Health
Reports

Volume 20, Number 1

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Health Information and Research Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 613-951-1765).

For information about this product or the wide range of services and data available from Statistics Canada, visit our website at www.statcan.gc.ca, e-mail us at infostats@statcan.gc.ca, or telephone us, Monday to Friday from 8:30 a.m. to 4:30 p.m., at the following numbers:

Statistics Canada's National Contact Centre

Toll-free telephone (Canada and United States):

Inquiries line	1-800-263-1136
National telecommunications device for the hearing impaired	1-800-363-7629
Fax line	1-877-287-4369

Local or international calls:

Inquiries line	1-613-951-8116
Fax line	1-613-951-0581

Depository Services Program

Inquiries line	1-800-635-7943
Fax line	1-800-565-7757

To access and order this product

This product, Catalogue no. 82-003-X, is available free in electronic format. To obtain a single issue, visit our website at www.statcan.gc.ca and select "Publications."

This product, Catalogue no. 82-003-X, is also available as a standard printed publication at a price of CAN$24.00 per issue and CAN$68.00 for a one-year subscription.

The following additional shipping charges apply for delivery outside Canada:

	Single issue	Annual subscription
United States	CAN$6.00	CAN$24.00
Other countries	CAN$10.00	CAN$40.00

All prices exclude sales taxes.

The printed version of this publication can be ordered as follows:

- Telephone (Canada and United States) 1-800-267-6677
- Fax (Canada and United States) 1-877-287-4369
- E-mail infostats@statcan.gc.ca
- Mail Statistics Canada
  Finance
  R.H. Coats Bldg., 6th Floor
  150 Tunney's Pasture Driveway
  Ottawa, Ontario K1A 0T6
- In person from authorized agents and bookstores.

When notifying us of a change in your address, please provide both old and new addresses.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "Providing services to Canadians."

HealthReports

A Canadian peer-reviewed journal of population health and health services research

---

Note of Appreciation

Symbols

The following standard symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for specific reference period
...	not applicable
p	preliminary
r	revised
x	suppressed to meet the confidentiality requirements of the Statistics Act
E	use with caution
F	too unreliable to be published

About Health Reports

*H*ealth Reports publishes original research on diverse topics related to the health of populations and the delivery of health care. The journal archives, for the research and policy communities and for the general public, discoveries from analyses of national/provincial surveys and administrative databases, as well as results of international comparative health research. *Health Reports* is also a forum for sharing methodological information by those using health surveys or administrative databases. *Health Reports* is produced by the Health Information and Research Division at Statistics Canada. Articles appear monthly in electronic format and quarterly in print, and are indexed in Index Medicus and MEDLINE.

For more information about *Health Reports*, contact the Editor-in-Chief, Health Information and Research Division, Statistics Canada, 24th Floor, R.H. Coats Building, Ottawa, Ontario, Canada K1A 0T6. Telephone: (613) 951-1765; fax: (613) 951-3959; email: Christine.Wright@statcan.gc.ca

In this issue

Research articles

Health matters

Methodological insights

ELECTRONIC PUBLICATIONS AVAILABLE AT
www.statcan.ca

Cancer prevalence in the Canadian population

by Larry F. Ellison and Kathryn Wilkins

Abstract

Background
The rising numbers of cancer diagnoses, together with improvements in survival, have led to increases in the prevalence of cancer in Canada. This article provides more precise and detailed estimates of cancer prevalence than have been available previously.

Data and methods
Based on incidence data from the Canadian Cancer Registry linked with mortality data from the Canadian Vital Statistics Death Database, direct estimates of cancer prevalence as of January 1, 2005 were calculated for an extensive list of cancers, by time since diagnosis, age and sex.

Results
Two-, five- and ten-year cancer prevalence counts were 217,089 (675 per 100,000), 454,149 (1,412 per 100,000) and 722,833 (2,248 per 100,000), respectively. Breast (20.6% of ten-year prevalent cases), prostate (18.7%) and colorectal cancer (12.9%) were the most prevalent, together accounting for just over half of all cases. Prevalence proportions for all cancers combined increased dramatically with age, peaking at ages 80 to 84; proportions were higher in females than in males before age 60, and higher in males thereafter.

Interpretation
Prevalence data tabulated according to type of cancer, age and time since diagnoses provide important information about the demand for cancer-related health care and social services.

Keywords
epidemiologic methods, neoplasms, registries, surveillance

Authors
Larry F. Ellison (613-951-5244; larry.ellison@statcan.gc.ca) is with the Health Statistics Division and Kathryn Wilkins (613-951-1769; Kathryn.wilkins@statcan.gc.ca) is with the Health Information and Research Division at Statistics Canada, Ottawa, ON K1A 0T6.

Information on the prevalence of cancer in a population is important to health care planning. People diagnosed with cancer require treatment, monitoring for recurrence, and screening for other cancers. As well, they run the risk of permanent impairment or disability.[1] Cancer prevalence data provide an overall indication of the demand for cancer-related health care and social services and can be used to plan the future allocation of these resources.[1,2]

Cancer prevalence can be defined as the number of previously diagnosed cases of cancer in a given population, among people alive on a specified date (index date). Total prevalence refers to prevalent cases diagnosed at any previous time, while limited-duration prevalence refers to prevalent cases diagnosed within a specified number of years.

Categorizing cancer prevalence estimates according to time since diagnosis provides more precise indications of health care needs.[3] Cases diagnosed within ten years represent the major demand for cancer care services; more specifically, different health care services are required by cases that were diagnosed two or fewer, two to five, or five to ten years ago.[2,3] Cancer-specific prevalence data are also useful in planning resource allocation for treatment, care and follow-up because the management of each type of cancer is different.

Cancer prevalence is a function of both the incidence of and survival from the disease. In Canada, the number of newly diagnosed cases continues to rise,[4] and survival is also increasing.[5-8] The combined result is an increase in the number of people living with cancer, which leads to a growing demand for health care services.

Since 2003, the annual *Canadian Cancer Statistics* monograph has reported indirect estimates of 15-year prevalence for the leading cancers (female breast, prostate, colorectal and lung), and for all cancers combined.[9] The estimates for Canada were obtained by applying observed survival proportions derived from Saskatchewan

Cancer Registry data—most recently from cases diagnosed from 1986 to 2001, with follow-up to the end of 2002[10]—to national cancer incidence counts. With the maturity of the Canadian Cancer Registry (CCR), which contains information on cases diagnosed from 1992 onward, it is now possible to derive Canadian estimates of prevalence directly, using individual information on vital status. In addition, the demand for prevalence figures by time since diagnosis, age and for more than the leading cancers can be met.

Using data from the CCR, this report provides Canadian cancer prevalence estimates as of January 1, 2005. National estimates were directly derived for all cases except those diagnosed in the province of Quebec, for which it was necessary to use indirect methodology. Results were calculated by cancer, age group, sex and duration since diagnosis (that is, two-, five-, and ten-year prevalence).

Methods

Data source

Cancer incidence data are from the January 2008 version of the Canadian Cancer Registry, a dynamic, person-oriented, population-based database maintained by Statistics Canada. The CCR contains information on cases diagnosed from 1992 onward, compiled from reports from every provincial/ territorial cancer registry. A detailed description of the CCR, including data sources, methodology and accuracy, is available on Statistics Canada's website.[11] Mortality data are from the Canadian Vital Statistics Death Database, also maintained by Statistics Canada. These data are based on information provided by the vital statistics registrars in each province and territory. Population estimates are from Statistics Canada's *Demographics Estimates Compendium 2007*.[12]

Analytical techniques

A file containing records of invasive cancer cases and *in situ* bladder cancer cases (the latter are reported for each province/territory except Ontario) was created using the multiple primary coding rules of the International Agency for Research on Cancer.[13] Cancer cases were classified based on the *International Classification of Diseases for Oncology, Third Edition* [14] and grouped using Surveillance, Epidemiology, and End Results (SEER) Program grouping definitions, with mesothelioma and Kaposi's sarcoma as separate groups.[15]

Mortality follow-up—complete through December 31, 2004—was determined through record linkage to the Canadian Vital Statistics Death Database, and from information reported by provincial/territorial cancer registries. For deaths reported by a provincial registry but not confirmed by record linkage, the date of death was assumed to be that submitted by the reporting registry. When the date of death was completely missing (n=77), the death was assumed to have occurred after the index date (January 1, 2005).

Prevalence can be calculated so as to estimate the number of people living with cancer on a specified date (person-based prevalence) or to estimate the total number of diagnoses of cancer among those alive on that date (tumour-based prevalence). The second method includes all qualifying cancers in the time-frame under consideration, regardless of whether they were first or subsequent primaries. Tumour-based prevalence is more useful in reflecting the demand for health care, because multiple cancers in an individual are usually treated independently.[1] Therefore, this report focuses on tumour-based prevalence. For completeness, however, estimates of person-based prevalence are provided in Appendix Tables A and B.

Prevalence was determined directly, using the counting method.[16,17] All primary invasive cancers (including *in situ* bladder cancer cases) among

persons alive on January 1, 2005 that had been diagnosed in the time-frame under consideration were counted. Two-year prevalence was estimated by counting the number of invasive primary cancers diagnosed from January 1, 2003 to December 31, 2004 in persons who were still alive on January 1, 2005. Similarly, five- and ten-year prevalence estimates were based on cases diagnosed back to 2000 and 1995, respectively.

Because of issues involved in ascertaining the vital status of cases diagnosed in Quebec, prevalence data for this province were determined indirectly. The probability of surviving until the index date was used to randomly assign the vital status of each incident case in Quebec. Survival probabilities were derived using the corresponding observed survival proportion calculated for the rest of Canada, stratified on age group (0 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, and 80 or older), sex, cancer and month of diagnosis. Monthly survival estimates were calculated through ten separate analyses—one for each year of follow-up—using the life-table (actuarial) approach. Each analysis was based on cases diagnosed over a four-year span covering the most pertinent year (for example, 1995 for those months in the tenth year of survival) and the three preceding years. For cancers for which there were sufficient data—colon, rectum, lung and bronchus (lung), skin melanoma, female breast, corpus uteri, prostate, bladder and other—the eldest age group was sub-divided: 80 to 84, 85 to 89, and 90 or older.

Age-specific prevalence estimates were derived using the age attained by each case as of January 1, 2005. Year of birth was missing for 105 cases. Because the exclusion of these cases from the analysis would have led to an underestimation of prevalence, the attained age group was randomly imputed using the sex-specific attained age-group distribution of prevalent cases in Ontario (where virtually all of the 105 affected cases had been

diagnosed). Because of the relatively large percentage of prostate cancer cases (25%) among the cases with unknown age, and the uniqueness of the age distribution for this cancer, two attained age-group distributions for imputation were used for males: one for prostate cases, and one for all other cancers combined. Crude prevalence proportions (per 100,000) were calculated by dividing prevalence counts by the appropriate population on January 1, 2005 and multiplying by 100,000. Population estimates for this date were derived by averaging the 2004 and 2005 mid-year population estimates.

Trends in prevalence proportions across age groups were determined using the Average Annual Percent Change (AAPC) feature of the Joinpoint Regression Program (v 3.3) distributed by the SEER program of the National Cancer Institute in the United States.[18] Tests of statistical significance were conducted with alpha=0.05.

Results

Among persons alive in Canada on January 1, 2005, an estimated 722,833 primary invasive cancer cases (or 2,248 per 100,000 persons) had been diagnosed from 1995 through 2004 (Tables 1 and 2). (Appendix Table C contains age-standardized results.) Five- and two-year prevalence case counts were 454,149 (or 1,412 per 100,000 persons) and 217,089 (or 675 per 100,000 persons), respectively. For all cancers and ages combined, prevalent cases were nearly evenly distributed between the sexes for each duration period; the percentage of cases in males ranged from 48.9% (ten-year duration) to 50.9% (two-year duration).

Nearly 40% of ten-year prevalent cancer cases were either breast (20.5%) or prostate (18.7%) (Table 1, Figure 1). Colorectal cancer was the next most common (12.9%), followed by lung cancer (5.1%), bladder cancer (5.0%), non-Hodgkin lymphoma (4.1%) and skin melanoma (4.1%). The relative contribution of both breast and prostate cancer decreased when shorter prevalence periods were considered; each comprised about 17% of the two-year prevalent cases. The opposite pattern was observed for lung cancer: the relative contribution increased with shorter prevalence periods (6.1% for five-year; 8.1% for two-year).

In men, prostate cancer accounted for the largest share of ten-year prevalent cases (38.2%), followed by

Table 1
Number of prevalent cases, by prevalence-duration, cancer and sex, Canada, January 1, 2005

Cancer	Prevalence-duration								
	Two-year			Five-year			Ten-year		
	Both sexes	Males	Females	Both sexes	Males	Females	Both sexes	Males	Females
All cancers	**217,089**	**110,595**	**106,494**	**454,149**	**228,318**	**225,831**	**722,833**	**353,508**	**369,325**
Oral cavity and pharynx	5,198	3,524	1,674	10,819	7,235	3,584	17,383	11,648	5,735
Esophagus	1,215	900	315	1,767	1,319	448	2,266	1,644	622
Stomach	2,890	1,911	979	4,791	3,056	1,735	6,875	4,229	2,646
Colorectal	29,554	16,048	13,506	60,700	32,736	27,964	93,489	49,603	43,886
Colon excluding rectum	19,438	9,838	9,600	40,005	20,070	19,935	62,123	30,781	31,342
Rectum and rectosigmoid	10,116	6,210	3,906	20,695	12,666	8,029	31,366	18,822	12,544
Anus	864	380	484	1,742	734	1,008	2,671	1,108	1,563
Liver	942	712	230	1,553	1,178	375	1,967	1,473	494
Pancreas	1,825	883	942	2,444	1,197	1,247	3,037	1,464	1,573
Larynx	1,861	1,521	340	3,908	3,215	693	6,701	5,479	1,222
Lung and bronchus	17,620	9,189	8,431	27,642	14,122	13,520	37,168	19,052	18,116
Soft tissue	1,442	793	649	3,001	1,645	1,356	4,891	2,623	2,268
Skin melanoma	7,722	4,078	3,644	17,459	8,961	8,498	29,602	14,697	14,905
Breast	37,391	306	37,085	86,552	598	85,954	148,542	947	147,595
Cervix uteri	2,346	...	2,346	5,433	...	5,433	9,995	...	9,995
Corpus uteri	6,894	...	6,894	15,373	...	15,373	26,467	...	26,467
Ovary	3,282	...	3,282	6,265	...	6,265	9,355	...	9,355
Prostate	37,583	37,583	...	85,956	85,956	...	135,065	135,065	...
Testis	1,557	1,557	...	3,817	3,817	...	7,091	7,091	...
Bladder (including in situ)	10,532	7,946	2,586	21,887	16,448	5,439	35,807	26,625	9,182
Kidney and renal pelvis	6,045	3,624	2,421	12,838	7,670	5,168	20,553	12,205	8,348
Brain	2,121	1,229	892	3,841	2,175	1,666	6,015	3,355	2,660
Thyroid	6,001	1,220	4,781	13,091	2,804	10,287	20,529	4,430	16,099
Hodgkin lymphoma	1,618	883	735	3,751	2,079	1,672	6,906	3,806	3,100
Non-Hodgkin lymphoma	9,253	4,930	4,323	18,991	10,015	8,976	29,619	15,316	14,303
Multiple myeloma	2,537	1,362	1,175	4,564	2,428	2,136	5,902	3,126	2,776
Leukemias	5,794	3,426	2,368	11,511	6,720	4,791	17,684	10,170	7,514
Other, unknown	13,002	6,590	6,412	24,453	12,210	12,243	37,253	18,352	18,901

... not applicable
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

Table 2
Prevalence proportions (per 100,000), by prevalence-duration, cancer and sex, Canada, January 1, 2005

Cancer	Two-year			Five-year			Ten-year		
	Both sexes	**Males**	**Females**	**Both sexes**	**Males**	**Females**	**Both sexes**	**Males**	**Females**
All cancers	**675.2**	**694.5**	**656.2**	**1,412.4**	**1,433.7**	**1,391.6**	**2,248.1**	**2,219.8**	**2,275.8**
Oral cavity and pharynx	16.2	22.1	10.3	33.6	45.4	22.1	54.1	73.1	35.3
Esophagus	3.8	5.7	1.9	5.5	8.3	2.8	7.0	10.3	3.8
Stomach	9.0	12.0	6.0	14.9	19.2	10.7	21.4	26.6	16.3
Colorectal	91.9	100.8	83.2	188.8	205.6	172.3	290.8	311.5	270.4
Colon excluding rectum	60.5	61.8	59.2	124.4	126.0	122.8	193.2	193.3	193.1
Rectum and rectosigmoid	31.5	39.0	24.1	64.4	79.5	49.5	97.6	118.2	77.3
Anus	2.7	2.4	3.0	5.4	4.6	6.2	8.3	7.0	9.6
Liver	2.9	4.5	1.4	4.8	7.4	2.3	6.1	9.2	3.0
Pancreas	5.7	5.5	5.8	7.6	7.5	7.7	9.4	9.2	9.7
Larynx	5.8	9.6	2.1	12.2	20.2	4.3	20.8	34.4	7.5
Lung and bronchus	54.8	57.7	52.0	86.0	88.7	83.3	115.6	119.6	111.6
Soft tissue	4.5	5.0	4.0	9.3	10.3	8.4	15.2	16.5	14.0
Skin melanoma	24.0	25.6	22.5	54.3	56.3	52.4	92.1	92.3	91.8
Breast	116.3	1.9	228.5	269.2	3.8	529.7	462.0	5.9	909.5
Cervix uteri	...	...	14.5	...	...	33.5	...	...	61.6
Corpus uteri	...	...	42.5	...	...	94.7	...	...	163.1
Ovary	...	...	20.2	...	...	38.6	...	...	57.6
Prostate	...	236.0	...	...	539.7	...	...	848.1	...
Testis	...	9.8	...	...	24.0	...	...	44.5	...
Bladder (including in situ)	32.8	49.9	15.9	68.1	103.3	33.5	111.4	167.2	56.6
Kidney and renal pelvis	18.8	22.8	14.9	39.9	48.2	31.8	63.9	76.6	51.4
Brain	6.6	7.7	5.5	11.9	13.7	10.3	18.7	21.1	16.4
Thyroid	18.7	7.7	29.5	40.7	17.6	63.4	63.8	27.8	99.2
Hodgkin lymphoma	5.0	5.5	4.5	11.7	13.1	10.3	21.5	23.9	19.1
Non-Hodgkin lymphoma	28.8	31.0	26.6	59.1	62.9	55.3	92.1	96.2	88.1
Multiple myeloma	7.9	8.6	7.2	14.2	15.2	13.2	18.4	19.6	17.1
Leukemias	18.0	21.5	14.6	35.8	42.2	29.5	55.0	63.9	46.3
Other, unknown	40.4	41.4	39.5	76.1	76.7	75.4	115.9	115.2	116.5

... not applicable
Note: Prevalence proportions were determined by a tumour-based analysis.
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

colorectal (14.0%), bladder (7.5%) and lung cancer (5.4%) (Figure 2). The corresponding most prevalent cancers in women were breast (40.0%), colorectal (11.9%), corpus uteri (7.2%) and lung (4.9%).

The increase with age in the prevalence proportion of cancer is striking. However, the patterns of increase differed between the sexes (Figure 3). In females, five-year prevalence exceeded that in males until just under age 60. Thereafter, mostly because of a sharp rise in prostate cancer, the proportions crossed over, and prevalence increased much more rapidly in males than in females. Prevalence proportions (per 100,000) peaked in the 80-to-84-year age group in both males (9,170) and females (5,179), and at older ages dipped to approximately the level at ages 70 to

74. The pattern was similar for two- and ten-year prevalence proportions (data not shown).

A statistically significant increase with age in sex-specific five-year prevalence proportions was observed for all cancers studied except testicular cancer among males, cervical and thyroid cancer among females, and Hodgkin lymphoma and brain cancer among both sexes (Table 3). Monotonic increases were observed across all age groups for some cancers, while for some others, the prevalence proportion rose with age and then fell in the oldest age group. For several cancers exhibiting the latter pattern, the proportion was at least 15% lower among persons aged 80 or older, compared with those aged 70 to 79. This was the case for thyroid, liver, and laryngeal cancer among males,

and ovarian, lung, corpus uterine and kidney and renal pelvis cancer among females. For testicular cancer, and for Hodgkin lymphoma among both sexes, five-year prevalence proportions were highest in young adults (aged 20 to 39), and the decrease with age was statistically significant. For cancer of the cervix uteri and thyroid cancer among females, prevalence peaked at ages 40 to 49 and then decreased monotonically at older ages.

