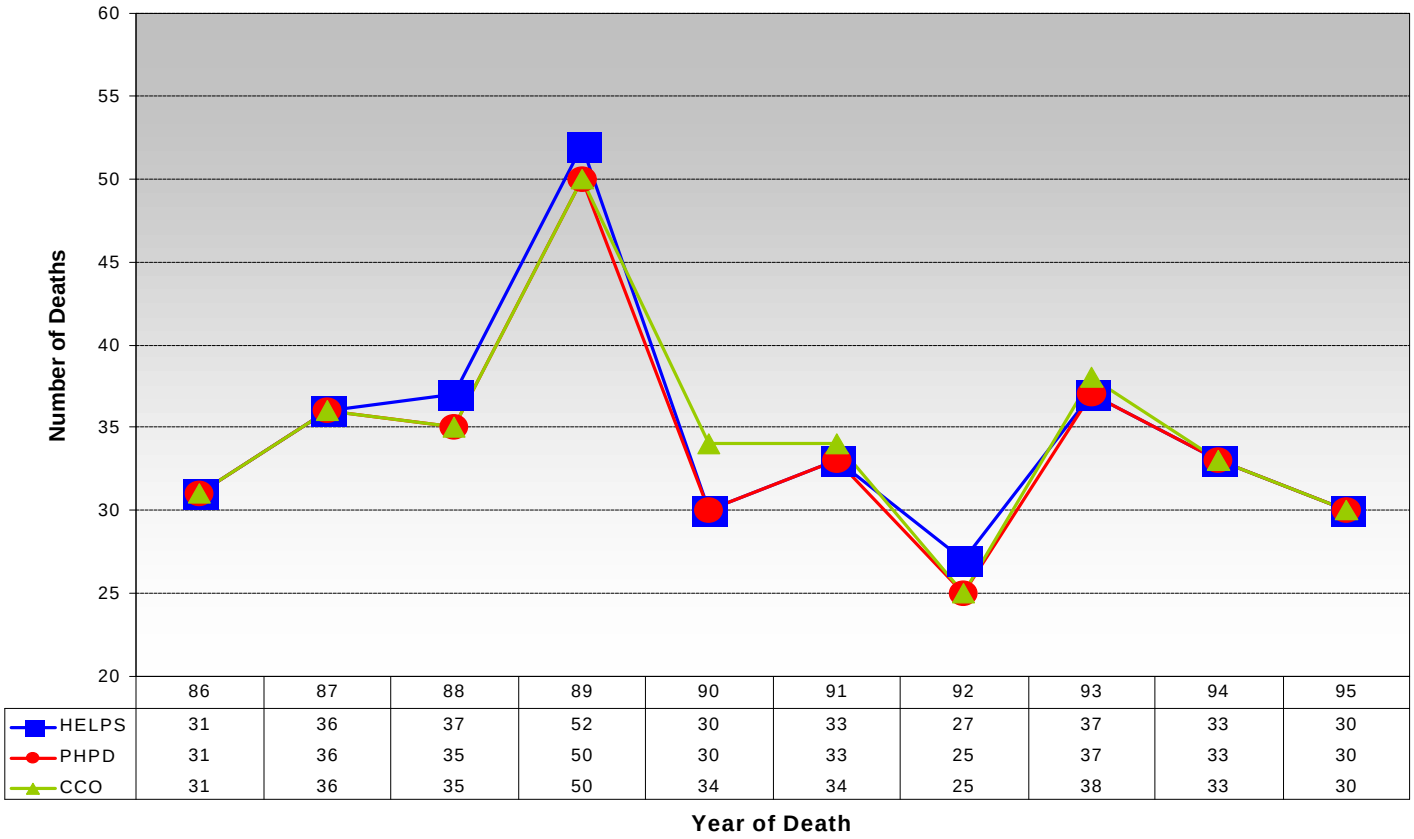
Chart 7. Deaths of Simcoe County residents due to ischemic heart disease.