Although the number of prevalent cancers varied greatly by age group, similarities emerged in the types of cancers that were most common. In the three oldest age groups, the most common cancers in terms of ten-year prevalence (prostate, breast, colorectal, lung, and bladder) were the same, and their ranking relative to one another was quite similar (Figure 4). Breast,

Statistics Canada, Catalogue no. 82-003-XPE • Health Reports, Vol. 20, no.1, March 2009
Cancer prevalence in the Canadian population • Research Article

11

Figure 1
Number of prevalent cases of ten leading cancers, by prevalence-duration, Canada, January 1, 2005
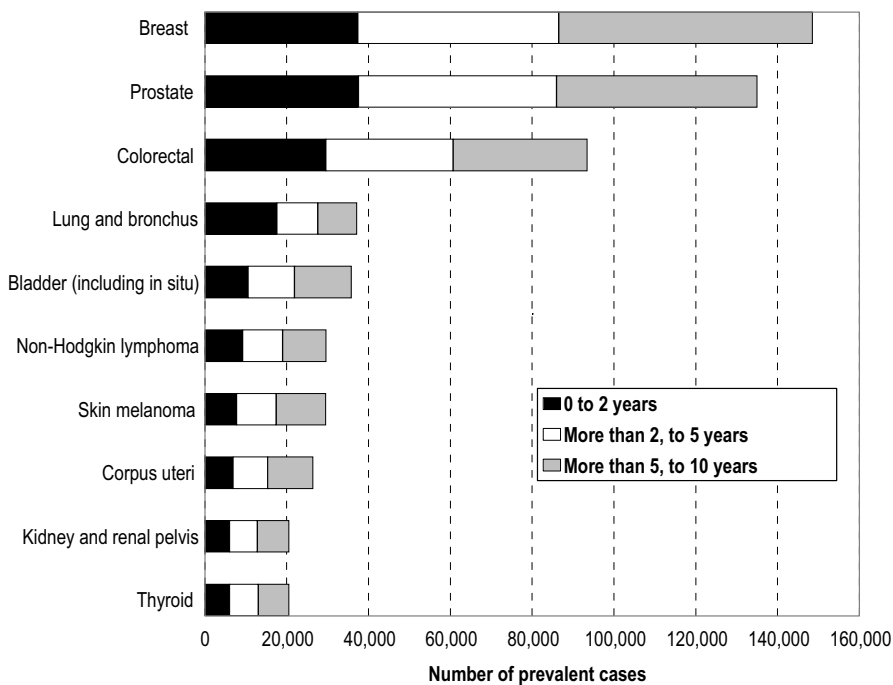
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

prostate and colorectal cancer were also the most prevalent cancers among 50- to 59-year-olds; in this age group, however, the number of breast cancers was triple that of either of the other two. Breast cancer was similarly dominant in the 40 to 49 age group, ahead of thyroid cancer and skin melanoma. From age 20 to 39, thyroid cancer was the most prevalent, followed by testicular cancer, Hodgkin lymphoma, breast cancer and skin melanoma.

Discussion

This study provides two-, five-, and ten-year Canadian prevalence estimates by sex for an extensive list of cancers. From 1995 to 2004, just under 723,000 primary cancers were diagnosed in Canada among approximately 695,000 people who survived until at least the end of 2004. Breast, prostate and colorectal cancer were the most prevalent, accounting for just over half of all ten-year cases. The sex-specific prevalence proportions for all cancers combined rose dramatically with age—

Figure 2
Distribution of ten-year prevalent cancer cases, by sex and cancer, January 1, 2005
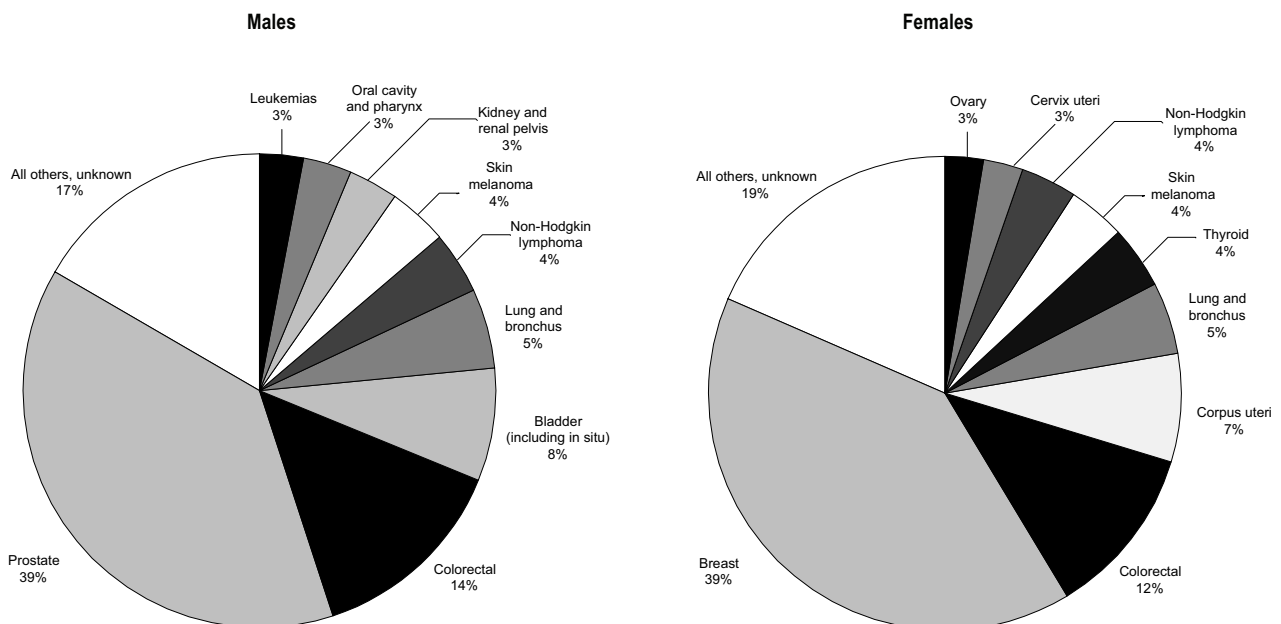
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

Table 3
Age-specific, five-year prevalence proportions (per 100,000) for all cancers combined, by sex, Canada, January 1, 2005

Cancer	Age group						
	All ages	20 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 or older
Males							
All cancers	**1,433.7**	**183.8**	**486.0**	**1,656.30**	**4,898.2**	**8,287.8**	**8,945.3**
Oral cavity and pharynx[†]	45.4	5.4	34.2	88.6	147.3	180.1	186.6
Esophagus[†]	8.3	0.3	2.6	11.4	32.6	47.1	42.5
Stomach[†]	19.2	1.1	7.7	23.6	60.5	113.9	137.9
Colorectal[†]	205.6	8.3	55.5	236.0	679.3	1,268.0	1,589.9
Colon excluding rectum[†]	126.0	5.1	30.9	129.3	397.0	801.0	1,091.8
Rectum and rectosigmoid[†]	79.5	3.2	24.6	106.7	282.2	467.0	498.2
Anus[†]	4.6	0.5	3.3	7.4	16.0	20.0	23.1
Liver[†]	7.4	0.7	3.7	13.8	25.5	33.7	24.6
Pancreas[†]	7.5	0.6	4.2	12.1	27.9	33.6	39.4
Larynx[†]	20.2	0.4	5.9	28.7	83.1	114.8	94.6
Lung and bronchus[†]	88.7	2.0	19.3	97.4	324.9	596.1	543.3
Soft tissue[†]	10.3	5.0	9.3	13.6	20.4	37.0	45.9
Skin melanoma[†]	56.3	17.5	47.7	94.4	145.4	225.8	269.3
Prostate[†]	539.7	0.3	31.6	500.1	2,222.2	3,727.0	3,512.4
Testis[‡]	24.0	48.3	38.0	15.2	8.3	4.3	6.7
Bladder (including in situ[†]	103.3	3.2	22.2	99.9	311.8	673.0	970.4
Kidney and renal pelvis[†]	48.2	4.8	31.0	84.1	158.9	219.3	209.4
Brain	13.7	11.1	16.3	17.6	20.6	18.2	15.0
Thyroid[†]	17.6	12.6	23.7	32.5	35.5	37.6	25.1
Hodgkin lymphoma[‡]	13.1	19.6	14.9	11.4	13.9	14.1	9.3
Non-Hodgkin lymphoma[†]	62.9	15.6	46.9	98.9	183.9	265.0	272.4
Multiple myeloma[†]	15.2	0.7	6.4	21.8	51.2	81.3	103.9
Leukemias[†]	42.2	10.2	20.7	51.3	111.9	176.1	243.1
Other, unknown[†]	76.7	15.5	39.4	91.9	205.4	380.9	547.7
Females							
All cancers	**1,391.6**	**293.8**	**1,067.8**	**2,229.0**	**3,643.7**	**4,742.7**	**4,934.9**
Oral cavity and pharynx[†]	22.1	4.8	17.3	33.5	53.6	78.0	85.9
Stomach[†]	10.7	1.0	4.9	11.4	24.2	47.5	68.3
Colorectal[†]	172.3	7.8	52.8	176.2	424.9	812.6	1,143.3
Colon excluding rectum[†]	122.8	5.2	32.8	114.9	290.0	592.4	873.4
Rectum and rectosigmoid[†]	49.5	2.6	20.1	61.2	134.9	220.2	269.9
Anus[†]	6.2	0.3	5.4	11.1	15.5	20.1	26.7
Pancreas[†]	7.7	0.7	3.6	9.4	21.9	33.0	37.7
Lung and bronchus[†]	83.3	2.6	35.3	116.6	282.9	398.9	270.7
Soft tissue[†]	8.4	4.8	7.8	10.0	15.5	19.8	26.8
Skin melanoma[†]	52.4	29.3	61.7	86.0	96.5	127.8	140.7
Breast[†]	529.7	59.3	477.9	1,033.4	1,473.9	1,638.6	1,568.0
Cervix uteri	33.5	32.3	61.2	49.9	45.6	40.4	34.5
Corpus uteri[†]	94.7	5.7	42.3	183.6	344.6	336.0	258.0
Ovary[†]	38.6	10.5	36.7	77.2	105.3	109.3	72.8
Bladder (including in situ)[†]	33.5	1.7	8.6	34.9	86.9	159.1	215.0
Kidney and renal pelvis[†]	31.8	4.5	20.6	46.7	87.0	127.6	108.0
Brain	10.3	9.1	11.6	12.5	14.7	12.3	6.2
Thyroid	63.4	64.2	105.6	103.8	93.3	68.0	39.7
Hodgkin lymphoma[‡]	10.3	18.4	8.7	8.2	7.4	10.9	7.2
Non-Hodgkin lymphoma[†]	55.3	12.4	35.4	76.3	152.4	211.2	206.4
Multiple myeloma[†]	13.2	0.5	4.5	15.7	36.1	62.8	71.5
Leukemias[†]	29.5	7.4	15.8	29.0	58.4	105.5	136.5
Other, unknown[†]	75.4	16.0	47.4	91.2	174.2	279.1	373.1

[†] statistically significant increasing trend in prevalence proportions across age groups
[‡] statistically significant decreasing trend in prevalence proportions across age groups
Note: Prevalence proportions were determined by a tumour-based analysis.
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

peaking at ages 80 to 84; proportions were higher in females than males before age 60 and higher for males thereafter.

For the first time, national figures are provided by age group and by prevalence-duration. The time elapsed since diagnosis provides a useful indicator of treatment need or follow-up services utilization. For example, just over 217,000 cancers had been diagnosed in 2003/2004 in persons who were still alive on January 1, 2005, and therefore, likely to be at a stage

What is already known on this subject?

■ Because of increases in the incidence of cancer and improving survival, the number of Canadians living with cancer is rising. This leads to a growing demand for cancer care services.

■ Cancer prevalence data provide an overall indication of the demand for cancer-related health care and social services in a population and can be used to plan the future allocation of these resources.

■ Prevalence estimates by time since diagnosis are especially relevant to resource planning; for example, estimates of five-year prevalence reflect fairly closely the number of cancers requiring active treatment or close follow-up care.

What does this study add?

■ This study presents direct estimates of the prevalence of cancer in Canada. Estimates derived directly are more precise than those based on indirect methodology.

■ For the first time, national figures are provided by age group and time since diagnosis.

■ In addition to the four leading cancers, for which indirectly derived estimates have been reported previously, this analysis offers data for 22 other cancers.

Figure 3

Age-specific, five-year prevalence proportions for all cancers combined, by sex, Canada, January 1, 2005
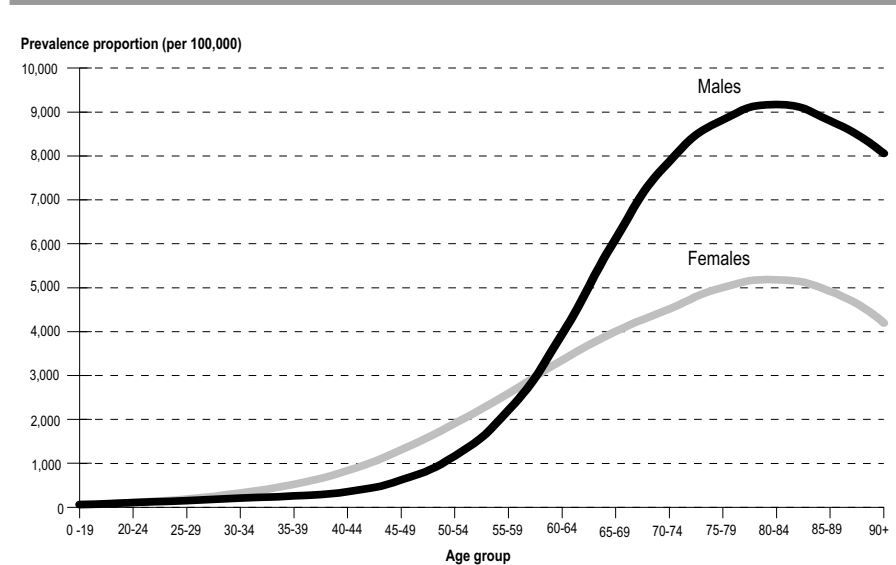
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

of the disease when they were undergoing primary treatment or recovering from its effects. In the period from two to five years since diagnosis—a time requiring close clinical follow-up for recurrence—the corresponding figure was just over 237,000. By age group, thyroid (ages 20 to 39), breast (ages 40 to 49 and 50 to 59), and prostate (ages 60 to 69, 70 to 79, and 80 or older) were the leading cancers.

Breast and prostate cancer were the most prevalent cancers in Canada—partly owing to their relatively high incidence,[4] but also because of favourable survival.[5,7] Despite the higher incidence of lung cancer during the period of study,[4] the number of prevalent colorectal cancer cases (ten-year) was over 2.5 times greater, reflecting the poor prognosis for those diagnosed with lung cancer.[5,7]

Sex-specific differences in the prevalence proportions for all cancers combined, before and after age 60, are also attributable to differences in incidence and survival. From age 25 to 54, incidence rates were considerably higher in females than in males, largely due to breast cancer. After age 60,

as a result of a dramatic increase in prostate cancer rates, overall incidence rates were higher in males.[4] Also, in persons younger than 65, overall survival from cancer was higher for females than for males.[5]

After age-adjustment to the European standard population, current Canadian estimates of the five-year prevalence proportion for all cancers combined were approximately 9% higher for both sexes than those recently reported for France for 2002. The French estimates were derived indirectly and based on cancer registry data covering 15% of the country.[19] Model-based estimates of five-year crude prevalence proportions in 2005 were also recently reported for Italians younger than 85.[3] Again, the overall estimates for Canada were higher—by approximately 11% among males and 4% among females. Compared with both the Italian and French estimates, Canadian prevalence proportions were higher for prostate and female lung cancer, but lower for breast and male lung cancer. For colorectal cancer, Canadian estimates were higher than those for France but lower than those for Italy. Comparisons of Canadian estimates with those for

Figure 4
Number of prevalent cases of leading cancers, by age group and prevalence-duration, Canada, January 1, 2005
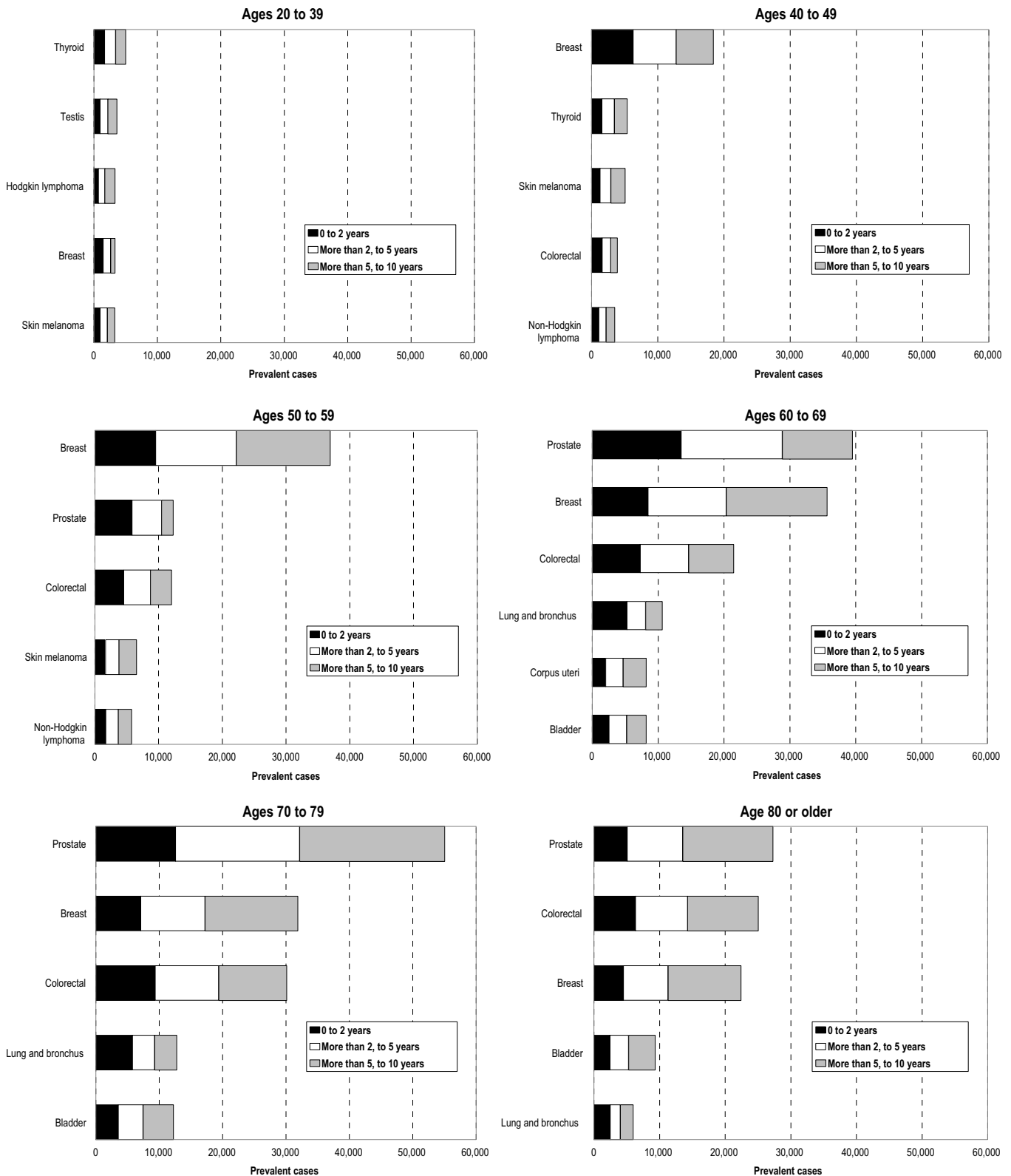
the United States could not be undertaken because prevalence proportions for the latter are not reported for durations of less than 15 years.[15]

Limitations

Except for cases diagnosed in the province of Quebec, prevalence estimates were determined directly by using individual information on vital status—resulting in greater precision of results than if they had been determined indirectly. The extent to which indirect estimates of prevalence for Quebec reflect direct prevalence depends on the degree of similarity in cancer survival between Quebec and the rest of Canada—which may vary. Quebec contributes about one-quarter of Canada's incident cancer cases.

Cancer prevalence will be underestimated if the registration of new cases is incomplete. In Canada, case registration by the provincial/ territorial cancer registries is generally considered to be quite complete.[20] In Quebec, however, because of the registry's dependence on hospital data, the numbers of microscopically confirmed prostate, melanoma and bladder cases have been estimated to be underreported by 32%, 35% and 14%, respectively.[21] In Ontario, the prevalence of bladder cancer was underestimated because *in situ* bladder cancer cases were not collected.

Persons whose cancer was documented in the CCR, but who moved out of the country and died before the index date, may have been erroneously considered as prevalent cases. However, they likely account for a minute proportion of the total number of prevalent cases, and their number may have been approximately offset by immigration of people with cancer into the country.

The possibility that some persons counted as prevalent cancer cases may have been cured was not considered in this study. To estimate the number of prevalent cases that have not been cured, statistical approaches have been applied to model "cure prevalence,"[22,23] but such analyses were beyond the scope of this study. Nonetheless, even among people who have been cured, cancer treatment can lead to long-term or permanent physical and psychological after-effects.

Conclusion

The current study presents the most precise and specific estimates of cancer prevalence in Canada that have been reported to date. The breakdown of prevalence estimates for an extensive list of cancers by time since diagnosis, age and sex provides much more detailed information about the cancer care needs of specific sub-populations than has been previously available. A valuable follow-up would be an examination of temporal trends in prevalence, leading to short- and long-term projections of these figures.■

Acknowledgement

References

1. Capocaccia R, Colonna M, Corazziari I, et al. Measuring cancer prevalence in Europe: the EUROPREVAL project. *Annals of Oncology* 2002; 13: 831-9.

2. Micheli A, Mugno E, Krogh V, et al. Cancer prevalence in European registry areas. *Annals of Oncology* 2002; 13: 840-65.

3. De Angelis R, Grande E, Inghelmann R et al. Cancer prevalence estimates in Italy from 1970 to 2010. *Tumori* 2007; 93: 392-7.

4. Statistics Canada. *Cancer Incidence in Canada* (Catalogue 82-231-XIE-2008001) Ottawa: Minister of Industry, 2008. Available at: http://www.statcan.gc.ca/pub/82-231-x/82-231-x2008001-eng.htm.

5. Statistics Canada. *Cancer Survival Statistics* (Catalogue82-226-XIE-2008001. Ottawa: Minister of Industry, 2008. Available at: http://www.statcan.gc.ca/pub/82-226-x/82-226-x2008001-eng.htm.

6. Ellison LF, Gibbons L. Leading cancers—changes in five-year relative survival. *Health Reports* (Statistics Canada, Catalogue 82-203) 2004; 15(2): 21-33.

7. Ellison LF, Gibbons L. Survival from cancer—up-to-date predictions using period analysis. *Health Reports* (Statistics Canada, Catalogue 82-003) 2006; 17(2): 19-30.

8. Ellison LF, Pogany L, Mery LS. Childhood and adolescent cancer survival: A period analysis of data from the Canadian Cancer Registry. *European Journal of Cancer* 2007; 43: 1967-75.

9. National Cancer Institute of Canada. *Canadian Cancer Statistics 2003.* Toronto: National Cancer Institute of Canada, 2003.

10. Canadian Cancer Society and National Cancer Institute of Canada. *Canadian Cancer Statistics 2008.* Toronto: Canadian Cancer Society and National Cancer Institute of Canada, 2008.

11. Statistics Canada. *Canadian Cancer Registry*. Ottawa: Health Statistics Division, Statistics Canada. Available at http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=3207&land=en&db=imdb&adm=8&dis=2. Accessed January, 2009.

12. Statistics Canada. *Demographic Estimates Compendium 2007* (Catalogue 91-213-SCB) Ottawa: Minister of Industry, 2008.

13. Parkin DM, Chen VW, Ferlay J, et al. *Comparability and Quality Control in Cancer Registration*. (IARC Technical Publications No. 19) Lyon: International Agency for Research on Cancer, 1994.

14. Fritz A, Percy C, Jack A, et al., eds. *International Classification of Diseases for Oncology, Third Edition*. Geneva: World Health Organization, 2000.

15. Ries LAG, Melbert D, Krapcho M et al., eds. *SEER Cancer Statistics Review, 1975-2005,*. Bethesda, Maryland: National Cancer Institute, 2008. Available at: http://seer.cancer.gov/csr/1975_2005/.

16. Feldman AR, Kessler L, Myers MH et al. The prevalence of cancer. Estimates based upon the Connecticut tumor registry. *The New England Journal of Medicine* 1986; 315: 1394-7.

17. Gail MH, Kessler L, Midthune D, et al. Two approaches for estimating disease prevalence from population-based registries of incidence and mortality. *Biometrics* 1999; 55: 1137-44.

18. Statistical Research and Applications Branch, National Cancer Institute. *Joinpoint Regression Program, Version 3.3 – April 2008*. Place of publication: Statistical Research and Applications Branch, National Cancer Institute, 2008.

19. Colonna M, Danzon A, Delafosse P et al. Cancer prevalence in France: Time trend, situation in 2002 and extrapolation to 2012. *European Journal of Cancer* 2008; 44: 115-22.

20. Howe HL, Lake A, Firth R, et al., eds. *Cancer in North America, 2001-2005. Volume Two: Registry-specific Cancer Incidence in the United States and Canada.* Springfield, Illinois: North American Association of Central Cancer Registries, Inc., 2008.

21. Brisson J, Major D, Pelletier E. *Évaluation de l'exhaustivité du fichier des tumeurs du Québec*. Quebec: Institut national de la santé publique du Québec, 2003.

22. Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registries data. *Statistics in Medicine* 1997; 16: 425-40.

23. Coldman AJ, McBride ML, Braun T. Calculating the prevalence of cancer. *Statistics in Medicine* 1992; 11: 1579-89.

24. Curado MP, Edwards B, Shin HR, et al., eds. *Cancer Incidence in Five Continents, Vol. IX*. (IARC Scientific Publications No. 160) Lyon: International Agency for Research on Cancer, 2007.

25. Segi, M. *Cancer mortality for selected sites in 24 countries (1950-1957)*. Sendai: Department of Public Health, Tohoku University of Medicine, 1960.

Appendix

Sometimes the unit of measure of interest is the number of people living with a past diagnosis of cancer rather than the number of cancers. In terms of individuals, 695,049 persons had been diagnosed with one or more primary invasive cancers in the previous ten years and were alive on January 1, 2005 (Appendix Table B). Of these, 441,155 (63%) had been diagnosed in the previous five years, and 212,606 (31%) in the previous two years (Appendix Table A). These figures correspond to 2.2% of the estimated population on January 1, 2005, or approximately 1 in 46 persons (ten-year); 1.4%, or 1 in 73 persons (five-year), and 0.7%, or 1 in 151 (two-year) (Appendix Table B). One in every 111 females alive on January 1, 2005 had been diagnosed with breast cancer in the previous 10 years; the figure for prostate cancer was 1 in every 118 males.

For all cancers combined, person-based prevalence was defined as the number of people who had been diagnosed with a primary invasive (or *in situ* bladder) cancer in a given time-frame and who were alive on January 1, 2005. For example, five-year person-based prevalence for all cancers combined refers to the number of people alive at the beginning of 2005 who had been diagnosed with cancer within the period 2000 to 2004.

Statistics Canada, Catalogue no. 82-003-XPE • Health Reports, Vol. 20, no.1, March 2009
Cancer prevalence in the Canadian population • Research Article

17

Cancer-specific, person-based prevalence was defined as the number of people diagnosed with a particular cancer in a given time-frame and who were alive on January 1, 2005. When cancers were combined for reporting purposes (for example, oral cancers, colorectal, leukemias and other and unknown), grouping was done before prevalence counts were determined.

The following example illustrates how cancers were counted in persons with more than one primary tumour diagnosis. For a person diagnosed with invasive primary breast cancers in 2001 and 2004, and also an invasive primary lung cancer in 2003, the 2001 breast cancer case and the lung cancer case would be counted in their respective five- and ten-year cancer-specific prevalence estimates. In calculating two-year prevalence estimates (based on cases diagnosed from January 1, 2003 to December 31, 2004), the lung cancer case and the 2004 breast cancer case would be counted in their respective cancer-specific prevalence estimates. There are other ways of determining person-based prevalence estimates in the case of multiple primaries, depending on the underlying question of interest; readers are invited to consider this when comparing person-based prevalence estimates from different sources. Note that, as defined above, the estimates of the prevalence of the individual cancers will not sum to the estimate given for all cancers combined.

Table A
Number of persons with cancer, by prevalence-duration, cancer and sex, Canada, January 1, 2005

| | Prevalence-duration | | | | | | | | |
| | Two-year | | | Five-year | | | Ten-year | | |
Cancer	Both sexes	Males	Females	Both sexes	Males	Females	Both sexes	Males	Females
All cancers	**212,606**	**108,099**	**104,507**	**441,155**	**221,180**	**219,975**	**695,049**	**338,912**	**356,137**
Oral cavity and pharynx	5,145	3,496	1,649	10,715	7,170	3,545	17,199	11,531	5,668
Esophagus	1,215	900	315	1,767	1,319	448	2,264	1,642	622
Stomach	2,888	1,909	979	4,788	3,054	1,734	6,872	4,227	2,645
Colorectal	29,173	15,816	13,357	59,766	32,184	27,582	91,934	48,725	43,209
Colon excluding rectum	19,269	9,748	9,521	39,592	19,861	19,731	61,450	30,446	31,004
Rectum and rectosigmoid	10,110	6,205	3,905	20,670	12,647	8,023	31,322	18,793	12,529
Anus	863	380	483	1,740	733	1,007	2,669	1,107	1,562
Liver	942	712	230	1,553	1,178	375	1,967	1,473	494
Pancreas	1,824	883	941	2,443	1,197	1,246	3,036	1,464	1,572
Larynx	1,861	1,521	340	3,908	3,215	693	6,700	5,478	1,222
Lung and bronchus	17,530	9,145	8,385	27,480	14,050	13,430	36,875	18,915	17,960
Soft tissue	1,440	792	648	2,997	1,642	1,355	4,886	2,620	2,266
Skin melanoma	7,692	4,059	3,633	17,378	8,911	8,467	29,401	14,590	14,811
Breast	37,069	304	36,765	85,734	595	85,139	146,635	943	145,692
Cervix uteri	2,344	...	2,344	5,428	...	5,428	9,989	...	9,989
Corpus uteri	6,888	...	6,888	15,366	...	15,366	26,458	...	26,458
Ovary	3,279	...	3,279	6,260	...	6,260	9,350	...	9,350
Prostate	37,582	37,582	...	85,953	85,953	...	135,061	135,061	...
Testis	1,557	1,557	...	3,814	3,814	...	7,068	7,068	...
Bladder (including in situ)	10,526	7,944	2,582	21,870	16,438	5,432	35,772	26,604	9,168
Kidney and renal pelvis	6,032	3,616	2,416	12,798	7,642	5,156	20,485	12,155	8,330
Brain	2,120	1,228	892	3,840	2,174	1,666	6,012	3,353	2,659
Thyroid	5,958	1,213	4,745	12,976	2,780	10,196	20,361	4,398	15,963
Hodgkin lymphoma	1,618	883	735	3,751	2,079	1,672	6,906	3,806	3,100
Non-Hodgkin lymphoma	9,253	4,930	4,323	18,991	10,015	8,976	29,617	15,315	14,302
Multiple myeloma	2,537	1,362	1,175	4,564	2,428	2,136	5,902	3,126	2,776
Leukemias	5,794	3,426	2,368	11,511	6,720	4,791	17,684	10,170	7,514
Other, unknown	12,963	6,568	6,395	24,360	12,153	12,207	37,124	18,281	18,843

... not applicable
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

Table B
Prevalence proportions (per 100,000), by prevalence-duration, cancer and sex, Canada, January 1, 2005

Cancer	Prevalence-duration								
	Two-year			Five-year			Ten-year		
	Both sexes	Males	Females	Both sexes	Males	Females	Both sexes	Males	Females
All cancers	**661.2**	**678.8**	**644.0**	**1,372.0**	**1,388.9**	**1,355.5**	**2,161.6**	**2,128.1**	**2,194.5**
Oral cavity and pharynx	16.0	22.0	10.2	33.3	45.0	21.8	53.5	72.4	34.9
Esophagus	3.8	5.7	1.9	5.5	8.3	2.8	7.0	10.3	3.8
Stomach	9.0	12.0	6.0	14.9	19.2	10.7	21.4	26.5	16.3
Colorectal	90.7	99.3	82.3	185.9	202.1	170.0	285.9	306.0	266.3
Colon excluding rectum	59.9	61.2	58.7	123.1	124.7	121.6	191.1	191.2	191.0
Rectum and rectosigmoid	31.4	39.0	24.1	64.3	79.4	49.4	97.4	118.0	77.2
Anus	2.7	2.4	3.0	5.4	4.6	6.2	8.3	7.0	9.6
Liver	2.9	4.5	1.4	4.8	7.4	2.3	6.1	9.2	3.0
Pancreas	5.7	5.5	5.8	7.6	7.5	7.7	9.4	9.2	9.7
Larynx	5.8	9.6	2.1	12.2	20.2	4.3	20.8	34.4	7.5
Lung and bronchus	54.5	57.4	51.7	85.5	88.2	82.8	114.7	118.8	110.7
Soft tissue	4.5	5.0	4.0	9.3	10.3	8.3	15.2	16.5	14.0
Skin melanoma	23.9	25.5	22.4	54.0	56.0	52.2	91.4	91.6	91.3
Breast	115.3	1.9	226.5	266.6	3.7	524.6	456.0	5.9	897.8
Cervix uteri	...	...	14.4	...	...	33.4	...	...	61.6
Corpus uteri	...	...	42.4	...	...	94.7	...	...	163.0
Ovary	...	...	20.2	...	...	38.6	...	...	57.6
Prostate	...	236.0	...	...	539.7	...	...	848.1	...
Testis	...	9.8	...	...	23.9	...	...	44.4	...
Bladder (including in situ)	32.7	49.9	15.9	68.0	103.2	33.5	111.3	167.1	56.5
Kidney and renal pelvis	18.8	22.7	14.9	39.8	48.0	31.8	63.7	76.3	51.3
Brain	6.6	7.7	5.5	11.9	13.7	10.3	18.7	21.1	16.4
Thyroid	18.5	7.6	29.2	40.4	17.5	62.8	63.3	27.6	98.4
Hodgkin lymphoma	5.0	5.5	4.5	11.7	13.1	10.3	21.5	23.9	19.1
Non-Hodgkin lymphoma	28.8	31.0	26.6	59.1	62.9	55.3	92.1	96.2	88.1
Multiple myeloma	7.9	8.6	7.2	14.2	15.2	13.2	18.4	19.6	17.1
Leukemias	18.0	21.5	14.6	35.8	42.2	29.5	55.0	63.9	46.3
Other, unknown	40.3	41.2	39.4	75.8	76.3	75.2	115.5	114.8	116.1

... not applicable
Note: Prevalence proportions were determined by a person-based analysis.
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

Statistics Canada, Catalogue no. 82-003-XPE • Health Reports, Vol. 20, no.1, March 2009
Cancer prevalence in the Canadian population • Research Article

19

Table C
Age-standardized* prevalence proportions (per 100,000), by prevalence-duration, cancer and sex, Canada, January 1, 2005

Cancer	Prevalence-duration								
	Two-year			Five-year			Ten-year		
	Both sexes	Males	Females	Both sexes	Males	Females	Both sexes	Males	Females
All cancers	**434.3**	**467.8**	**411.6**	**899.7**	**955.2**	**868.0**	**1,402.1**	**1,448.9**	**1,393.4**
Oral cavity and pharynx	10.7	15.4	6.4	22.0	31.3	13.5	34.7	49.6	21.3
Esophagus	2.3	3.8	1.0	3.4	5.5	1.5	4.3	6.8	2.0
Stomach	5.3	7.7	3.2	8.7	12.3	5.6	12.3	16.8	8.4
Colorectal	53.5	64.7	43.5	107.8	130.2	88.5	161.5	193.7	134.3
Colon excluding rectum	34.4	39.1	30.2	69.3	78.5	61.5	104.4	118.1	93.1
Rectum and rectosigmoid	19.1	25.6	13.3	38.5	51.7	27.0	57.0	75.6	41.1
Anus	1.7	1.6	1.8	3.4	3.1	3.7	5.1	4.6	5.6
Liver	2.0	3.1	0.9	3.4	5.3	1.6	4.3	6.8	2.1
Pancreas	3.5	3.8	3.3	4.7	5.1	4.4	5.8	6.2	5.5
Larynx	3.7	6.5	1.3	7.7	13.5	2.5	12.9	22.5	4.4
Lung and bronchus	33.9	37.5	31.2	52.7	57.2	49.6	69.9	76.3	65.6
Soft tissue	3.4	3.8	3.0	7.0	7.9	6.3	11.3	12.5	10.3
Skin melanoma	16.2	17.6	15.2	36.5	38.5	35.2	61.3	62.9	60.9
Breast	75.6	1.2	144.9	173.3	2.4	331.9	291.4	3.7	555.9
Cervix uteri	...	...	10.7	...	...	24.6	...	...	44.2
Corpus uteri	...	...	27.2	...	...	59.5	...	...	98.9
Ovary	...	...	13.6	...	...	26.0	...	...	38.6
Prostate	...	156.4	...	...	348.9	...	...	525.8	...
Testis	...	8.8	...	...	20.9	...	...	37.2	...
Bladder (including in situ)	18.6	30.9	8.4	38.4	63.5	17.4	61.7	101.6	28.9
Kidney and renal pelvis	12.5	15.8	9.5	26.3	33.2	20.2	41.6	52.3	32.1
Brain	5.8	6.7	4.9	10.9	12.3	9.5	17.1	19.1	15.2
Thyroid	14.4	5.8	23.0	31.0	13.2	48.7	47.9	20.7	74.8
Hodgkin lymphoma	4.6	5.1	4.2	10.5	11.6	9.5	18.9	20.6	17.2
Non-Hodgkin lymphoma	18.9	21.7	16.3	39.0	44.2	34.1	60.4	67.5	53.9
Multiple myeloma	4.7	5.5	3.9	8.4	9.9	7.1	10.9	12.8	9.2
Leukemias	13.1	16.4	10.2	26.4	32.5	20.9	41.0	49.6	33.3
Other, unknown	25.7	28.2	23.8	48.9	52.8	46.0	74.4	79.4	70.7

*   age-standardized to the World Standard Population[24,25]
... not applicable
Note: Analysis was based on tumour-specific data.
Source: Canadian Cancer Registry, Statistics Canada and Provincial/Territorial Cancer Registries.

Income and psychological distress: The role of the social environment

by Heather M. Orpana, Louise Lemyre and Ronald Gravel

Abstract

Background
This article examines the relationship between lower income and the risk of experiencing high psychological distress over twelve years.

Data and methods
Data from the first 12 years of the longitudinal National Population Health Survey (1994/1995 through 2006/2007) were analysed. Proportional hazards modelling was conducted to determine whether lower household income was associated with a greater risk of experiencing high distress, when adjusting for sociodemographic characteristics and baseline health status. It was also used to examine the relationship between reporting a stressor and experiencing a subsequent episode of distress.

Results
Overall, 11% of the initial sample experienced at least one episode of high distress during the 12 years of the study. Low-income respondents were at a significantly higher risk of becoming psychologically distressed, and many of the stressors were associated with a significantly higher risk of becoming distressed. Stressors accounted for 22% of the relationship between low income and distress for men, and more than a third of this relationship for women.

Interpretation
Low income is an important risk factor for becoming psychologically distressed, and stressors account for part of this increased risk.

Keywords

psychological distress, socio-economic status, income, stressors, stress, longitudinal studies

Authors

Heather M. Orpana (1-613-951-1650; Heather.Orpana@statcan.gc.ca) is with the Health Information and Research Division and Ronald Gravel (1-613-951-2295; Ronald.Gravel@statcan.gc.ca) is with the Health Statistics Division at Statistics Canada, Ottawa, Ontario, K1A 0T6. Louise Lemyre is with the University of Ottawa.

A large body of research has focused on the poorer physical health of individuals with low income, and important differences in the mental health of these groups can also be observed.[1,2] Much of this research, however, has been cross-sectional, making it difficult to determine whether low income or poor mental health comes first. As well, few studies have looked at this relationship in the Canadian context.

The primary purpose of this analysis is to examine whether people in lower-income groups had a greater risk of developing high levels of psychological distress over a twelve-year period, compared with those in higher-income groups. A second goal is to determine whether a higher level of stressors explains part of the relationship between lower income and high levels of psychological distress. Psychological distress is a non-specific psychological state characterized by feelings consistent with depressed mood or anxiety, such as feeling sad and nervous.[3] High levels of distress may indicate more serious disorders, such as clinical depression.[4]

Two directions for the association between income and mental health are possible: poorer mental health may lead to lower income, or lower income may cause poorer mental health.[5] The first is a health selection process, whereby individuals with mental health problems are less likely to be able to complete education or to engage in occupations associated with higher incomes. If this is the case, interventions should focus on improving mental health so that individuals can achieve higher education and income. Past research has shown, however, that health selection processes do not explain a large proportion of the relationship between low income and poorer mental health.[6]

Social causation, the second hypothesis, suggests that the poor overall mental health of low-income individuals is rooted in the negative social environment to which they are exposed. The social environment encompasses "the groups to which we belong, the

neighbourhoods in which we live, the organization of our workplaces, and the policies we create to order our lives."[7] One characteristic of a negative social environment is a higher prevalence of stressors, which are demands from the environment that tax or exceed the adaptive capacity of the individual.[8] Examples of stressors include marital discord, living in a neighbourhood that requires heightened vigilance because of crime, or high demands at work without adequate resources to meet the demands.

If the social causation hypothesis is correct, interventions aimed at improving a negative social environment would be expected to reduce inequities in health. Not only are individuals in lower income groups exposed to a high number of stressors, but they also have fewer resources, such as social support,[9] with which to cope and mitigate the negative effects of stressors. Higher stressors and few coping resources can increase feelings of distress, which have been linked to physical and mental disorders.[10]

In earlier Canadian research, Turner, Wheaton and Lloyd (1995) demonstrated a greater burden of stressors among low socio-economic status individuals in Toronto.[11] They reported that life events (major acute stressors such as the death of a spouse or the loss of one's job), chronic stressors (expected to continue over a period of months or years), and childhood traumas accounted for approximately a third of the relationship between occupation and depressive symptoms.

Analysing data from the first cycle of the National Population Health Survey (1994/1995), Cairney and Krause examined determinants of distress and depression among adults aged 50 or older.[9] They found higher levels of distress among those with less than a high school education, while no significant differences by education were observed for 12-month major depressive disorder. Accounting for stressors explained over a third of the

relationship between education and distress.

Matthews et al., who studied psychological distress at age 33 in the 1958 British Birth Cohort,[12] found that work factors, specifically job strain and job insecurity, had an important role in occupational class differences in psychological distress. This effect was more pronounced for men than women.

Finally, Myer et al. found the mediating effect of life events between lower socio-economic status and psychological distress to be significant, although they did not report the magnitude of the effect.[13]

While these studies support the hypothesis that the social distribution of stressors contributes to the poorer mental health of lower socio-economic groups, the research is limited in that the studies were cross-sectional. The sequencing of the relationship between income, stressors and psychological distress is unclear. Longitudinal data are better suited to identifying risk factors for developing mental health problems and identifying appropriate targets of intervention.

This article, based on data from the 1994/1995 to 2006/2007 National Population Health Survey (NPHS), examines the income gradient in *new cases of high psychological distress* in a representative sample of Canadians over a *twelve-year period* from 1994/1995 to 2006/2007. Stressors in the social environment are also investigated, because they are expected to be associated with subsequent experience of high distress, and are expected to explain part of the association between lower income and the risk of becoming distressed. This research provides important evidence, which has not been reported elsewhere, of the temporal ordering of income, stressors and psychological distress based on longitudinal analysis.