	86	87	88	89	90	91	92	93	94	95
HELPS	4005	3811	3875	3756	3380	3190	3322	3331	3270	3162
PHPD	3952	3744	3832	3720	3347	3158	3295	3331	3270	3162
CCO	3948	3739	3830	3713	3370	3180	3297	3377	3233	3150

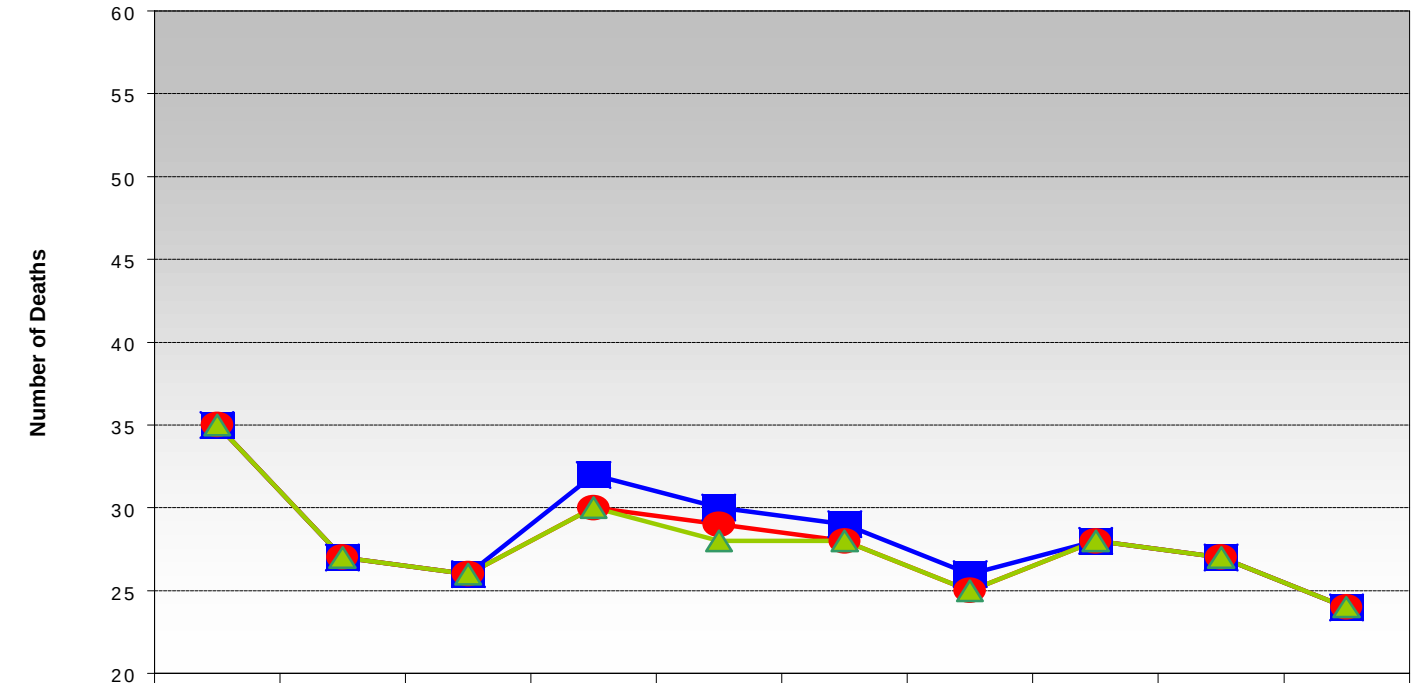
Year of Death

Chart 8. Deaths of Toronto residents due to ischemic heart disease.