Data and methods

Data source

This analysis is based on data from seven cycles (cycles 1 to 7) of the household component of the National Population Health Survey (NPHS), conducted by Statistics Canada from 1994/1995 to 2006/2007. Every two years since 1994/1995, the NPHS has collected data about health status, health behaviours, and other determinants of health. This survey is representative of the household population of all provinces in 1994/1995, excluding members of the regular Canadian Forces and residents of Indian reserves, Crown Lands, health institutions, Canadian Forces bases (military and civilian) and some remote areas in Ontario and Quebec. Although the NPHS also has an institutional component covering residents of health institutions such as nursing homes, that sample was not analysed in this article.

In 1994/1995, 20,095 household residents were selected to be members of the NPHS longitudinal panel. Of these, 86.0% agreed to participate (17,276), and 83.6% provided responses to the in-depth questionnaire. Response rates in subsequent cycles were 92.8% in 1996/1997; 88.3% in 1998/1999; 84.9% in 2000/2001; 80.8% in 2002/2003; 77.6% in 2004/2005; and 77.0% in 2007/2008. More detailed descriptions of the NPHS design, sample and interview procedures are available elsewhere.[14,15] In 1994/1995, data were collected primarily through computer-assisted personal interviews; thereafter, primarily through computer assisted-telephone interviews.

Measures

Household income
Income was based on self-reported household income. To account for inflation, household size and the cost of living in different urban and rural areas, self-reported household income was divided by a low income cut-off (LICO), as determined by the respondent's place of residence and

household size. (For detailed information about this income variable, refer to the NPHS Cycle 7 Derived Variable documentation.)

A variable was created representing the ratio of the respondent's household income to a corresponding LICO. A household with a ratio below 1 is more likely to be in a difficult financial situation because its spending on necessities tends to be a greater proportion of its income. With a ratio above 1, a household is more likely in a better financial situation, because it spends a smaller proportion of its income on necessities.

From this ratio, a categorical variable representing lower and higher income was created. Respondents with values of 1.5 or less were considered to be *lower income*. Those with values more than 1.5 were considered to be *higher income*. For example, in a large urban centre, members of a four-person household with a total income of $46,400 or less in 1994 would be considered lower income, as would people living alone in a rural area if their income was $17,200 or less. Individuals with missing income values were included in the analyses as a separate income group, in order to retain as many respondents as possible. Although not reported here, analyses using three income categories yielded similar results, with middle-income respondents having a lower relative risk (RR) of experiencing an episode distress than did those in the lower-income group, and higher RR than did those in the higher-income group.

The same cut-offs were used in subsequent cycles to categorize household income. Not surprisingly, as the survey continued and the participants aged, the proportion of respondents in the lower income group decreased, and the proportion in the higher income group increased.

This method of measuring income accounts for numerous factors that affect the tangible meaning of household income, which is the ability to access goods and services in one's area.

Although this derived variable has LICO as its base, it is not meant to determine poverty, to measure income adequacy, or to estimate the number of Canadians in households whose total income is above or below the LICO.

Stressors
Recent life events, chronic stressors and job strain were assessed in the first, fourth, and subsequent NPHS cycles. Because stressors were not measured in the second and third cycles, the scores from 1994/1995 were imputed forward for these two cycles. Analyses using several methods of imputation yielded similar results; the strategy of imputing forward was adopted because it minimized the possibility of reverse causation. Recent life events and chronic stressors were measured with questions developed by Turner, Wheaton and Llyod.[11] Job strain was measured using an abbreviated version of the Job Content Questionnaire.[16]

The chronic stressors scale consisted of a series of 18 questions about situations that respondents reported they faced. The scale has been validated by Wheaton,[17] and demonstrates good convergent validity with indicators of difficult social circumstances, and discriminant validity with measures of psychological distress. The questions covered several domains: personal stress, such as "trying to take on too many things at once"; relationship problems with a spouse or partner; problems with one's children; family health problems; living in an undesirable neighbourhood; and financial problems. A dichotomous variable was created for each stressor domain, with 1 indicating a positive response to any question on the subscale, and 0 for those who answered negatively to all questions on that subscale.

Seven questions were used to measure the concept of job strain, which is based on Karasek's work with the demand control model and the US Quality of Employment Surveys.[16] A job is considered stressful if the incumbent must meet high demands without adequate resources and decision-making authority. Five questions measured decision latitude, and two measured psychological demands. Individuals falling in both the top third of demands and the lowest third of decision latitude based on the distribution of these variables in 1994/1995 were coded as experiencing job strain.

Distress
Distress was based on respondents' K6 scores. The K6 is a non-specific psychological distress measure developed by Kessler,[3,18] which has been used in numerous population-based surveys. The K6 measures distress through answers to six Likert-type questions scored from 0 to 6, which are summed to form a scale score ranging from 0 to 24. Kessler proposed a cut-off of 13 or higher as representing likely serious mental illness, based on requirements of the Substance Abuse and Mental Health Services Administration in the United States.[18] However, we conducted psychometric analyses to identify a less restrictive cut-off for the K6 if the outcome of interest was major depressive disorder or an anxiety disorder. Based on these analyses, a score of 9 or more was considered to indicate high psychological distress and possible mood or anxiety disorder. Recent analyses by Cairney et al. show that the K6 is a good screening tool for depression, and that a cut-off of 9 or more results in a stratum-specific likelihood ratio of more than 6 for 12-month major depressive disorder.[4] To control for the graded nature of the distress score for scores below the cut-off point of 9, a low distress score variable was included in the analysis.

The sociodemographic variables were: age, marital status (married/with partner/with spouse versus single or widowed/separated/divorced), urban/rural residence, birthplace, labour force status (employed versus unemployed or out of the labour force), and presence of children in the household. Because of the association between physical and mental health problems, self-rated health was included in models predicting

distress (scored on a 5-point scale ranging from poor to excellent).

Analysis

To be included in the dataset for analysis, respondents had to have a low distress score (see *Measures*) in 1994/1995 and be aged 18 to 75. Respondents contributed another observation to the dataset for each cycle through which they survived without experiencing high distress. If an individual's distress score was missing for the end of a studied interval, that interval was their last observed period, and they received a score of 0 on the event variable.

Proportional hazards modelling was conducted to determine if lower household income was associated with a greater risk of experiencing an episode of high distress. The first model (Model 1) examined income and cases of high distress, while controlling for sociodemographic predictors (age, marital status, urban/rural residence, birthplace, labour force status, presence of children in the household) and self-rated health and baseline low distress score. The predictors in the model were time-varying covariates: they reflected the value for the respondent two years before observation of the event (that is, becoming distressed). The second model (Model 2) included stressors, which were also time-varying. Because the stressors were not measured at cycles 2 and 3, values for cycle 1 were imputed to cycles 2 and 3. The analysis was conducted in SAS using the bootstrap weights developed by Statistics Canada.[19,20] Because of the discrete nature of the time of measurement, logistic regression using a complementary log-log link was conducted, excluding the intercept and including an indicator variable for time of observation.

Since one of the goals was to determine if stressors were a mediating factor linking lower income with higher distress, a mediational analysis was conducted.[21] A mediational analysis examines whether a proposed mediator (in this case, stressors) may explain part of the relationship between and

independent variable (IV: income) and the dependent variable (DV: high psychological distress). Mediation is demonstrated by a reduction of the relationship between the IV and the DV when the proposed mediator is taken into account. A detailed explanation is provided in Baron and Kenny.[21] To ensure consistency across models, only individuals with full data on all variables were included, resulting in the exclusion of 1,077 records, or slightly more than 2% of observations.

The percentage reduction in the relative risk (RR) associated with income between the model excluding (Model 1) and including stressors (Model 2) was calculated: (RR1-RR2)/(RR1-1).

Results

Prevalence of high distress and stressors

Among an initial cohort of 11,058 individuals aged 18 to 75 with low distress scores in 1994/1995, 1,191 first episodes of high psychological distress were observed over the period to 2006/2007. Thus, 11% of the sample experienced high distress at some point during those twelve years. Among the sample with no missing data (n=10,948), 47,379 person-periods at risk were observed, with 1,152 episodes of high psychological distress. There were fewer cases of high distress among

Table 1
Prevalence of selected characteristics of sample, by sex, Canada 1994/1995

	Men		Women	
	Number	Percent	Number	Percent
Sample size	5,014	100.0	5,934	100.0
Household income				
Lower	1,607	29.9	2,396	37.4
Higher	3,100	62.7	3,188	56.5
Missing	307	7.3	350	6.1
Age (mean)	42	...	43	...
Marital status				
Married	3,111	68.2	3,535	67.4
Single	1,330	24.4	1,109	17.0
Divorced/Widowed/ Separated	573	7.4	1,290	15.7
Residence				
Urban	3,692	81.7	4,565	83.7
Rural	1,322	18.3	1,369	16.3
Birthplace				
Canada	4,336	81.5	5,141	80.9
Foreign-born	678	18.5	793	19.1
Labour force status				
Employed	3,139	65.7	2,729	47.1
Unemployed	640	12.5	804	14.4
Out of labour force	1,235	21.8	2,401	38.6
Children in household				
Children	3,145	64.7	4,342	74.3
Children younger than 12	1,128	28.1	1,701	32.7
Stressors				
Personal stress	2,641	55.9	3,566	61.5
Financial problems	1,986	39.6	2,139	35.0
Neighborhood problems	1,177	24.6	1,375	24.0
Problems with children	1,071	21.0	1,496	24.7
Family health problems	1,049	19.9	1,492	24.0
Relationship problems	509	11.2	712	14.1
Job strain	358	7.0	536	9.0
Recent life events (mean number)	0.47	...	0.49	...

... not applicable
Source: 1994/1995 to 2006/2007 National Population Health Survey.

men than among women (407 versus 745).

Some stressors were more prevalent than others in 1994/1995 (Table 1). Personal stress, such as having too many things expected by others, was the most common stressor, reported by the majority of men (56%) and women (61%). A quarter of men and women reported problems with their neighbourhood. Relationship problems were relatively uncommon— 11% of men and 14% of women. Job strain was even less prevalent, with 7% of men and 9% of women in the sample experiencing it.

Lower income and risk of becoming distressed

Among both men and women, lower household income was significantly associated with a higher risk of becoming distressed (Table 2, Model 1). Lower-income men were 1.58 times as likely

Table 2

Relative risk of becoming distressed, by household income group, sociodemographic characteristics, stressors and sex, household population, Canada excluding territories, 1994/1995 to 2006/2007

	Men						Women					
	Model 1			Model 2			Model 1			Model 2		
	Relative risk	95% confidence interval		Relative risk	95% confidence interval		Relative risk	95% confidence interval		Relative risk	95% confidence interval	
		from	to		from	to		from	to		from	to
Household income												
Lower	1.58*	1.15	2.17	1.45*	1.07	1.98	1.25*	1.02	1.53	1.16	0.95	1.43
Higher†	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Age‡	0.99	0.98	1.00	0.99	0.99	1.00	0.98*	0.97	0.99	0.99	0.98	1.00
Marital status												
Married†	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Single	0.85	0.56	1.28	0.86	0.57	1.30	1.13	0.83	1.53	1.16	0.85	1.59
Divorced/Widowed/Separated	1.28	0.86	1.91	1.30	0.87	1.94	1.09	0.85	1.38	1.08	0.84	1.38
Residence												
Urban†	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Rural	1.11	0.85	1.45	1.15	0.87	1.51	0.94	0.76	1.18	0.94	0.76	1.18
Birthplace												
Canada†	1.00	...	...	1.00	...	...	1.00			1.00	...	...
Foreign-born	1.22	0.86	1.74	1.24	0.87	1.77	1.13	0.86	1.48	1.13	0.86	1.48
Labour force status												
Employed†	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Unemployed	1.00	0.65	1.53	1.44	0.97	2.12	1.43	1.00	2.04	1.56*	1.06	2.28
Out of labour force	1.18	0.81	1.70	1.11	0.71	1.73	1.11	0.86	1.44	1.29	0.99	1.69
Children in household												
None	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Children	0.65*	0.46	0.91	0.68*	0.47	0.98	1.28	0.92	1.78	1.05	0.75	1.47
Children younger than 12	0.96	0.62	1.49	0.92	0.60	1.41	0.89	0.69	1.14	0.90	0.71	1.14
Baseline distress	1.53*	1.38	1.71	1.46*	1.30	1.64	1.59*	1.44	1.75	1.50*	1.36	1.66
Baseline self-rated health	1.51*	1.32	1.74	1.47*	1.27	1.69	1.39*	1.25	1.55	1.33*	1.19	1.49
Stressors§												
Job strain	...	...	...	1.95*	1.28	2.97	...	...	...	1.34	0.97	1.85
Financial problems	...	...	...	1.32	1.00	1.75	...	...	...	1.19	0.98	1.45
Personal stress	...	...	...	1.26	0.96	1.66	...	...	...	1.61*	1.26	2.07
Relationship problems	...	...	...	1.31	0.92	1.87	...	...	...	1.20	0.93	1.56
Neighbourhood problems	...	...	...	1.26	0.95	1.66	...	...	...	1.04	0.84	1.29
Problems with children	...	...	...	0.88	0.62	1.23	...	...	...	1.23	0.98	1.54
Family health problems	...	...	...	1.07	0.79	1.47	...	...	...	0.96	0.79	1.17
Recent life events	...	...	...	1.06	0.91	1.23	...	...	...	1.13*	1.04	1.24
- 2 log likelihood	3929.27			3873.61			5669.42			5594.51		

† reference category
‡ use as a continuous variable
§ reference category is absence of stressor
* significantly different from estimate for reference category (p < 0.05)
... not applicable
Source: 1994/1995 to 2006/2007 National Population Health Survey.

Table 3
Relative risk of becoming distressed by stressors and sex, controlling for sociodemographic factors, Canada 1994/1995 to 2006/2007

	Men			Women		
		95% confidence interval			95% confidence interval	
	Relative risk	from	to	Relative risk	from	to
Job strain†	2.04*	1.34	3.11	1.44*	1.04	1.99
Financial problems†	1.50*	1.14	1.96	1.34*	1.11	1.63
Personal stress†	1.38*	1.05	1.82	1.74*	1.37	2.21
Relationship problems†	1.51*	1.07	2.12	1.40*	1.07	1.82
Neighbourhood problems†	1.41*	1.08	1.84	1.19	0.97	1.47
Problems with children†	0.99	0.70	1.39	1.39*	1.11	1.73
Family health problems†	1.18	0.88	1.57	1.11	0.91	1.35
Recent life events‡	1.13	0.98	1.31	1.19*	1.09	1.30

† reference category = stressor not reported
‡ relative risk = increase in risk with one unit increase in life events
* significantly different from estimate for reference category (p < 0.05)
Source: 1994/1995 to 2006/2007 National Population Health Survey.

as higher-income men to become distressed, even when the influence of factors such as age, urban/rural residence, and immigrant, parental and labour force status was controlled. Lower-income women were at a 25% greater risk of becoming distressed than were higher-income women.

Other risk factors

For both sexes, other characteristics significantly associated with a heightened risk of becoming distressed were poorer self-rated health and higher baseline distress scores. For men, having children in the household was associated with a significantly reduced risk of becoming distressed, compared with not having children. For women, younger age and being unemployed rather than employed was associated with a significantly higher risk of becoming distressed.

Stressors and risk of distress

When studied individually (Table 3), most of the stressors were associated with a significantly higher risk of becoming distressed. Men experiencing job strain were twice as likely to become distressed as were those in low-strain jobs; women experiencing job strain had a 44% greater risk. Financial problems, personal stress and relationship problems were each

significantly related to the risk of becoming distressed for both sexes. For men, neighbourhood problems increased the risk by 41%. By contrast, problems with children were associated with becoming distressed for women but not for men: women reporting problems with their children were 39% more likely to become distressed than were women without this stressor. Recent life events were also significant for women, but not for men. With almost every life event reported, women were about 20% more likely to develop distress. Family health problems were not a significant risk factor for either sex.

Stressors mediate income-distress relationship

When stressors were included in the model, the relationship between household income and distress diminished. The reduction in the -2 log likelihood from Model 1 to Model 2 was significant for both sexes, indicating that the addition of stressors improved model fit. The reduction in the relative risk associated with low income was 22% for men and 36% for women.

Discussion

This study demonstrates that lower income is associated with a higher risk of becoming distressed, but that this risk is partially accounted for by the higher prevalence of stressors in the lives of lower-income individuals. The study supports the social causation hypothesis of the income gradient in health, because lower income preceded the development of high psychological distress.

The social environment also appears to be implicated in this relationship. Because a fairly high level of household income was chosen to define the lower-income group (1.5 times or less the low income cut-off), a substantial proportion of respondents were not living in material deprivation. Thus, we conclude that factors beyond material deprivation are contributing to income differences in mental health. While financial and neighbourhood problems are more closely related to material resources, stressors such as difficulties with children and relationships are more clearly social in nature. Because the association between income and developing psychological distress weakened when stressors were taken into account, this study provides support for the mediating role of social stressors in income-related health inequalities.

The results of this analysis parallel those of studies examining stressors as mediators of the income-physical health relationship, using cross-sectional[22] and longitudinal analyses.[23] While other research has demonstrated the role of stressors in the income gradient in mental health,[9,11,12] the longitudinal nature of the data analysed in this article provides evidence for the temporal sequencing of the income-distress relationship. In this study, low self-rated health and the presence of even low distress at baseline also predicted episodes of high psychological distress, consistent with findings from cross-sectional studies. And among the participants in this study, lower income and stressors preceded distress.

What is already known on this subject?

- Lower income is strongly related to individuals' mental health, including their feelings of psychological distress.
- Stressors are more prevalent among lower income groups and may explain part of this relationship.
- The temporal ordering of these factors has not been demonstrated in a Canadian population survey.

What does this study add?

- This paper shows that lower income is significantly related to future episodes of high psychological distress, and that stressors mediate a modest part of this relationship.
- The everyday social environments of low-income Canadians are implicated in health disparities.

Limitations

This analysis has a number of limitations. First, all data were self-reported. It is well known that negative affectivity is associated with reporting stressors, poor health and distress, and thus, the results may reflect confounding by this unmeasured variable.[24] However, other research has demonstrated that negative affectivity may be on the causal pathway linking stressors and poor health, and consequently, this may not be an instance of confounding but of mediation.[25]

Respondents were interviewed at two-year intervals; no information is available about what happened to them between interviews. It is possible that they experienced an unmeasured episode of high distress between observations.

As well, stressors were measured only at cycles 1, 4, 5 and 6, and were imputed for cycles 2 and 3 from cycle 1. It is possible that stressors resolved themselves between cycles 1 and 4, or that new stressors appeared that were not taken into account.

Finally, not all respondents were followed up in all cycles. Although the attrition rate in the NPHS is quite low compared with other surveys, a substantial number of respondents were lost from the study. If those who continued in the survey differed systematically from those who dropped out, systematic bias may have been have introduced into the results. Even so, other analyses of NPHS data have demonstrated that those lost to attrition were no more or less likely to experience episodes of distress than were those who were always present.[26] Thus, it is unlikely that attrition biased the results.

Conclusion

Future research should examine the predictors of multiple episodes of distress and the chronicity of distress, neither of which were analysed in the present study. Changes in income may also be important predictors of distress and should be explored further. Finally, including other components of a stress process model, such as mastery, self-esteem and social support, in an expanded model may further clarify the complex pathways between income, stressors and poor mental health. While more research is clearly needed, the present results point to the need to examine the social environment as one of the drivers in income-related disparities in mental health among Canadians. ■

References

1. Turner RJ, Lloyd D. The stress process and the social distribution of depression. *Journal of Health and Social Behavior* 1999; 40: 374-404.

2. Adler N, Boyce T, Chesney M, et al. Socioeconomic status and health: The challenge of the gradient. *American Psychologist* 1994; 49: 15-24.

3. Kessler RC, Andrews G, Colpe LJ, Hiripi R, Mroczek DK, Normand S-LT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine* 2002; 32: 959-76.

4. Cairney J, Veldhuizen B, Wade TJ, et al. Evaluation of 2 measures of psychological distress as screeners for depression in the general population. *Canadian Journal of Psychiatry* 2007; 52: 111-20.

5. Elstad J. *Social Inequalities in Health and Their Explanations.* Oslo: NOVA, 2000.

6. Chandola T, Bartley M, Sacker A, et al. Health selection in the Whitehall II study, UK. *Social Science and Medicine* 2003; 56: 2059-72.

7. Yen IH, Syme SL. The social environment and health: A discussion of the epidemiological literature. *Annual Review of Public Health* 1999; 20: 287-308.

8. Selye H. *The Stress of Life*. New York: McGraw Hill, 1956.

9. Cairney J, Krause N. The social distribution of psychological distress and depression in older adults. *Journal of Aging and Health* 2005; 17(6): 807-35.

10. McEwen B. Protective and damaging effects of stress mediators. *New England Journal of Medicine* 1998; 338: 171-9.

11. Turner RJ, Wheaton B, Lloyd D. The epidemiology of social stress. *American Sociological Review* 1995; 60(1): 104-25.

12. Matthews S, Power C, Stansfeld SA. Psychological distress and work and home roles: a focus on socio-economic differences in distress. *Psychological Medicine* 2001; 31(4): 725-36.

13. Myer L, Stein DJ, Grimsrud A, et al. Social determinants of psychological distress in a nationally-representative sample of South African adults. *Social Science and Medicine* 2008; 66: 1828-40.

14. Tambay J-L, Catlin G. Sample design of the National Population Health Survey—its longitudinal nature. *Health Reports* (Statistics Canada, Catalogue 82-003) 1995; 7(1): 29-38.

15. Statistics Canada. *NPHS, Household Component, Cycles 1 to 7 (1994/1995 to 2006/2007), Longitudinal Documentation.* Ottawa: Statistics Canada, 2008.

16. Schwartz JE, Pieper CF, Karasek RA. A procedure for linking psychosocial job characteristics data to health surveys. *American Journal of Public Health* 1988; 78(8): 904-9.