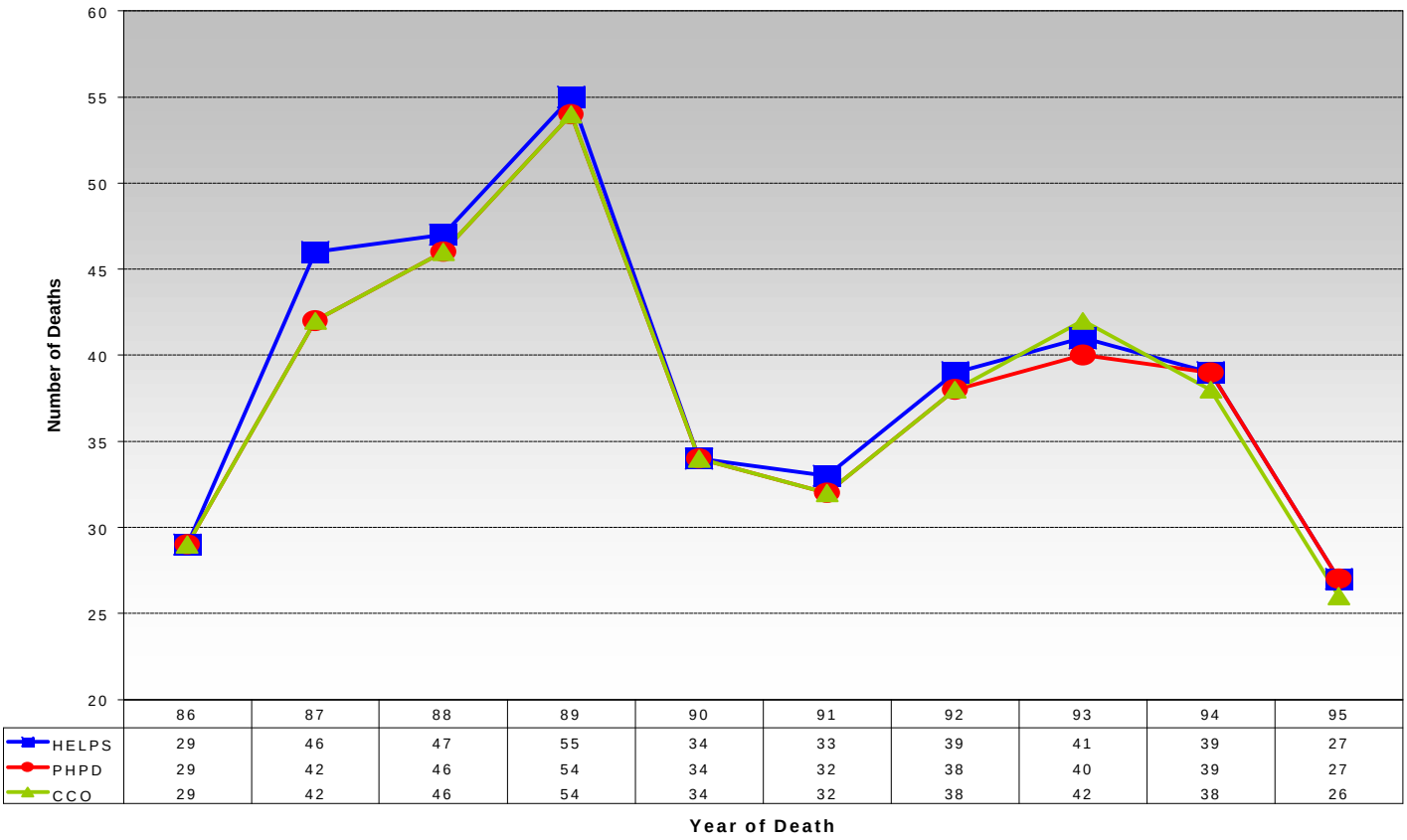
	86	87	88	89	90	91	92	93	94	95
HELPS	31	36	37	52	30	33	27	37	33	30
PHPD	31	36	35	50	30	33	25	37	33	30
CCO	31	36	35	50	34	34	25	38	33	30

Chart 9. Deaths of Durham Region residents as a result of motor vehicle collisions.