17. Wheaton B. Sampling the stress universe. In: Avison WR, Gotlib I, editors. *Stress and Mental Health: Contemporary Issues and Prospects for the Future.* New York: Plenum Press, 1994: 77-114.

18. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. *Archives of General Psychiatry* 2003; 60: 184-9.

19. Rao J, Wu C, Yue K. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue 12-001)1992; 18(2): 209-17.

20. Rust K, Rao J. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 1996; 5: 281-310.

21. Baron R, Kenny D. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 1986; 51(6): 1173-82.

22. Orpana H, Lemyre L. Explaining the social gradient in health in Canada: Using the National Population Health Survey to examine the role of stressors. *International Journal of Behavioral Medicine* 2004; 11(3): 143-51.

23. Orpana H, Lemyre L, Kelly S. Do stressors explain the relationship between income and declines in self-rated health? A longitudinal analysis of the National Population Health Survey. *International Journal of Behavioral Medicine* 2007; 14(1): 40-7.

24. Watson D, Pennebaker J. Health complaints, stress and distress: Exploring the central role of negative affectivity. *Psychological Review* 1989; 96(2): 234-54.

25. Spector P, Zapf D, Chen P, Frese M. Why negative affectivity should not be controlled for in job stress research: Don't throw the baby out with the bath water. *Journal of Organizational Behavior* 2000; 21: 79-95.

26. Orpana H. Using the National Population Health Survey to identify factors associated with patterns of psychological distress over 10 years. *Healthcare Policy* 2007; 394): 55-63.

The natural history of health-related quality of life: A 10-year cohort study

by Heather M. Orpana, Nancy Ross, David Feeny, Bentson McFarland, Julie Bernier and Mark Kaplan

Abstract

Background
Taking account of the impacts of institutionalization and death, this study describes the normative trajectories of health-related quality of life (HRQL) in Canada as individuals age from mid - to late life.

Methods
A nationally representative sample of 7,915 community-dwelling adults aged 40 and older in 1994/1995 was studied using 10 years of data from the longitudinal National Population Health Survey. Growth curve models of HRQL over age were fitted to describe the evolution of HRQL. Successive models were tested, first including only those living in a household throughout the entire period, then adding those who moved to an institution, and finally, including those who had died.

Results
HRQL remained generally stable until approximately age 70, when it began to decline. Excluding individuals when they were institutionalized, or ignoring the impact of death resulted in overly optimistic trajectories of HRQL in later years.

Interpretation
These results demonstrate the importance of following individuals into institutions and accounting for death in the production of realistic health estimates in aging populations.

Keywords

aging, health status, health surveys, life expectancy, longevity, longitudinal studies

Authors

Heather M. Orpana (613-951-1650; Heather.Orpana@statcan.gc.ca) and Julie Bernier (613-951-4556; Julie.Bernier@statcan.gc.ca) are with the Health Information and Research Division at Statistics Canada, Ottawa, Ontario, K1A 0T6; Nancy Ross is with McGill University; David Feeny is with Kaiser Permanente Northwest; Bentson McFarland is with the Oregon Health and Science University; and Mark Kaplan is with Portland State University.

In Canada, as in most developed countries, the average age of the population and life expectancies are increasing.[1] The resulting demographic shift toward a population with a larger proportion of older adults has directed attention to understanding how health evolves among adults from mid- to later life. Whether the population is experiencing a compression of morbidity, with ill health being confined to the last few years before death,[2,3] or an expansion of morbidity with the additional years of life lived with disease, disability and loss of quality of life, has implications for society as a whole and for the health care system.[4]

Few studies have employed a growth curve approach to describe health using longitudinal data.[5] However, estimates from longitudinal data of how health changes as people age are likely to differ considerably from those derived from cross-sectional studies.[6] As well, the studies that have employed a growth curve model approach often do not use representative samples, and thus, have limited external validity. An exception is research by Liang et al. who analyzed data for a sample of 2,200 older Japanese adults over 12 years.[7]

They observed a slight decline in self-rated health (SRH) from age 60 to 85, after which SRH actually improved. However, this effect may have been due to the deaths of individuals with lower SRH. In another study, McCullogh et al.[8] examined the participants in the Terman Life Cycle Study of Children with High Ability over 59 years. They show a model of SRH decline for both men and women beginning at about age 50 and becoming steeper around age 70, but it is doubtful that these results are generalizable to

entire populations, given that this was a select group who were likely more affluent and, in turn, healthier than most populations.

The purpose of the present study is to describe the pattern of HRQL from mid- to late adulthood in a representative sample of Canadians, while taking into account institutionalization and mortality. The longitudinal National Population Health Survey (NPHS) offers a unique opportunity to study HRQL trajectories in a large sample of the Canadian adult population.[9] Two advantages of the NPHS for this type of analysis are that it follows individuals through the transition from the household to institutions (a common experience among the oldest old),[10] and it provides information about deaths occurring in the sample.

The impacts of institutionalization and mortality on population estimates of health have frequently been overlooked in studies describing normative patterns of health.[11] Indeed, institutionalization is an important consideration when estimating population health, given that in 1992, 24% of the Canadian population aged 80 or older was living in a health care institution.[10] Many longitudinal studies are confined to household samples and thereby exclude the sickest members of society who often live in health care institutions. Such analyses present overly optimistic estimates of the health of the population as it ages. Ignoring the effect of mortality can also result in an overestimate of the health status of the older population, as generally only the healthier individuals survive.[11,12]

Because of differences in morbidity, institutionalization and mortality between the sexes, this analysis shows separate trajectories for men and women. Although some evidence demonstrates that women live longer than men but have a higher burden of morbidity,[13] sex differences in morbidity among those who remain alive may not be as great as previously suggested. In Canada, women's life expectancy surpasses that of men at birth (reference year 2002) and at age 65 (reference year 2001): 82.1 and 20.5 years for women versus 77.2 and 17 years for men.[14,15] As well, women have higher health-adjusted life expectancy (reference year 2001) at birth and at age 65: 70.8 and 14.4 years for women versus 68.3 and 12.7 years for men. However, at age 65, fewer men (77%) than women (85%) have at least one chronic condition, and women are far more likely to require help with instrumental activities of daily living (29% of women versus 15% of men).[16] From ages 45 to 79, men are more likely than women to live in an institution such as a nursing home; thereafter, women are more likely to be institutionalized.[10]

Methods

Sample and data

This analysis is based on longitudinal data from the first six cycles (1994/1995 through 2004/2005) of the National Population Health Survey (NPHS). The target population of the NPHS Household component includes household residents in the ten Canadian provinces in 1994/1995, excluding persons living on Indian Reserves and Crown Lands, and residents of health institutions, Canadian Forces Bases and some remote areas in Ontario and Quebec.

In 1994/1995, 20,095 households were selected for the NPHS longitudinal panel. Of these, 86% completed the general component of the questionnaire (17,276) and 83.6% of selected respondents provided responses to the in-depth health questionnaire. Response rates in subsequent cycles based on the 17,276 selected respondents were 92.8% in 1996/1997; 88.3% in 1998/1999; 84.9% in 2000/2002; 80.8% in 2002/2003, and 77.6% in 2004/2005. More detailed descriptions of the NPHS design, sample and interview procedures are available elsewhere.[9] Data were collected primarily through computer-assisted personal interviews in 1994/ 1995 and primarily through computer assisted-telephone interviews thereafter

NPHS respondents were followed up every two years. In the first NPHS cycle, the majority of respondents were interviewed in person; in later cycles, the majority were interviewed by telephone. For this study, 10 years of data were analyzed (1994/1995 through 2004/2005). Attrition due to non-response increased with subsequent cycles, although after 10 years only 17% of respondents aged 40 or older had been lost to follow-up, which is modest compared with attrition in other longitudinal studies of older adults.[17] Because growth curve modelling can be accomplished even with missing data (under certain conditions), the effect of non-response is expected to be minimal in this analysis.[5]

Measures

Health-related quality of life (HRQL) has been defined as "the value assigned to duration of life as modified by the impairments, functional states, perceptions and social opportunities that are influenced by disease, injury, treatment or policy."[18] HRQL was measured by the Health Utilities Index Mark 3 (HUI3). The HUI3 describes health status using eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain and discomfort.[19] Each attribute has 5 or 6 levels that range from severely impaired (for instance, blind for vision) to no impairment. HUI3 health states are scored using utility functions based on preference scores obtained from a Canadian sample. Thus, each individual has an HUI3 score for each measurement time, which reflects an overall level of HRQL based on the combination of attribute levels they experience.

Overall HUI3 scores can range from -0.36 to 1.00. A score of 1.00 is considered perfect health, while a score of 0 represents the state of being dead, and a score less than 0, a state "worse than dead." Scores less than 0 are possible because certain combinations

of health attributes were considered by the Canadian preference scoring sample to be less preferable than being dead. A score of -0.36 represents the health state comprised of the lowest level of every attribute. For analyses including participants for the first cycle after their death, an HUI3 score of 0 was imputed for that cycle. For example, someone who died in 1997 would have a value of 0 ascribed for the 1998/99 cycle, and be excluded from the analyses thereafter.

Age in years was centered on age 40 (by subtracting 40 from each participant's stated age) to improve the interpretability of estimates. For participants who died, the record for the cycle after their death included an imputed age variable equal to their age at death. Each of these age measures was also squared and cubed in each analysis to allow for the testing of quadratic and cubic effects, because many health states show an increasing rate of decline at older ages, without an initial increase which would be observed in a quadratic model. A cubic model also appeared to be the better fit to plotted raw data.

Analysis

Three analytical groups were created. To demonstrate the effect of confining the analysis to the healthiest Canadians, only data for individuals living in a household were analyzed in Model 1. To demonstrate the effect of accounting for those who became ill enough to be institutionalized, Model 2 also included residents of institutions. Finally, to show that descriptions of population health are heavily affected by ignoring death as a health state, Model 3 included data not only for all living participants (residents of household and institutions), but also for those who had died—for the first cycle at which death was recorded, their age at death and an HUI3 score of 0 were the data values. Data for decedents in subsequent cycles were not included.

Multilevel growth curve models were estimated to describe the normative trajectory of HRQL in order to answer the question, "How does HRQL change as adults age from mid-life?" A multilevel growth curve model is a hierarchical linear model in which observations over time (level 1) are nested within an individual (level 2). The analysis accounts for the non-independence between observations of the same individual at multiple time points.[5]

A person-period dataset was created, with one record per participant for each cycle at which their HUI3 score was available. A two-level unconditional growth curve model was built in MPLUS[20] predicting HUI3 score from age, age squared and age cubed, and specifying random intercept, slope, and quadratic and cubic terms for each of the three analytical groups. The first level was the within-individual growth model, specified as:

$$HUI_{ij} = \alpha_{0i} + B_{1i}(age_{ij}) + B_{2i}(age^2_{ij}) + B_{3i}(age^3_{ij}) + r_{ij}$$

The second level allows for model parameters to vary between individuals (random effects). It was specified as:

$$\alpha_{0i} = \gamma_{00} + u_{0i}$$
$$B_{1i} = \gamma_{10} + u_{1i}$$
$$B_{2i} = \gamma_{20} + u_{2i}$$
$$B_{3i} = \gamma_{30} + u_{3i}$$

Normalized weights were used to ensure that the sample reflected the Canadian population. Because of the complex sample design of the NPHS, which can result in artificially small variance estimates,[21] a conservative p value of 0.001 was chosen as the threshold for significance to reduce the risk of Type I errors.

Results

The sample of 7,915 community-dwelling adults reflected the Canadian household population aged 40 or older in 1994/1995, and was comprised of 52% men and 48% women. Their mean age was 57 years (range from 40 to 102 in 1994/1995). Most respondents were married or living with a partner. The age and sex distribution by 10-year age group in 1994/1995 can be seen in Table 1. Over the 10 years of the study, 1,562 respondents died. At any cycle, a small number of respondents were institutionalized, ranging from 62 in cycle 2 to 160 in cycle 5.

In 1994/1995, respondents' mean HUI3 score was 0.833, and their modal score was 0.973 (data not shown). The HUI3 score was negatively skewed (skewness = -2.52). As expected, mean HUI3 score decreased with age (Table 1). The intraclass correlation coefficients (ICC), an indicator of the average autocorrelation of the dependent variable across observations, were moderate. The ICC for the HUI3 over time for men living in households was 0.48; for women, 0.52. For those living

Table 1
Characteristics of sample in 1994/1995 and observations over study period, by 10-year age group, population aged 40 or older in 1994/1995, Canada excluding territories

Age group	% female	n in 1994/1995	Observations over study period	Mean HUI3 in 1994/1995
40 to 49	48.5	2,511	7,037	0.88
50 to 59	51.0	1,829	10,903	0.85
60 to 69	53.0	1,655	8,678	0.82
70 to 79	58.6	1,340	7,326	0.77
80 to 89	61.3	518	3,562	0.66
90 or older	61.3	62	562	0.44

Note: N are unweighted; percents and means are weighted estimates.
Source: 1994/1995 to 2004/2005 National Population Health Survey.

Table 2
Parameter estimates for growth curve models of HUI3 over age for men and women aged 40 or older in 1994/1995, Canada excluding territories

	Model 1 (household only)				Model 2 (household and institutions)				Model 3 (household, institutions and deceased)			
	Men		Women		Men		Women		Men		Women	
	Estimate	t-value	Estimate	t-value	Estimate	t-value	Estimate	t-value	Estimate	t-value	Estimate	t-value
Fixed effects												
Intercept	0.918722	119.99	0.911904	88.48	0.926754	115.0	0.917889	80.47	0.926585	106.95	0.921052	95.97
Age	-0.005125	-3.75	-0.009304	-4.69	-0.007832	-5.55	-0.011380	-5.61	-0.008415	-4.91	-0.012727	-6.76
Age squared	0.000279	3.62	0.000502	5.02	0.000483	6.0	0.000664	6.71	0.000548	5.38	0.000768	8.05
Age cubed	-0.000007	-5.96	-0.000010	-6.95	-0.000011	-8.53	-0.000013	-9.35	-0.000015	-8.96	-0.000016	-11.89
Random effects												
Variance intercept	0.010974	5.68	0.017180	8.75	0.017049	3.11	0.032053	3.46	0.007577	3.40	0.015725	7.79
Variance age	0.000028	8.74	0.000022	7.43	0.000052	5.05	0.000058	4.25	0.000054	11.51	0.000034	10.43

Note: All reported estimates are significant at the 0.001 level. t-values (estimate/standard error of estimate) are reported instead of standard error to reduce number of digits in table.
Age is centered on age 40.
Source: 1994/1995 to 2004/2005 National Population Health Survey.

in households or in institutions, the ICC was 0.47 for men and 0.49 for women. This indicates an important degree of autocorrelation in the data, with about half of the variation in the HUI3 scores over time being within individuals, and half between individuals,[5] thus making a multilevel growth curve model appropriate for these data.

Compared to those with complete data at all six cycles (n=3,375), women for whom data were missing for one to three cycles (n=687) were similar in terms of age in 1994/1995 and baseline HUI3 score. However, women for whom data were missing for four to six cycles (n=300) were four years younger and marginally healthier than women with complete data. Among men with one to three cycles of missing data (n = 650), average age was three years younger than that of those with complete data (n = 2,613); their HUI3 scores were almost identical. For men with four to six cycles of missing data (n=290), average age was almost five years younger than that of those with complete data, and as would be expected for a younger group, their health was marginally better. These relatively small differences between groups lead to the conclusion that missing data likely had little effect on the findings.

## Normative growth curve for HRQL

The first growth curve model examined the trajectories of HRQL for men and women separately. The growth curves illustrate the pattern of HRQL of Canadian men and women from age 40 on, given the HRQL observed in the cohort from 1994/1995 to 2004/ 2005. Model parameters are shown in Table 2, and the modeled normative trajectories are shown in Figure 1 for men, and in Figure 2 for women. For both sexes in all analyses, only the intercept and the linear function of age had significant random effects (that is, significant variation between participants). Thus, only a random intercept and age term were included in the model specification. Because no interindividual predictors were included in the model, these terms are not interpreted further in this paper. The significant variability in these components of the model indicates that future work should examine the determinants of interindividual differences in HUI3 trajectories. In all models, the covariance between the intercept and the linear function of age was non-significant. The fixed effects of age squared and age cubed were significant in all models.

According to Model 1 (Table 2), at age 40, men's average HUI3 score was 0.92. With a one-year increase in age, this value declined by the sum of a decrease of 0.005*(age – 40), an increase of 0.0003*(age – 40)^2, and a decrease of 0.000007*(age – 40)^3. Although the coefficients in the equation predicting HUI3 from age, age squared and age cubed appear small, at increasing ages their effects are quite large, as a decrease in an overall HUI3 score of 0.03 or more is considered to be clinically important.[22,23]

The growth curves for men and women living in a household (Model 1) were relatively similar, with men having better HRQL than women before age 74, and the trend reversing after age 74 (Table 2, Figures 1 and 2). An important difference between the sexes was the decrease in HUI3 among the youngest women in the cohort. From age 40 to age 50, women's average HUI3 fell by 0.06, twice the threshold considered clinically important. After this initial downturn, the HRQL of female household residents remained relatively stable until about age 70, and then declined to approximately 0.70 at age 80.

The results of Model 2, which included respondents in both households and institutions, were less optimistic than those of Model 1. After age 75

Figure 1
HUI 3 trajectories, by age, men aged 40 or older in 1994/1995, Canada excluding territories
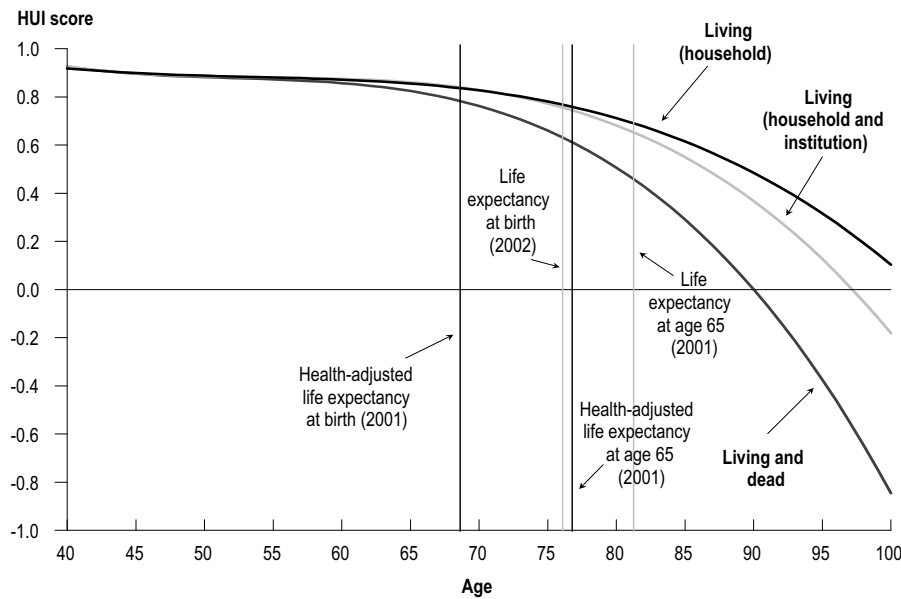
Source: 1994/1995 to 2004/2005 National Population Health Survey.

Figure 2
HUI 3 trajectories, by age, women aged 40 or older in 1994/1995, Canada excluding territories
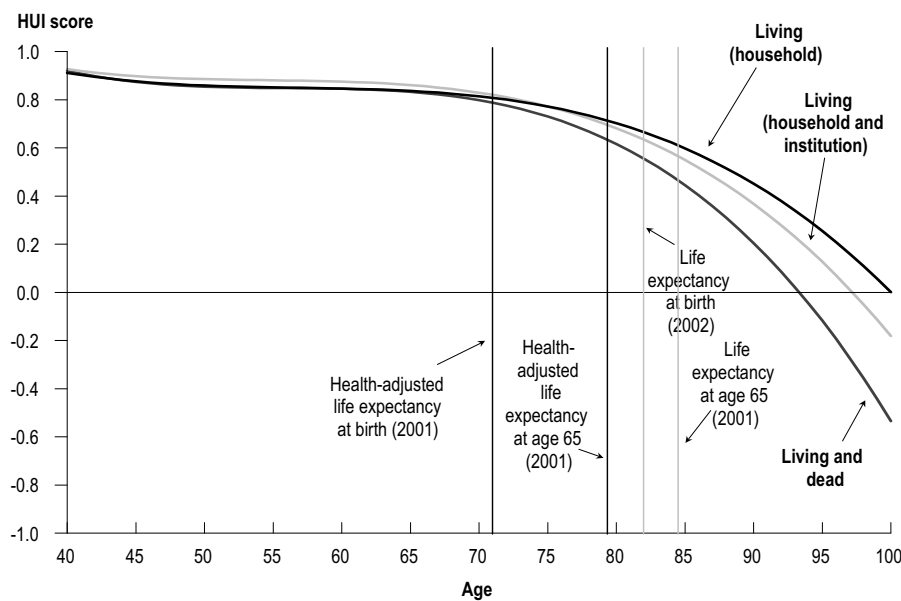
Source: 1994/1995 to 2004/2005 National Population Health Survey.

for men and after age 80 for women, HRQL declined more sharply in Model 2 than Model 1. This is not surprising, because the mean HUI3 score among the non-institutionalized sample was 0.83, while the mean HUI3 score among the institutionalized sample was 0.14, and institutionalization rates increase with age.

As well as residents of households and institutions, Model 3 included respondents who had died for the first cycle after their death, with a value of 0 for their HRQL at their age of death. This model showed a slow decline from scores close to 1 to around 0.8 for men and women until about age 70, after which HRQL declined more rapidly for men than for women. This reflects the effect of men's higher mortality at older ages. Taking death into account had a large effect on the growth curves: for men, HRQL was lower after age 60 when accounting for death compared with only institutionalization; for women, this discrepancy emerged around age 70.