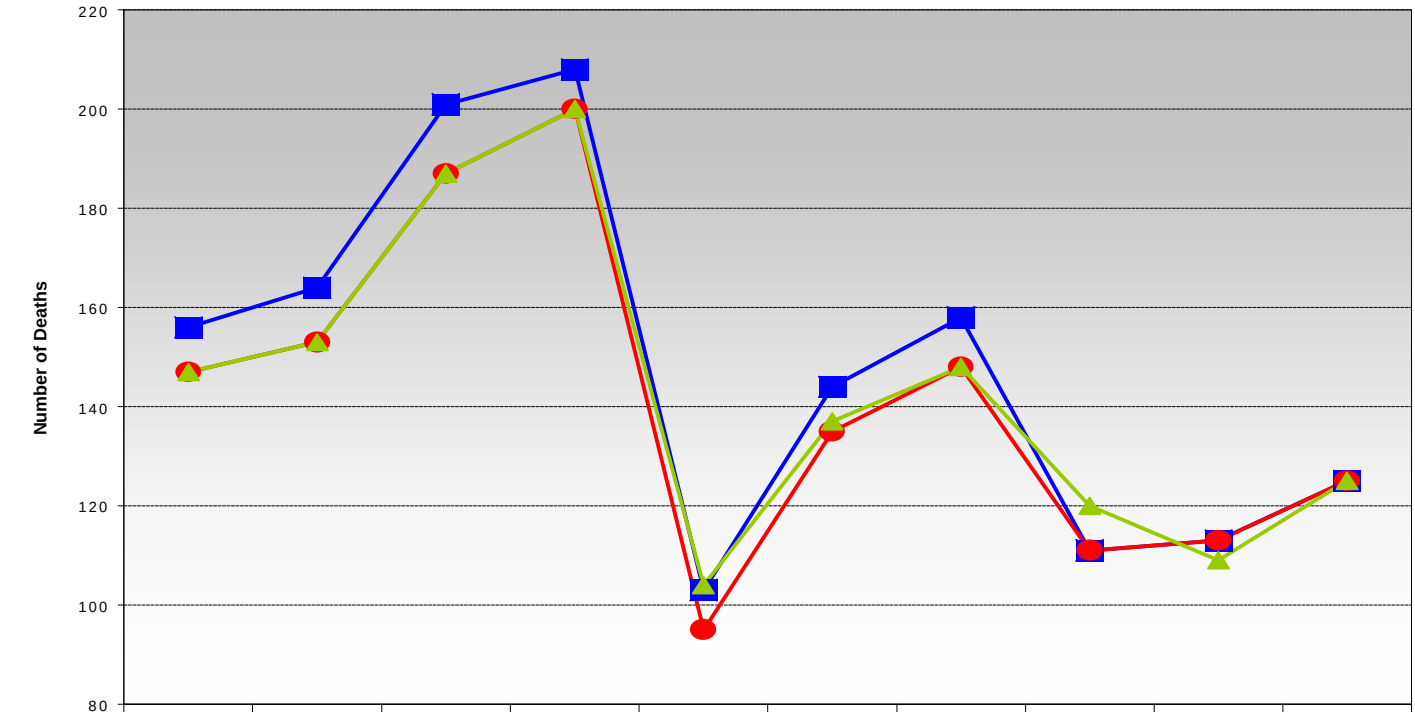
	86	87	88	89	90	91	92	93	94	95
HELPS	35	27	26	32	30	29	26	28	27	24
PHPD	35	27	26	30	29	28	25	28	27	24
CCO	35	27	26	30	28	28	25	28	27	24

Chart 10. Deaths of Halton Region residents as a result of motor vehicle collisions.



	86	87	88	89	90	91	92	93	94	95
HELPS	29	46	47	55	34	33	39	41	39	27
PHPD	29	42	46	54	34	32	38	40	39	27
CCO	29	42	46	54	34	32	38	42	38	26

Chart 11. Deaths of Simcoe County residents as a result of motor vehicle collisions.



	86	87	88	89	90	91	92	93	94	95
HELPS	156	164	201	208	103	144	158	111	113	125
PHPD	147	153	187	200	95	135	148	111	113	125
CCO	147	153	187	200	104	137	148	120	109	125

Chart 12. Deaths of Toronto residents as a result of motor vehicle collisions.

Table 1. Maximum and minimum difference among data sets by county/region and cause of death.