Discussion

On average, HRQL remains relatively high from mid- to later life, suggesting that the older population is well and enjoying high quality of life in the years leading up to normal Canadian life expectancy. This is consistent with findings from international comparisons, which indicate that Canada ranks high on measures of life expectancy and disability-adjusted life years.[24,25]

Perhaps most important, this study demonstrates that excluding data for institutionalized individuals presents a biased view of the aging process, as does the failure to take mortality into account when describing the health of the population.[12] Inclusion of the institutionalized elderly results in less optimistic, but more accurate, estimates of population health. The effect of death on men's HRQL is greater than that on women's until quite late in life, reflecting men's earlier average mortality and women's additional years of life

What is already known on this subject?

- By global standards, the Canadian population is generally healthy, with long life expectancies for both men and women.
- Cross-sectional studies indicate that health and health-related quality of life tend to decline with age.

What does this study add?

- Canadians enjoy high health-related quality of life from mid-life into advanced ages.
- Men and women have similar trajectories of health-related quality of life with the exception of a notable decline among women in their 40s.
- Excluding data for institutionalized individuals and decedents presents an unduly optimistic view of the aging process.

lived in disability.[15,16] Either postponing mortality or reducing health problems should result in curves that maintain a higher level of HRQL longer through the lifespan. Future research can compare curves of successive cohorts to identify whether, when accounting for institutionalization and death, the health of the population is, indeed, improving.

The findings of the present analysis are consistent with some studies of self-rated health (SRH) trajectories, where SRH has been found to decline with advancing age,[8] but contrast with results from a study that showed an increase in SRH at older ages.[7] The results of the latter likely reflect a "survivor effect," whereby individuals with the worst SRH evaluations die and are thus removed from the analysis. Although we did not observe an increase in scores in older people, the difference between models 1 and 3 shows a "survivor effect."

The differences between the findings of the present study and those of others may reflect differences between HRQL as measured by the HUI3 and the use of SRH as the outcome measure. The HUI3 (in self-reports of health states) and SRH both contain a component of subjectivity, however at different levels. The results of this study highlight the importance of avoiding selection bias by following subjects into health care institutions and by including in the analyses those who die during the follow-up period.

Limitations

This study is based on self-reported health states that were transformed into a health utility score determined from societal preferences for different health states. Health states may not be accurately reported, and societal preferences for different health states may change over a long period, such as the 10 years covered in this study. Furthermore, the sample suffered attrition over time—in the last cycle analyzed, almost one in five respondents was lost to follow-up (and had not died or entered an institution). If the health status of the group lost to follow-up differed systematically from that of respondents who remained in the study, this could introduce bias into the results. However, respondents who dropped out were relatively similar to those who remained, and were younger. Because of the large number of younger participants in this study, the effect of attrition is likely to be diluted.

The results presented here are descriptive. The growth curve models employed are useful for describing patterns over time, but may not be appropriate for analyses aimed at explaining those patterns. Because age was centred on 40, the intercept is interpreted as the value of HUI for a 40-year-old. For explanatory models, other methods of centering age may be more appropriate and improve interpretability. Even so, analyses centering age on its mean value (57) did not differ substantially from the models presented here.

Conclusion

In summary, Canadians are, on average, quite healthy as they age from mid- to later life. While patterns of HRQL are similar for men and women, these deviate when institutionalization and death are considered. Furthermore, ignoring institutionalization and death portrays a healthier population than is actually the case, at least at advanced ages.

This research highlights the importance of moving beyond cross-sectional and household surveys in the study of successful aging. It also illustrates the need for data or surveys that include residents of institutions.

Future research should examine inter-individual variation in healthy aging, and focus on the predictors of successful aging as defined by HRQL. Such research will allow us to better understand aging in terms of health, broadly defined, and to identify ways in which policy and programs can promote healthy aging. ■

Funding and conflict of interest

References

1. Martel L, Malenfant E. *Portrait of the Canadian Population, 2006, by Age and Sex, 2006 Census.* (Statistics Canada, Catalogue 97-551-XIE) Ottawa: Statistics Canada, 2008.

2. Fries JF. Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 1980; 303: 130-6.

3. Fries JF. The compression of morbidity: Near or far? *Milbank Quarterly* 1989; 67(2): 208-32.

4. Jagger C. Compression or expansion of morbidity—what does the future hold? *Age and Ageing* 2000; 29: 93-4.

5. Singer JB, Willett JB. *Applied Longitudinal Data Analysis.* Toronto: Oxford University Press; 2003.

6. Hofer SM, Sliwinski MJ. Understanding aging: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology* 2001; 47: 341-52.

7. Liang J, Shaw BA, Krause N, et al. How does self-assessed health change with age? A study of older adults in Japan. *Journals of Gerontology: Series B* 2005; 60B(4): S224-32.

8. McCullogh ME, Laurenceau J-P. Gender and the natural history of self-rated health: a 59-year longitudinal study. *Health Psychology* 2004; 23(6): 651-5.

9. Tambay JL, Catilin G. Sample design of the National Population Health Survey. *Health Reports* (Statistics Canada, Catalogue 82-003) 1995; 7: 29-38.

10. Smith G. *The Population in Collective Dwellings: Canada 1961-1991.* (Statistics Canada, Catalogue 91F0015MIE1996002) Ottawa: Statistics Canada, 1996.

11. Diehr P, Patrick DL. Trajectories of health for older adults over time: Accounting fully for death. *Archives of Internal Medicine* 2003; 139: 416-20.

12. Diehr P, Johnson LL, Patrick DL, Psaty B. Methods for incorporating death into health-related variables in longitudinal studies. *Journal of Clinical Epidemiology* 2005; 58: 1115-24.

13. Belanger A, Martel L, Berthelot J-M, Wilkins R. Gender differences in disability-free life expectancy for selected risk factors and chronic conditions in Canada. *Journal of Women and Aging* 2002; 14(1/2): 61-83.

14. St-Arnaud J, Beaudet M, Tully P. Life expectancy. *Health Reports* (Statistics Canada, Catalogue 82-003) 2005; 17(1): 43-6.

15. Statistics Canada. Health-adjusted life expectancy, at birth and at age 65, by sex and income group, Canada and provinces, occasional. CANSIM Table 101-0121. Ottawa: Statistics Canada, 2001.

16. Gilmour H, Park J. Dependency, chronic conditions, and pain in seniors. *Health Reports* (Statistics Canada, Catalogue 82-003) 2005; 16(Suppl.): 21-31.

17. Rhodes AR. *Attrition in Longitudinal Studies Using Older Adults: A Meta-analysis.* Denton, Texas: University of North Texas, 2005.

18. Patrick DL, Erickson J. *Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation.* New York: Oxford University Press, 1993.

19. Feeny D, Furlong W, Torrance GW, set al. Multi-attribute and single-attribute utility functions for the Health Utilities Index Mark 3 System. *Medical Care* 2002; 40(2): 113-28.

20. *Mplus* [computer program]. Version 4.1. Los Angeles: Muthen & Muthen, 2006.

21. Rao J, Wu C, Yue K. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue 82-003) 1992; 18(2): 209-17.

22. Drummond M. Introducing economic and quality of life measurements into clinical studies. *Annals of Medicine* 2001; 33(5): 344-9.

23. Grootendorst P, Feeny D, Furlong W. Health Utilities Index Mark 3: Evidence of construct validity for stroke and arthritis in a population health survey. *Medical Care* 2000; 38(3): 290-9.

24. Mathers CD, Sadana R, Salomon JA, et al. *Estimates of DALE for 191 Countries: Methods and Results.* Geneva: World Health Organization, 2000.

25. World Heath Organization. *World Health Report: Statistical Annex.* Geneva: World Health Organization, 2006.

ELECTRONIC PUBLICATIONS AVAILABLE AT
www.statcan.ca

Medication use among senior Canadians

by Pamela L. Ramage-Morin

Abstract

Based on data from the 1996/1997 (institutional component) and 1998/1999 (household component) National Population Health Survey, seniors were major consumers of prescription medications, over-the-counter (OTC) products, and natural and alternative medicines. Almost all (97%) seniors living in long-term health care institutions were current medication users (medication use in the two days before their interview), as were 76% of those living in private households. Over half (53%) of seniors in institutions and 13% of those in private households used multiple medications (currently taking five or more different medications). Both medication and multiple medication use were associated with morbidity. Medications for the nervous system, the alimentary tract and metabolism, and the cardiovascular system were reported most frequently. Among seniors in institutions, those with Alzheimer's disease were less likely to take multiple medications than were those without this condition.

Keywords

aging, cross-sectional studies, morbidity, non-prescription drugs, polypharmacy, prescription drugs

Author

Pamela L. Ramage-Morin (613-951-1760, Pamela.Ramage-Morin@statcan.gc.ca) is with the Health Information and Research Division at Statistics Canada, Ottawa, Ontario, K1A 0T6.

Prescription medications, over-the-counter (OTC) products, and natural and alternative medicines are widely used in Canada, especially by seniors. But while medications play an important role in health care and disease management, their use is not without risk.[1]

Multiple medication use (prescription or OTC) is a common concern in relation to seniors' health.[2-5] In 2005, pharmacists dispensed an average of 35 prescriptions per person aged 60 to 79, and 74 prescriptions per person aged 80 or older, compared with an overall average of 14 prescriptions per Canadian.[6] People who take several medications at once are more likely to have adverse drug reactions; seniors are particularly vulnerable because of co-morbidity and physiological changes that come with age.[3,4,7-9]

There is a need in Canada for information about the number of seniors who use medications and those who have an elevated risk of drug-related adverse events from multiple medication use. The current study addresses these issues with data from nationally representative samples of seniors living in private households and in long-term health care institutions (see *The data*). In addition, the study reveals the most commonly reported types of medications and health-related factors associated with seniors' medication use.

Medication use

Nearly all residents of health care institutions were current medication users: 97% had taken some form of medication in the past two days (Table 1). As well, a large majority (76%) of seniors living in private households had done so.

Among the household population, women were more likely than men to have taken medications in the past two days, and seniors aged 75 to 84 were more likely to have done so than those aged 65 to 74. Differences based on educational attainment or the use of proxy reporters were not evident.

For seniors in institutions, the likelihood of taking medication did not differ by sex, age group, educational attainment, or proxy response.

Multiple medication use

Multiple medication use (taking five or more different drugs in the past two days) was reported for 53% of seniors in health care institutions and 13% of those in private households (Table 1, Figure 1). The likelihood of multiple

The data

Estimates of seniors' medication use were based on the latest cycles of the National Population Health Survey (NPHS) that collected information on medication use in the two days before their interview. Detailed documentation on the NPHS can be found on Statistics Canada's website (http://www.statcan.ca), and descriptions of the survey design, sample, and interview procedures are available in published reports.[10,11]

The NPHS *household file* covers household residents in all provinces, except persons living on Indian reserves, on Canadian forces bases, and in some remote areas. The data in this analysis are from 1998/1999 (cycle 3), which has a household response rate of 87.6% and a selected person response rate of 98.5%. Eleven percent of the senior household sample (317) relied on proxy reporters (Appendix Table A).

The NPHS *health institutions file* covers people living in hospitals, nursing homes, and facilities for people with disabilities. The data in this analysis are from 1996/1997 (cycle 2), which has institutional and individual response rates of 100% and 89.9%, respectively. Fifty-nine percent of the institutionalized respondents (1,013) relied on proxy reporters—49% were family members, and 10% were staff of the institutions.

Demographic distributions of the samples and populations used in this analysis are presented in Appendix Table A.

The primary outcome variables in this study are *medication use* and *multiple medication use. Medication use* refers to prescription and OTC medications including natural and alternative medicines. Household residents, who were usually interviewed by telephone, were asked to gather their medications and read the names from the containers. For institutionalized respondents, staff members of the institutions provided this information; these medications would all be classified as "prescribed," because seniors in institutions usually do not have the option of self-medicating. *Current users* were those who had taken medication in the two days before their interview.

The terms *multiple medication use* and *polypharmacy* are sometimes used interchangeably. The latter has been defined in the literature in relative terms (for example, the administration of an excessive number of drugs) and in absolute terms, ranging from two to more than six simultaneous medications.[5,12-15] In this study, preference is given to the term, *multiple medication use,* defined as *currently* taking five or more different medications. The threshold of five is relatively conservative alongside other absolute definitions of polypharmacy, and is consistent with an earlier Statistics Canada study based on the NPHS.[16]

NPHS respondents reporting current medication use were asked the names of their medications; data were recorded for a maximum of 12 medications. Drugs were listed in the order that they were reported, and so could not be ranked according to strength or importance. The drugs were coded using the Canadian edition of the *Anatomical Therapeutic Chemical (ATC) Classification System for Human Medications*.

Self-perceived health was based on the question, "In general, would you say your health is: ...." The five response categories were combined into two groups: good/very good/excellent health was defined as "positive" self-perceived health, and fair/poor health as "negative" self-perceived health.

Chronic pain was defined as a response of "no" to the question, "Are you usually free of pain or discomfort?"

The presence of *chronic conditions* was established by asking respondents if they had been diagnosed by a health professional with a long-term chronic health condition—one that had lasted, or was expected to last, at least six months. Respondents were read a list of conditions. Individual conditions included in this study were incontinence, arthritis, diabetes, heart disease, stroke, Alzheimer's disease or other dementia, and cataracts. For the institutional population, *incontinence* included urinary incontinence and difficulty controlling bowels, but for the household population, was limited to urinary incontinence. A more comprehensive list of chronic conditions was used to estimate the overall *number of chronic conditions* each respondent experienced (Appendix Table B). The count of chronic conditions was categorized into three groups: none or 1, 2, and 3 or more.

The analysis was based on independent samples from households and institutions. Data were weighted to reflect the age and sex distribution of the appropriate target populations. Weighted frequencies and cross-tabulations were used to estimate the proportion of people who had used medication/multiple medications in the past two days by selected characteristics. Logistic regression was used to model associations between indicators of ill health (chronic pain and number of chronic conditions) and multiple medication use while controlling for sex, age, education, and proxy reporting status. To account for survey design effects, standard errors and coefficients of variation were estimated with the bootstrap technique.[17-19]

The current study has a number of limitations. The data on institutions are from the 1996/1997 NPHS, whereas the household data are from the 1998/1999 NPHS. These surveys are the most recent from which multiple medication use can be established. The count of chronic conditions may vary between household and institutional residents, in part, because the lists of conditions were not identical in the two surveys (Appendix Table B). As well, chronic conditions were self-reported and were not verified by any other source.

Respondents who reported medication use in the past two days were limited to providing the names of 12 different drugs. Seven of the 2,851 household sample and 27 of the 1,711 institutional sample reported more than 12 different medications. For an additional 18 household and 73 institutional respondents, data on the number of different medications taken in the past two days were missing.

It is possible that respondents may not consider certain OTC products such as vitamins and natural/herbal products to be drugs, in which case the true number of medications taken would be under-reported.

A substantial share of the respondents—11% of the household sample and 59% of the institutional sample—relied on proxy reporters. However, excluding these respondents (the most seriously ill or cognitively impaired seniors) would have biased the results.

Table 1
Percentage using medications and multiple medications in past two days, by selected characteristics, household and institutional populations aged 65 or older, Canada excluding territories, 1998/1999 (households) and 1996/1997 (institutions)

| | Medication use | | | | | | Multiple medication use | | | | | |
| | Households | | | Institutions | | | Households | | | Institutions | | |
| | | 95% confidence interval | | | 95% confidence interval | | | 95% confidence interval | | | 95% confidence interval | |
	Percentage	from	to	Percentage	from	to	Percentage	from	to	Percentage	from	to
Total	**76.3**	**74.2**	**78.4**	**96.7**	**95.8**	**97.6**	**12.8**	**11.0**	**14.7**	**53.1**	**49.7**	**56.5**
Sex												
Men	71.9 *	68.8	74.9	96.5	94.9	98.1	11.6	9.2	14.0	51.3	45.9	56.7
Women†	79.7	77.0	82.4	96.8	95.7	97.8	13.8	11.5	16.2	53.8	49.8	57.8
Age group												
65 to 74†	73.4	70.4	76.3	97.2	95.3	99.2	10.5	8.5	12.5	52.6	45.8	59.5
75 to 84	80.8 *	77.4	84.2	96.9	95.3	98.4	16.3*	13.3	19.4	56.9	51.9	61.9
85 or older	77.9	71.2	84.5	96.4	95.2	97.6	14.4E	7.3	21.5	50.7	46.3	55.1
Education												
Less than secondary graduation†	76.9	74.1	79.8	96.7	95.6	97.7	14.2	11.6	16.7	54.5	50.6	58.5
Secondary graduation or more	75.7	72.7	78.7	97.0	95.5	98.5	11.6	9.4	13.8	51.1	44.9	57.2
Proxy reporter												
Yes	75.8	69.7	81.9	96.8	95.6	97.9	14.4	9.8	19.1	45.4*	41.0	49.8
No†	76.4	74.2	78.5	96.6	95.2	98.0	12.6	10.6	14.5	64.9	60.3	69.5

† reference category
* significantly different from estimate for reference category (p<0.05)
Source: National Population Health Survey, cross-sectional files, 1998/1999 (households) and 1996/1997 (institutions).

Figure 1
Percentage using medications in past two days, by number of medications, household and institutional populations aged 65 or older, Canada excluding territories, 1998/1999 (households) and 1996/1997 (institutions)
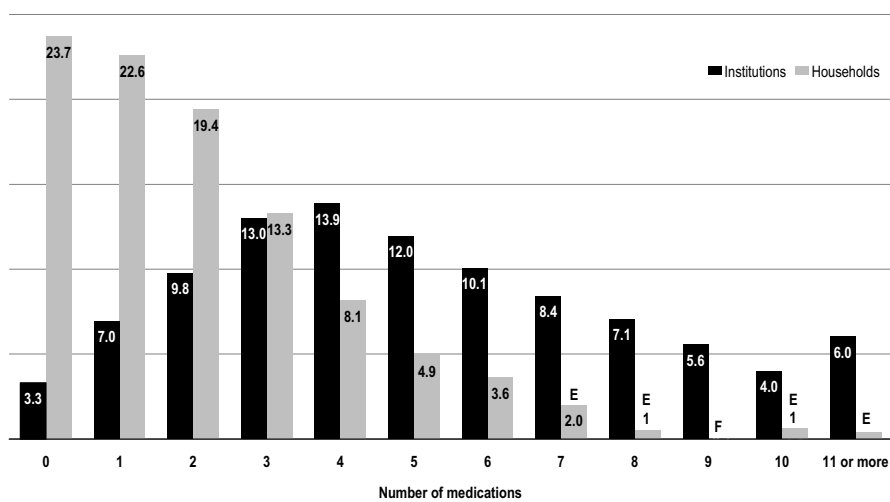
E use with caution (coefficient of variation 16.6% to 33.3%)
F too unreliable to be published (coefficient of variation greater than 33.3%)
Source: National Population Health Survey, cross-sectional files, 1998/1999 (households) and 1996/1997 (institutions).

medication use did not differ by sex, regardless of whether seniors lived in private households or in institutions.

Only among household residents was there a difference by age group—those aged 75 to 84 were more likely to be multiple medication users than were younger seniors (16 % compared with 11%). For institutionalized seniors, those who relied on proxy reporters were significantly less likely to report multiple medication use (45%) than were those who responded on their own behalf (65%).

Medication use, including the use of multiple medications, tended to be positively associated with indicators of ill health. For example, seniors who assessed their general health as poor were more likely to have used five or more different medications in the past two days than were those in better health (Table 2). As well, the number of chronic conditions reported, the presence of chronic pain, and having specific chronic conditions were

associated with medication use (Table 2). Even when sex, age, education, and proxy reporter were taken into account, the presence of chronic pain and a greater number of chronic conditions were each independently associated with higher odds of using multiple medications in the fully controlled model (Table 3).

A notable exception to the positive associations between chronic conditions and medication use emerged for institutionalized seniors who had Alzheimer's disease or other dementia (Table 2). They were less likely to be multiple medication users than were institutionalized seniors without this condition: 40% compared with 62%. This is the reverse of the association for seniors in households, among whom a higher proportion with Alzheimer's disease reported using multiple medications (34%), compared with those without this condition (13%). Among institutionalized seniors, the average

Table 2

Percentage using medications and multiple medication in past two days, by self-perceived health, chronic pain and other major chronic conditions, household and institutional populations, aged 65 or older, Canada excluding territories, 1998/1999 (households) and 1996/1997 (institutions)

	Medication use						Multiple medication use					
	Households			Institutions			Households			Institutions		
		95% confidence interval			95% confidence interval			95% confidence interval			95% confidence interval	
	Percentage	from	to	Percentage	from	to	Percentage	from	to	Percentage	from	to
Total	**76.3**	**74.2**	**78.4**	**96.7**	**95.8**	**97.6**	**12.8**	**11.0**	**14.7**	**53.1**	**49.7**	**56.5**
Poor self-perceived health												
Yes	91.1 *	88.2	94.0	97.5	96.5	98.5	32.2*	27.3	37.2	58.8*	54.6	63.0
No†	71.9	69.5	74.4	95.6	93.9	97.3	7.1	5.7	8.5	45.6	40.9	50.3
Chronic pain												
Yes	89.0 *	85.9	92.1	97.9*	96.8	98.9	23.6*	19.0	28.2	66.6*	61.4	71.8
No†	71.7	69.3	74.2	96.2	94.9	97.4	8.9	7.2	10.6	45.8	41.5	50.0
Incontinence												
Yes	87.5 *	81.8	93.3	97.0	95.8	98.2	26.6*	19.0	34.3	52.5	48.6	56.4
No†	75.4	73.2	77.5	96.3	94.8	97.7	11.7	9.8	13.7	54.0	49.2	58.7
Arthritis												
Yes	84.7 *	82.2	87.3	97.3	96.0	98.7	18.8*	15.8	21.9	62.2*	58.6	65.9
No†	69.4	66.6	72.3	96.1	94.9	97.4	8.0	6.1	9.9	45.3	40.9	49.6
Diabetes												
Yes	92.8 *	89.3	96.4	100.0*	100.0	100.0	31.6*	24.4	38.8	74.6*	69.7	79.5
No†	74.2	71.9	76.5	96.1	95.1	97.1	10.4	8.8	12.1	49.3	45.6	53.1
Heart disease												
Yes	97.9 *	96.6	99.2	98.3*	96.7	99.9	37.1*	31.3	42.9	67.7*	62.7	72.7
No†	71.8	69.4	74.2	96.1	95.0	97.2	7.8	6.3	9.3	47.7	43.6	51.9
Stroke												
Yes	93.4 *	87.6	99.3	99.0*	98.0	100.0	30.6*E	20.5	40.7	61.5*	56.7	66.2
No†	75.5	73.3	77.7	96.0	94.9	97.1	12.0	10.2	13.8	50.6	46.6	54.5
Alzheimer's disease or other dementia												
Yes	87.1	75.1	99.0	96.0	94.4	97.5	33.7*E	14.4	53.1	39.7*	34.9	44.4
No†	76.2	74.1	78.3	97.2	96.2	98.1	12.6	10.8	14.4	61.7	58.0	65.3
Cataracts												
Yes	85.4 *	81.3	89.5	96.4	94.8	98.0	17.0*	12.8	21.3	57.4*	51.8	63.0
No†	74.3	72.0	76.6	96.8	95.9	97.8	11.9	10.0	13.8	51.3	47.8	54.8
Number of chronic conditions												
None or one†	57.5	54.0	61.0	93.4	90.8	96.6	2.4E	1.3	3.4	43.0	35.7	50.4
Two	85.0 *	81.5	88.5	94.9	91.7	98.2	8.2*E	5.6	10.9	38.2	31.0	45.5
Three or more	93.9 *‡	91.9	95.8	97.9*	97.0	98.8	29.7*‡	25.6	33.8	58.6*‡	55.0	62.3

† reference category
* significantly different from estimate for reference category (p<0.05)
‡ significantly different from estimate for two chronic conditions (p<0.05)
E Use with caution (Coefficient of variation between 16.6% and 33.3%)
Note: "Incontinence" includes urinary and bowel incontinence for the institutional population, but is limited to urinary incontinence for the household population.
Source: National Population Health Survey, cross-sectional files, 1998/1999 (households) and 1996/1997 (institutions).

Statistics Canada, Catalogue no. 82-003-XPE • Health Reports, Vol. 20, no. 1, March 2009
Medication use among senior Canadians • Health Matters

41

age of those with and without Alzheimer's disease did not differ, although those with Alzheimer's disease had, on average, more chronic conditions (4.3 versus 3.4). Despite the tendency toward a greater number of chronic conditions, institutionalized seniors with Alzheimer's disease were less likely have chronic pain (28% versus 44%), less likely to be on pain medication (44% versus 58%), and received, on average, fewer medications overall than did institutionalized seniors without Alzheimer's disease or other dementia (4.4 medications versus 5.9).

Types of medications

The medications reported most commonly by seniors were those that act on the nervous system, the alimentary tract and metabolism, and the cardiovascular system (Table 4).

Almost four out of every five (78%) seniors in institutions and 37% of those in households took medications for the nervous system. Of these medications, analgesics were the most common, followed by psycholeptics, which include antipsychotics, anxiolytics, hypnotics and sedatives.

Medications for the alimentary tract and metabolism were used by 71% of seniors in institutions and 23% of those in households. Of these medications, laxatives were commonly reported for seniors in institutions (62%), followed by antacids, and drugs for the treatment of peptic ulcers and flatulence (34%). For seniors in households, peptic ulcer and flatulence medications were most common (40%), followed by drugs for diabetes (34%); laxatives were rarely reported (9%).

More than half (55%) of seniors in institutions and 44% of those in households reported taking cardiovascular medications. This group of medications is composed of diuretics, cardiac therapy drugs, antihypertensives, calcium channel blockers, and beta-blocking agents. Diuretics were most commonly reported for seniors in institutions, and

Table 3

Adjusted odds ratios relating multiple medication use to selected characteristics, household and institutional populations, aged 65 or older, Canada excluding territories, 1998/1999 (households) and 1996/1997 (institutions)

	Multiple medication use					
	Households			Institutions		
	Adjusted odds ratio	95% confidence interval		Adjusted odds ratio	95% confidence interval	
		from	to		from	to
Indicators of ill health						
Chronic pain						
Yes	1.8*	1.3	2.4	2.1*	1.5	3.0
No[†]	1.0	...	...	1.0	...	...
Number of chronic conditions						
None or one[†]	1.0	...	...	1.0	...	...
Two	3.4*	1.9	6.0	0.9	0.6	1.4
Three or more	14.8*	8.8	24.9	2.3*	1.5	3.4
Socio-demographic characteristics						
Sex						
Men	1.1	0.8	1.5	0.9	0.7	1.2
Women[†]	1.0	...	...	1.0	...	...
Age (continuous)	1.0	0.99	1.04	0.99	0.97	1.00
Education						
Less than secondary graduation[†]	1.0	...	...	1.0	...	...
Secondary graduation	0.9	0.7	1.3	0.9	0.7	1.2
Proxy reporter						
Yes	1.4	0.8	2.2	0.4*	0.3	0.6
No[†]	1.0	...	...	1.0	...	...
Model information						
Sample size	2,820			1,364		
Sample using multiple medications	365			747		
Records dropped because of missing values	31			347		

... not applicable
[†] reference category
* significantly different from estimate for reference category (p<0.05)
Source: National Population Health Survey, cross-sectional files, 1998/1999 (households) and 1996/1997 (institutions).

antihypertensives were prominent for seniors in private households.

Conclusion

This is the first nationwide, population-based study to provide benchmarks of medication and multiple medication use among all Canadian seniors, covering not only private households, but also long-term health care institutions. The stringent collection process for medication information minimizes the potential for recall bias: household residents were asked to read the names of their medications to the interviewer; staff members provided the information for institutionalized

respondents. This approach has an advantage over some administrative data in cases when discrepancies emerge between medications that are prescribed and those that are actually used.

Medication use by seniors is common—almost all seniors in institutions and over three-quarters of those in households reported using at least one medication in the past two days. Concurrent use of five or more medications was reported by 53% of seniors in institutions and 13% of those in households. This amounts to over a half million seniors taking multiple medications: approximately 94,000 in institutions and 445,000 in households.

Table 4
Prevalence of type of medication used in past two days, by anatomical system and sub-category, household and institutional populations aged 65 or older, Canada excluding territories, 1998/1999 (households) and 1996/1997 (institutions)

Anatomical system[†]	Sub-category[‡]	Households		Institutions	
		Estimated population	Prevalance	Estimated population	Prevalance
		thousands	%	thousands	%
Nervous system		**1,222**	**36.6**	**138**	**78.3**
	Analgesics	939	76.9	88	64.3
	Psycholeptics	276	22.6	76	55.3
	Psychoanaleptics	131	10.7	34	24.4
	Anti-Parkinson drugs	25[E]	2.1[E]	20	14.8
	Antiepileptics	65[E]	5.3[E]	16	11.6
Alimentary tract/metabolism		**781**	**23.5**	**124**	**70.8**
	Laxatives	74[E]	9.5[E]	78	62.5
	Antacids, drugs for peptic ulcer/flatulence	311	39.8	42	33.7
	Mineral supplements	123	15.7	22	17.8
	Vitamins	95	12.2	22	17.5
	Drugs used in diabetes	269	34.4	19	14.9
Cardiovascular system		**1,455**	**43.5**	**96**	**54.9**
	Diuretics	536	36.9	56	58.3
	Cardiac therapy	289	19.9	37	39.0
	Antihypertensives	653	44.9	29	30.6
	Calcium channel blockers	330	22.7	21	21.5
	Beta-blocking agents	377	25.9	9	9.5

† percent based on all respondents who reported taking at least one medication in past two days
‡ percent based on respondents who reported taking medications for specific anatomical system in previous two days
E use with caution (coefficient of variation between 16.6% and 33.3%)
Note: "Incontinence" includes urinary and bowel incontinence for the institutional population, but is limited to urinary incontinence for the household population.
Source: National Population Health Survey, cross-sectional files, 1998/1999 (households) and 1996/1997 (institutions).

Although medication use was generally associated with morbidity, this study identified unique circumstances for institutionalized seniors with Alzheimer's disease or other dementia. They tended to have more chronic conditions than institutionalized seniors without Alzheimer's disease, but were less likely to report pain, to have taken pain medications or to be multiple medication users. These findings are consistent with earlier studies that suggest higher levels of cognitive impairment are associated with undetected pain and subsequent under-treatment.[20-23]

Finally, people take several medications at once for many reasons including comorbidity, multiple prescribing physicians, inappropriate prescribing, access to different pharmacies, as well as self-medication with OTC and alternative products.[4,5,24,25] Regardless of the reason, those who take multiple medications have an elevated risk of adverse events. ■

References

1. Routledge PA, O'Mahony MS, Woodhouse KW. Adverse drug reactions in elderly patients. *British Journal of Clinical Pharmacology* 2003; 57(2): 121-6.

2. Courtman BJ, Stallings SB. Characterization of drug-related problems in elderly patients on admission to a medical ward. *The Canadian Journal of Hospital Pharmacy* 1995; 48(3): 161-6.

3. Patel RB. Polypharmacy and the elderly. *Journal of Infusion Nursing* 2003; 26(3): 166-9.

4. Wooten J, Galavis J. Polypharmacy: keeping the elderly safe. *RN* 2005; 68(8): 44-50.

5. Hippisley-Cox J, Pringle M, Ryan R. Polypharmacy in the elderly: Analysis of QRESEARCH data. QRESEARCH 2007. Available at: www.qresearch.org. Accessed 16 January, 2008.

6. IMS Health Canada. *Canada Rx Report 2005*. Table 3: Retail prescriptions dispensed in Canada per capita by age, 2005. Personal contact: Christine D. Albino, Senior Analyst, calbino@ca.imshealth.com.

7. Brazeau S. Polypharmacy and the elderly. *The Canadian Journal of CME*. August, 2001. Available at: www.stacommunications.com/journals/cme/images/cmepdf/august01/polypharmacyapvd.pdf. Accessed 16 January, 2008.

8. Williams CM. Using medications appropriately in older adults. *American Family Physician* 2002; 66(10): 1917-24.

9. Akner G, Boréus L. Drug treatment (Chapter 16). In: *Geriatric Care and Treatment: A Systematic Compilation of Existing Scientific Literature*. Place of publication: The Swedish Council on Technology Assessment in Health Care, 2003.

10. Swain L, Catlin G, Beaudet MP. The National Population Health Survey – its longitudinal nature. *Health Reports* (Statistics Canada, Catalogue 82-003) 1999; 10(4): 62-82.

11. Tambay J-L, Catlin G. Sample design of the National Population Health Survey. *Health Reports* (Statistics Canada, Catalogue 82-003) 1995; 7(1): 29-38.

12. World Health Organization (WHO) Centre for Health Development. *Ageing and Health Technical Report, Volume 5: A Glossary of Terms for Community Health Care and Services for Older Persons*, 2004. Available at: http://whqlibdoc.who.int/wkc/2004/WHO_WKC_Tech.Ser._04.2.pdf. Accessed 24 June 2008.

13. Frazier SC. Health outcomes and Polypharmacy in elderly individuals: an integrated literature review. *Journal of Gerontological Nursing* 2005; 31(9): 4-11.

14. Woods DJ, Walker T, Ross S. *Problems with Polypharmacy: Focus on the Elderly at Risk*. Dunedin, New Zealand: Best Practice Advocacy Centre, date of publication. Available at: http://cap.org.jm/News/polypharmacy_carib.pdf Accessed 24 June 2008.

15. Lococo KH, Staplin L. *Literature Review of Polypharmacy and Older Drivers: Identifying Strategies to Collect Drug Usage and Driving Functioning among Older Drivers.* (Report no. DOT HS 810 558) Place of publication: U.S. Department of Transportation, National Highway Traffic Safety Administration, 2006.

16. Millar W. Multiple medication use among seniors. *Health Reports* (Statistics Canada, Catalogue 82-003) 1998; 9(4): 11-7.

17. Rao JNK, Wu CFJ, Yue K. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 1992; 18(2): 209-17.

18. Rust KF, Rao JN. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 1996; 5(3): 283-310.

19. Yeo D, Mantel H, Liu T-P. Bootstrap variance estimation for the National Population Health Survey. *American Statistical Association: Proceedings of the Survey Research Methods Section.* Baltimore: August, 1999.

20. Victorino C, Maxwell C, Hogan D. The prevalence and pharmacological management of pain among older home care clients. *Senior Care Canada*. Available at: http://seniorcarecanada.com/articles/2003/q4/pain.management/. Accessed 1 October, 2008.

21. Proctor WR, Hirdes JP. Pain and cognitive status among nursing home residents in Canada. *Pain Reearch and Management* 2001; 6(3): 119-25.

22. Morrison RS, Siu AL. A comparison of pain and its treatment in advanced dementia and cognitively intact patients with hip fracture. *Journal of Pain and Symptom Management* 2000; 19(4): 240-8.

23. Sengstaken EA, King SA. The problems of pain and its detection among geriatric nursing home residents. *Journal of the American Geriatric Society* 1993; 41(5): 541-4.

24. Fugh-Berman A. Herb-drug interactions. *Lancet* 2000; 355(9198): 134-8.

25. Roehl B, Talati A, Parks S. Medication prescribing for older adults. *Annals of Long-term Care* 2006; 14(6): 33-9.

Table A
Sample size, estimated population and prevalence of selected characteristics, household and institutional populations aged 65 or older, Canada excluding territories, 1998/1999 (households) and 1996/1997 (institutions)

	Households			Institutions		
	Sample size	Estimated population	Prevalance	Sample size	Estimated population	Prevalance
		thousands	%		thousands	%
Total	2,851	3,488	100.0	1,711	185	100.0
Sex						
Men	1,108	1,518	43.5	457	50	27.2
Women	1,743	1,970	56.5	1,254	135	72.8
Age group						
65 to 74	1,539	1,999	57.3	261	28	15.2
75 to 84	1,037	1,222	35.0	603	64	34.8
85 or older	275	266	7.6	847	93	50.0
Education						
Less than secondary graduation	1,473	1,748	50.2	1,104	120	69.7
Secondary graduation or more	1,374	1,734	49.8	481	52	30.3
Proxy reporter						
Yes	317	522	15.0	1,013	111	60.0
No	2,534	2,966	85.0	698	74	40.0
Current medication use						
Yes	2,181	2,642	76.3	1,582	171	96.7
No	652	821	23.7	56	6	3.3
Multiple medication use						
Yes	367	445	12.8	876	94	53.1
No	2,466	3,018	87.2	762	83	46.9

Note: Detail may not add to totals because of missing values for some variables.
Source: National Population Health Survey, cross-sectional files, 1998/1999 (households) and 1996/1997 (institutions).

Table B
Chronic diseases used to estimate overall *Number of chronic conditions* for household and health institutions components of National Population Health Survey, 1998/1999 (households) and 1996/1997 (institutions)

	Households	Institutions
Asthma	√	√
Arthritis or rheumatism	√	√
Back problems excluding arthritis	√	...
High blood pressure	√	√
Chronic bronchitis or emphysema	√	√
Diabetes	√	√
Epilepsy	√	√
Heart disease	√	√
Cancer	√	...
Stomach or intestinal ulcers	√	√
Suffers from effects of stroke	√	√
Urinary incontinence	√	√
Difficulty controlling bowels	...	√
Bowel disorder/Crohn's Disease or colitis	√	√
Alzheimer's disease or other dementia	√	√
Cataracts	√	√
Glaucoma	√	√
Thyroid condition	√	√
Suffers from partial or complete paralysis	...	√
Osteoporosis or brittle bones	...	√
Kidney failure or disease	...	√
Other chronic condition	√	√

... not available
Note: For health institutions, "Difficulty controlling bowels" and/or urinary incontinence was counted as one chronic condition.
Source: National Population Health Survey, cross-sectional files, 1998/1999 (households) and 1996/1997 (institutions).

Evaluation of the coverage of linked Canadian Community Health Survey and hospital inpatient records

by Michelle Rotermann

Abstract

Background
Evaluation of the coverage that results from linking routinely collected administrative hospital data with survey data is an important preliminary to undertaking analyses based on the linked file.

Data and methods
To evaluate the coverage of the linkage between data from cycle 1.1 of the Canadian Community Health Survey (CCHS) and in-patient hospital data (Health Person-Oriented Information or HPOI), the number of people admitted to hospital according to HPOI was compared with the weighted estimate for CCHS respondents who were successfully linked to HPOI. Differences between HPOI and the linked and weighted CCHS estimate indicated linkage failure and/or undercoverage.

Results
According to HPOI, from September 2000 through November 2001, 1,572,343 people (outside Quebec) aged 12 or older were hospitalized. Weighted estimates from the linked CCHS, adjusted for agreement to link and plausible health number, were 7.7% lower. Coverage rates were similar for males and females. Provincial rates did not differ from those for the rest of Canada, although differences were apparent for the territories. Coverage rates were significantly lower among people aged 75 or older than among those aged 12 to 74.

Keywords
coverage, data collection, databases, health surveys, hospital records, inpatients, medical records, probabilistic linkage

Author

Michelle Rotermann (1-613-951-3166; Michelle.Roterman@statcan.gc.ca) is with the Health Information and Research Division at Statistics Canada, Ottawa, Ontario, K1A 0T6

Record linkage is used in health studies to obtain more complete information, to fill gaps in existing datasets, and/or to improve data quality.[1,2] For instance, prospective death clearance of survey respondents, study cohorts or administrative data sources, such as inpatient hospital records, have made it possible to study associations between death and factors such as lifestyle, occupation, treatment modalities, patient histories and geography.[3-20]

Similarly, linking birth and stillbirth records with death registrations and/or hospitalization data has enabled the study of maternal, fetal and infant morbidity and mortality by maternal and infant characteristics.[21-24] Record linkage has also been used to validate self-reported information,[25,26] describe the characteristics of unmatched records,[27] assess the comparability or quality of data files generated using probabilistic and deterministic linkage approaches,[28] reduce under-ascertainment of disease prevalence,[29] and monitor health system performance.[30,31] In the absence of disease registries, record linkage is a cost-effective and efficient way to monitor disease incidence and prevalence.[32-35]

This study was motivated by the need to assess the coverage of the linkage between the Canadian Community Health Survey (CCHS) and Health Person-Oriented Information (HPOI), *an administrative database of hospital records*. Initial research on the rate of linkage between the CCHS and HPOI estimated the proportion of CCHS respondents who had been hospitalized during the 1994/1995 to 2004/2005 period, but coverage has yet to be assessed.[36] Evaluation of the coverage is essential if the linked file is to be used for epidemiologic research. It is important to know if findings will be biased, that is, if survey respondents with certain characteristics are more likely than others to have been linked.

HPOI and the CCHS are complementary sources of data. HPOI does not have information about non-medical determinants of health, such as socio-economic and lifestyle factors.

For example, hospital records do not contain information about smoking status or body mass index (BMI), two important risk factors. The CCHS, by contrast, is a rich source of information about health status and determinants of health, but lacks the detail needed to study hospitalization. Combining HPOI with the CCHS reduces many of the limitations of each source, and thereby facilitates a more complete understanding of what brings Canadians in contact with the health care system and how they fare within the system.

The two main objectives of this study were to: 1) evaluate the coverage of the linked CCHS and HPOI by calculating coverage rates; and 2) identify characteristics of CCHS cycle 1.1 respondents who were less likely to be in the linked file.

Methods

Data sources

Canadian Community Health Survey

The Canadian Community Health Survey is a cross-sectional survey that collects information about health status, health care use and health determinants. It covers the household population aged 12 or older in the provinces and territories, except members of the regular Forces and residents of institutions, Indian reserves and other Aboriginal settlements, and some remote areas. The rate of coverage is in the 98% range in the provinces, 97% in the Northwest Territories, 90% in the Yukon, and 71% in Nunavut.

Data for cycle 1.1 were collected from September 1, 2000 through November 3, 2001 from a sample of 131,535 people; the response rate was 84.7%. All CCHS information, including provincial health care numbers (HNs) and postal codes, is self-reported by respondents, and the extent of error in these variables is unknown. However, data capture applications used by interviewers contain features that check for inconsistent answers, out-of-range responses or invalid alpha-numeric sequences. More information about the CCHS is available in a published report.[37]

CCHS respondents were asked for permission to link information collected during the interview with their provincial health information, including past and continuing use of services such as hospitals, clinics, doctor's offices or other services provided by the province; 91% of respondents gave permission. The sample used for this study consists of 72,354 (66.5%) respondents aged 12 or older in all provinces and territories except Quebec, who agreed to link and provided a valid health number (HN) (Appendix A). Quebec HPOI records cannot be linked to CCHS records because the Quebec hospital records provided to Statistics Canada contain scrambled HNs, no date of birth and incomplete postal codes.

Survey weights were used so that estimates produced from the CCHS data were representative of the target population, not just the sample itself. The survey weight is the number of people in the population represented by each respondent. Survey weights reflect the differing probabilities of selection and response. Each record is, therefore, weighted by the inverse of the probability of selecting the person and getting a response from him or her.[38] Additional survey weights are required for record linkage because not all respondents agree to link and not all those who agree to link, provide a valid HN. For this study, survey weights, adjusted for agreement to link and provision of a valid HN, were calculated.