County/Region and Cause (Chart Number)	Number of Years of Complete Agreement Among Data Sets	Minimum Difference Among Data Sets	Maximum Difference Among Data Sets	
		%	%	Year
Breast Cancer				
Durham (1)	9	0	2.3	1988
Halton (2)	8	0	12.5	1990
Simcoe (3)	8	0	2.4	1991
Toronto (4)	1	0	0.7	1995
Ischemic Heart Disease				
Durham (5)	0	0.2	1.8	1993
Halton (6)	1	0	3.3	1990
Simcoe (7)	1	0	1.7	1988
Toronto (8)	0	0.4	1.9	1987
MVCs				
Durham (9)	4	0	11.8	1990
Halton (10)	6	0	6.7	1990
Simcoe (11)	2	0	8.7	1987
Toronto	1	0	8.7	1990

Discussion

Overall, very little difference was found when the results of the three data source were compared. Slight differences did exist in the total number of deaths for many of the years examined, but these differences, especially if converted to age-adjusted mortality rates, would not lead the agencies to arrive at different planning or research conclusions. Thus, the differences do not appear to be severe from a research or planning perspective. However, the fact that differences are present suggests that further investigation into the formatting and editing of this file by the different agencies is warranted.

It is evident from our results that none of the three sources have the exact same data set. None of the charts revealed complete agreement among the data sets over the ten years examined. Disagreement was greatest for the City of Toronto which may be a due to the fact that they also had

significantly more deaths from each cause than the other areas i.e. more events provides greater opportunity for an error. However, the number of inconsistencies among the data sets regarding MVC deaths was relatively large even though the absolute number of deaths was the smallest of the three causes. As well, disagreement was greatest for the two causes that included more than one three digit ICD-9 code.

Differences between the HELPS and PHPD results are almost entirely due to a single difference in the data editing procedures, specifically the inclusion/exclusion criteria for events occurring outside of the province of Ontario. The mortality data within the PHPD contains records for events pertaining to Ontario residents that occur within Ontario. The HELPS data from 1986 up to and including 1992 contains records for deaths of Ontario residents that occurred in Ontario as well as those that occurred outside of Ontario. HELPS data for 1993 to 1995 includes only deaths of Ontario residents that occurred in Ontario.

When we re-examined the HELPS data and selected only deaths occurring in Ontario, all differences between HELPS and PHPD for the years 1986 through 1992 disappeared. Slight differences between these two databases during 1993 through 1995 exist in two instances (see Charts 7 and 11). The origin of these two differences is unclear, but it is clear that this simple editing procedure has had a significant impact. Users of the PHPD who are unaware of the omission of events occurring outside of Ontario will arrive at different totals than HELPS users. Inclusion and exclusion criteria of the databases should be made apparent to all users of the data.

The remaining differences among the data sets, mainly CCO versus HELPS/PHPD, may have a number of different origins. Statistics Canada and CCO may not receive identical files from the Registrar General. Upon receipt of the mortality files, Statistics Canada and CCO may carry out different data quality checks and editing procedures. Further investigation involving the Registrar General's original morality data file is recommended to help determine the origin of these differences.

Conclusions

With the increasing emphasis on the use of population health data for local research and health planning purposes by health units and District Health Councils, it is important to assure that the data are of the highest possible quality. From this overview of mortality data, three conclusions can be drawn. First, there should be one set of criteria for inclusion/exclusion of cases and for coding/labeling that all users agree to and apply consistently. Second, there should be one original

vital statistics data file which all researchers and planners have access to. Third, there should be regular communication between the users of the vital statistics data and the Registrar General, the collector and owner of these data.

Collaboration between the owners and users of vital statistics data will result in the development of a more reliable data file and will enhance the value of the research and health planning projects that make use of the data.

References

McGurran, J.J. 1997. Review of Vital Statistics Data Files Provided to CEHIP for Analysis & Distribution. Central East Health Information Partnership.

McGurran, J.J., Woodward, G.L., Shortt, G.W., and Knight, A. 1998. Low Birth Weight in Central East Ontario, 1981-1995: A CEHIP Discussion Paper. Central East Health Information Partnership.

World Health Organization. 1977. International Classification of Diseases. 1975 Revision. Volume 1. World Health Organization, Geneva.