Statistics Canada does not have access to provincial health insurance databases against which the HNs provided by CCHS respondents could be verified. Instead, all provinces and territories provide check-digit formulas that are used to verify that the HNs are at least plausible. Although check-digits are not a substitute for databases that contain first and last names, birth dates, addresses and HNs, they can detect accidental transcription errors, such as the inversion of two numbers, and offer a simple method of distinguishing meaningful numbers from strings of random digits.

Hospital data

The Health Person-Oriented Information (HPOI) database, maintained by Statistics Canada, contains information about inpatient hospital separations (discharges and in-hospital deaths) from virtually all acute-care and some psychiatric, chronic and rehabilitative hospitals.

HPOI is a person-level dataset derived from discharge records (which can reflect multiple discharges of the same person) in the Hospital Morbidity Database (HMDB). Sequential person-level HPOI records can be used to construct each patient's hospitalization history. During the linkage process, records belonging to the same individual are identified from the patient's HN and demographic and diagnosis/intervention information (for example, sex, birth date, sex-specific procedures).[39]

Hospital records pertaining to the past fiscal year are added to HPOI annually. With each additional year of data, the entire HPOI process is rerun to ensure internal consistency of the demographic information at the person-level for patients with multiple hospital discharges.

Reabstraction studies, which validate the accuracy of hospital records, have found that the non-medical administrative data elements (essential for record linkage) are of high quality. For example, 99% of a random sample of discharge records for hospital stays from September through November 2000 had correct HNs, and 91% of postal codes were error-free.[40]

Statistics Canada has hospital data with HNs for all provinces (except Quebec) and the Northwest Territories from fiscal year 1994/1995 onwards;

data for 1992/1993 and 1993/1994 are available for some provinces.

While the HPOI database includes the vast majority of records from HMDB, about 3% of records for patients aged 12 or older (the target population of this study) were excluded because of missing or invalid HNs.[39]

From September 1, 2000 through November 3, 2001, there were 2.3 million discharges of 1,624,972 people aged 12 or older from acute-care hospitals outside Quebec. Discharges from non-acute hospitals were excluded from this study because coverage of such hospitals is inconsistent across provinces.

The target populations of the Canadian Community Health Survey (CCHS) and HPOI differ somewhat. The CCHS excludes full-time members of the Canadian Forces and residents of Indian Reserves, of institutions (for instance, nursing homes and prisons) and of some remote areas. HPOI is a census and, therefore, these groups are included among hospitalizations. In an effort to match the target populations of the CCHS and HPOI more closely, hospitalizations that could be identified as pertaining to the on-reserve or the institutionalized population were removed from this analysis.

The on-reserve population is a derived census variable created by identifying census sub-division (CSD) type according to criteria established by Indian and Northern Affairs Canada (INAC), as well as selected CSDs that correspond to northern communities in Saskatchewan, the Northwest Territories, and the Yukon.[41] The postal code conversion file (PCCF+)[42] and a list of facilities used by the Residential Care Facility survey[43] were used to identify institutional residents. Hospitalizations pertaining to 31,330 residents of Reserves and associated lands were removed from HPOI, as were hospitalizations of 21,299 residents of institutions. Removal of these 52,629 records, which amounted to about 3% of the HPOI patients hospitalized during the study period, brought the population covered by HPOI more in line with the CCHS target population.

Analytical techniques

Probabilistic record linkage
Probabilistic record linkage was used to identify CCHS respondents who were hospitalized. The linkage between the CCHS and HPOI was done with Generalized Record Linkage software (GRLS) developed at Statistics Canada. The two data sources contain many variables, but only a few fields appear in both and are distinct enough to be useful in matching for linkage. A CCHS respondent was considered to have been hospitalized if a record containing an HN and/or similar demographic characteristics (for example, birth date, sex, postal code) and an admission date to an acute-care facility between September 1, 2000 and November 3, 2001 was found in HPOI.

Probabilistic linkage does not require complete agreement on the matching variables. Rather, the quality of the match between pairs of records is rated with algorithms that evaluate the likelihood of a correct match[1,44] (Figure 1). Points were given or subtracted depending on the similarity of the values between fields. For instance, high positive scores were assigned if the HNs were identical and the issuing province of the HN matched; if the values were similar but not exact, a lower positive score was assigned, reflecting partial agreement; if the values

Figure 1
Example of how pairs of Canadian Community Health Survey (CCHS) and Health Person-oriented Information (HPOI) records were assessed and scored using Generalized Record Linkage Software (GRLS)

ID	Province	Birthdate	Postal code	Health number (HN)	Sex	Score calculated by GRLS	Match?	Commentary
A	Ontario	11/06/1964	L9Y3B9	3512345678	Female			
1	Ontario	11/06/1964	L9Y3B9	3512345678	Female	447	yes	All fields match
F	Manitoba	24/07/1927	R0A0T0	55667788	Male			
1	Ontario	11/06/1964	L9Y3B9	3512345678	Female	-308	no	Nothing matches
B	Manitoba	21/05/1945	R0A0T0	missing	Male			
7	Manitoba	21/05/1945	R0A0T0	4624252627	Male	268	yes	HN missing; everything else matches
B	Manitoba	21/05/1945	R0A0T0	missing	Male			
1	Ontario	11/06/1964	L9Y3B9	3512345678	Female	-244	no	HN missing; nothing else matches

◻ CCHS record   ◻ HPOI record

on the two records were totally different, points were subtracted.

The number of points assigned to each pair of linking variables reflected their importance as matching variables, which typically was related to uniqueness. For example, because there are only two possible values for the sex of the respondent/patient, matches on this field scored fewer points than if the postal codes or HNs matched.

Total linkage weights for each pair of CCHS-HPOI records were calculated by summing the scores assigned to each pair of linking variables. The higher the total linkage weight, the more likely the two records pertained to the same individual. Total linkage weights ideally form a bi-modal distribution. When pairs of records scored above the selected threshold, they were accepted as "true" matches; pairs below the threshold were rejected. To eliminate the need for manual review, the cut-off points chosen for this study were identical, which meant that each pair of records could have only one of two values: match or non-match.

Results

To evaluate the coverage of the linkage between cycle 1.1 of the CCHS and HPOI, the number of people admitted to hospital according to each data source was compared. Survey weights, adjusted for agreement to link and HN validity, were applied to the records of CCHS respondents for whom records were also found in the HPOI database. The HPOI count of hospitalizations was regarded as the standard. The coverage rate was calculated by dividing the weighted estimates of CCHS respondents who successfully linked to HPOI by HPOI counts, minus records identified as pertaining to residents of Indian Reserves or associated lands or of institutions and then multiplying by 100. Differences between the HPOI counts and the weighted estimates from the CCHS were examined. Standard errors and 95% confidence intervals were calculated for the coverage rates

Table 1

Number hospitalized in acute-care hospitals and coverage rates, September 1, 2000 through November 3, 2001, by selected characteristics and data source, population aged 12 or older, Canada excluding Quebec

| | Health Person-Oriented Information (HPOI) | Canadian Community Health Survey (CCHS) | | CCHS/HPOI Coverage rate | | |
| | | | | | 95% confidence interval | |
	Number	Unweighted number	Weighted number	%	from	to
Total	1,572,343	6,785	1,451,272	92.3	88.9	95.7
Province/Territories						
Newfoundland and Labrador	41,394	272	40,445	97.7	83.6	111.8
Prince Edward Island	11,784	237	11,061	93.9	79.6	108.1
Nova Scotia	67,226	348	60,419	89.9	78.0	101.7
New Brunswick	67,542	423	62,203	92.1	81.7	102.5
Ontario	753,970	2,230	694,463	92.1	86.6	97.6
Manitoba	82,386	567	69,739	84.6	73.6	95.7
Saskatchewan	82,778	659	78,664	95.0	86.4	103.7
Alberta	202,498	863	186,301	92.0	83.3	100.7
British Columbia	258,883	1,062	241,647	93.3	85.3	101.3
Territories	3,882	124	6,331	163.1*	139.3	186.9
Sex						
Female†	971,087	4,343	904,318	93.1	88.8	97.5
Male	601,249	2,442	546,955	91.0	85.4	96.5
Age group						
12 to 24†	173,009	711	165,093	95.4	81.5	101.3
25 to 34	276,150	1,041	270,164	97.8	85.0	105.8
35 to 44	210,848	830	209,447	99.3	88.3	110.4
45 to 54	183,878	773	174,194	94.7	84.9	104.6
55 to 64	181,041	854	169,235	93.5	83.2	103.7
65 to 74	227,410	1,090	219,259	96.4	86.9	106.0
12 to 74	1,252,336	5,299	1,207,392	96.4	92.4	100.4
75 or older	320,007	1,486	243,881	76.2*	70.2	82.2

† reference category

* significantly different from reference category (p<0.05); for provincial comparison, significantly different from rest of Canada, for example, Ontario compared with Canada minus Ontario

Source: 2000/2001 Canadian Community Health Survey; Health Person-oriented Information, 2000/2001 to 2001/2002.

using the bootstrap technique. Statistical significance was tested using the t-test ($p<0.05$).[45,46]

According to HPOI, from September 1, 2000 through November 3, 2001, 1,572,343 people were admitted to an acute-care hospital (excluding Quebec) (Table 1). Weighted estimates from the CCHS, adjusted for agreement to link and valid HN, were 7.7% lower (1,451,272).

Coverage rates were similar for males and females (91.0% and 93.1%). Provincial rates did not differ significantly from the rate for the rest of Canada. However, based on the CCHS, the estimated number of residents of the territories who were hospitalized was considerably higher than the HPOI

number. As a result, the coverage rate for the territories exceeded 100%.

Coverage rates for most age groups were similar. The exception was seniors aged 75 or older whose rate (76.2%) was significantly lower than that of people aged 12 to 74 (96.4%).

Discussion

The significantly lower coverage rate for seniors aged 75 or older was anticipated because the two data sources did not pertain to exactly the same populations. The CCHS excludes residents of institutions, but they are included in the hospital data (HPOI). Institutionalization is considerably more common among seniors than among

younger people: overall, fewer than 2% of Canadians live in an institution, but at age 75 or older, the figure is 16%.[47]

In the absence of direct information in HPOI records about patients' place of residence, the postal code in combination with the PCCF+ and the Residential Care Facilities list was used to determine if patients lived in an institution.   More than 20,000 institutional residents were identified and subsequently removed from HPOI using the PCCF+. Nonetheless, the coverage rate for seniors aged 75 or older remained significantly below the rates for younger people.

Use of the PCCF+ and the Residential Care Facilities list to identify institutions based only on the postal code is not ideal. Institutions that accounted for the majority of the population sharing a postal code had a higher chance of being identified and subsequently removed from the HPOI counts. As well, institutions in urban areas have more precise postal codes, and therefore, residents of such institutions were more likely to have been removed from HPOI. Rural and outlying suburban areas and smaller towns often have the same postal code for multiple enumeration/ dissemination areas. Consequently, the coding is far less precise than for

centralized urban postal codes, which are usually linked to a single enumeration/dissemination area. Therefore, residents of institutions in rural and outlying suburban areas and smaller towns likely remained in the HPOI counts.

The coverage rate in the territories is also problematic, in that the linked CCHS-HPOI estimates exceeded the standard (HPOI). This, however, is less of a concern, because the small number of CCHS records linking to HPOI (124) precludes future analyses featuring this subpopulation. Before the removal of on-reserve residents from the HPOI count, the coverage rate for the territories was 113%; after their removal, the rate was 163%. It is unclear why the linked HPOI-CCHS estimate is so much higher than HPOI. Records of CCHS respondents identified as living in the territories were reviewed to determine if some had high survey weights, which might explain the discrepancy between the HPOI and HPOI-CCHS counts. No discrepant weights were found; the average weight was 51, with weights ranging in value from 11 to 178.

In   addition,   hospitalizations pertaining to military personnel could not be identified and removed from HPOI. Full-time members of the Armed

Forces are excluded from CCHS, and their inclusion may affect the coverage rate.

Conclusion

The value of record linkage is well established in epidemiological studies of population health.   Linking information from routinely collected administrative health data such as HPOI with survey data like the CCCHS holds promise for discoveries about health determinants, different types of health care use and health outcomes. Coverage evaluation is a fundamental pre-requisite to analyses that integrate health-related information from multiple sources based on the CCHS-HPOI linked file.

This evaluation shows that the overall coverage rate is high, often over 90%, although some CCHS respondents, notably seniors, had lower rates. Even this limitation is manageable, however, as long as users of the file explicitly acknowledge that findings pertain only to the general household population (the target population of the CCHS), and not to the total population, particularly residents of institutions. ◼

References

1. Fair ME, Whitridge P. *Tutorial on Record Linkage Slides Presentation.* 1997 Federal Committee on Statistical Methodology. Available at http://www.fcsm.gov/working-papers/RL_chap12.html.

2. Fair M. Generalized record linkage system-Statistics Canada's Record linkage software. *Austrian Journal of Statistics* 2004; 33(1 and 2): 37-53.

3. Villeneuve PJ, Morrison HI, Craig CL, et al. Physical activity, physical fitness, and risk of dying. *Epidemiology* 1998; 9(6): 626-31.

4. Weller I, Corey P. The impact of excluding non-leisure energy expenditure on the relation between physical activity and mortality in women. *Epidemiology* 1998; 9(6): 632-5.

5. Bata IR, Gregor RD, Eastwood BJ, et al. Trends in the incidence of acute myocardial infarction between 1984 and 1993 - The Halifax County MONICA Project. *Canadian Journal of Cardiology* 2000; 16(5): 589-95.

6. Roos NP, Havens B. Predictors of successful aging: a twelve-year study of Manitoba elderly. *American Journal of Public Health* 1991; 81(1): 63-8.

7. Khaw KT, Wareham N, Bingman S, et al. Combined impact of health behaviors and mortality in men and women: the EPIC-Norfolk prospective population study. *PLOS Medicine* 2008. Available at: http://medicine.plosjournals.org/perlserv/?request=get-document&10.1371/journal.pmed.0050012.

8. Weston TL, Aronson KJ, Howe GR, et al. Cancer mortality among males in relation to exposures assessed through a job-exposure matrix. *International Journal of Occupational and Environmental Health* 2000; 6(3): 194-202.

9. Aronson KJ, Howe GR, Carpenter M, et al. Surveillance of potential associations between occupations and causes of death in Canada, 1965-91. *Occupational and Environmental Medicine* 1999; 56(4): 256-69.

10. Wei L, Lang CC, Sullivan FM, et al. Impact on mortality following first acute myocardial infarction of distance between home and hospital: Cohort study. Available at: http://www.heart.bmj.com/onlinefirst.dtl.

11. Schmaltz HN, Southern D, Ghali WA, et al. Living alone, patient sex and mortality after acute myocardial infarction. *Journal of General Internal Medicine* 2007; 22(5): 572-8.

12. Ellison LF, Gibbons L, Canadian Cancer Survival Analysis Group. Five-year survival from prostate, breast, colorectal and lung cancer. *Health Reports* (Statistics Canada, Catalogue 82-003) 2001; 13(1): 23-34.

13. Ellison LF, Gibbons L. Leading cancers-changes in five-year relative survival. *Health Reports* (Statistics Canada, Catalogue 82-003) 2004; 15(2): 19-32.

14. Kaul P, Armstrong PW, Change W, et al. Long-term mortality of patients with acute myocardial infarction in the United States and Canada: Comparison of patients enrolled in global utilization of streptokinase and t-PA for occluded coronary arteries (GUSTO)-I. *Circulation* 2004; 110: 1754-60. Available at: http://circ.ahajournals.org/cgi/content/full/110/13/1754

15. Fischbacher CM, Bhopal R, Povey C, et al. Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: myocardial infarction in South Asians. *BMC Public Health* 2007; 7: 142

16. Jebamani LS, Burchill CA, Martens PJ. Links using data linkage to identify First Nations Manitobans: technical, ethical, and political issues. *Canadian Journal of Public Health* 2005; 96(Suppl.): S23-32.

17. Knoops KT, de Groot LC, Kromhout D, et al. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. *The Journal of the American Medical Association 2004;* 292: 1433-9.

18. Myint PK, Welch AA, Bingham SA et al. Fruit and vegetable consumption and self-reported functional health in men and women in the European Prospective Investigation into Cancer-Norfolk (EPIC-Norfold): a population-based cross-sectional study. *Public Health Nutrition 2007*; 10: 34-41.

19. Schnatter, A Robert. An analysis of death ascertainment and follow-up through Statistics Canada's Mortality Data Base System. *Canadian Journal of Public Health* 1990; 81(1): 60-5.

20. Schoenman JA, Sutton JP, Elixhauser A, et al. Understanding and enhancing the value of hospital discharge data. *Medical Care Research and Review* 2007; 64(4): 449-68.

21. Herrchen B, Gould JB, Nesbitt TS. Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies. *Computing Biomedical Research* 1997; 30(4): 290-305.

22. Chen J, Fair M, Wilkins R, et al. Maternal education and fetal and infant mortality in Quebec. *Health Reports* (Statistics Canada, Catalogue 82-003) 18; 10(2): 53-64.

23. Wen SW, Kramer MS, Liu S, et al. Infant mortality by gestational age and birth weight in Canadian provinces and territories, 1990-1994 births. *Chronic Diseases in Canada* 2000; 21(1): 14-22.

24. Kramer MS, Demissie K, Yang H, et al. The contribution of mild and moderate preterm birth to infant mortality. *Journal of the American Medical Association* 2000; 284(7): 843-9.

25. Bingham SA, Cassidy M, Cole TJ, et al. Validation of weighed records and other methods of dietary assessment using the 24h urine nitrogen technique and other biological markers. *British Journal of Nutrition* 1995; 73: 531-50.

26. Quigley MA, Hockley C, Davidson LL. Agreement between hospital records and maternal recall of mode of delivery: evidence from 12 391 deliveries in the UK Millennium Cohort Study. *International Journal of Obstetrics and Gynaecology* 2007; 114(2): 195-200.

27. Ford JB, Herrchen B, Gould JB, et al. Characteristics of unmatched maternal and baby records in linked birth and hospital discharge data. *Health Statistics Quarterly* 2007; 33: 25-33.

28. Fair M, Cyr M, Allen A, et al. *Validation Study for a Record Linkage of Births and Infant Deaths in Canada* (Statistics Canada, Catalogue 84-F0013) Ottawa: Statistics Canada, 1999.

29. Manuel DG, Schultz SE. Using linked data to calculate summary measures of population health: Health-adjusted life expectancy of people with Diabetes Mellitus. *Population Health Metrics* 2004. Available at: http://www.pophealthmetrics.com/content/2/1/4. Accessed: May 13, 2008.

30. Calver J, Brameld KJ, Preen DB, et al. High-cost users of hospital beds in Western Australia: a population-based record linkage study. *Medical Journal of Australia* 2006; 184(8): 393-7.

31. Houle C, Berthelot JM, David P, et al. Matching census database and Manitoba health care files. *Analytical Branch Studies Research Paper Series* (Statistics Canada, Catalogue 11F0019MPE) No. 91. Available at: http://www.fcsm.gov/working-papers/choule.pdf.

32. Brocco S, Visentin C, Fedeli U, et al. Monitoring the occurrence of diabetes mellitus and its major complications: the combined use of different administrative databases. *Cardiovascular Dialectology* 2007;15(6): 5.

33. Young TK, Kliewer E, Blanchard J, et al. Monitoring disease burden and preventive behavior with data linkage: cervical cancer among aboriginal people in Manitoba, Canada. *American Journal of Public Health* 2000; 90(9): 1466-8.

34. Ringland C, Correll PK, Lim K,et al. Hospital readmissions for asthma: a feasibility study comparing strategies for linking hospital morbidity data. *Australian and New Zealand Journal of Public Health* 2006; 30(5): 435-9.

35. Aronson KJ, Howe GR, Carpenter M, et al. *Occupational Surveillance in Canada: Cause-specific Mortality among Workers, 1965-1991* (Statistics Canada, Catalogue 84-546) Ottawa: Minister of Industry, 2000.

36. Nadeau C, Beaudet MP, Marion J. Deterministic and probabilistic record linkage. *Proceedings of Statistics Canada Symposium-Gatineau.* Ottawa: Minister of Industry, 2006.

37. Béland Y. Canadian Community Health Survey -methodological overview. *Health Reports* (Statistics Canada, Catalogue 82-003) 2002; 13(3): 9-14.

38. Statistics Canada. *Canadian Community Health Survey—Estimation.* Available at: www.statcan.ca/cgi-in/imdb/p2SV.pl?Function=getSurvey&SDDS=3226&lang=en&db=IMDB&dbg=f&adm=8&dis=2#b6. Accessed June 27, 2008

39. Statistics Canada, Household Surveys Methodology Division. External linkage production report: Data years: F1992 to F2004. (Unpublished), 2006.

40. Richards J, Brown A, Homan C. The data quality study of the Canadian Discharge Abstract Database. *Proceedings of Statistics Canada Symposium 2001. Achieving Data Quality in a Statistical Agency: A Methodological Perspective.* Available at: www.statcan.ca/english/freepub/11-522-XIE/2001001/session16/s16a.pdf. Accessed: July 2, 2008.

41. Statistics Canada. *2001 Census Dictionary* (Statistics Canada, Catalogue 92-378). Available at: www12.statcan.ca/english/census01/Products/reference/dict/appendices/92-378-XIE2002.pdf. Accessed June 26, 2008.

42. Statistics Canada. *Postal Code Conversion File (PCCF): Update.* Available at : www.statcan.ca/bsolc/english/bsolc?catno=92-153-UCB. Accessed June 28, 2008.

43. Statistics Canada. *Residential Care Facilities (RCF).* Available at: www.statcan.ca/bsolc/english/bsolc?catno=83-237-X. Accessed July 2, 2008

44. Fellegi IP, Sunter AB. A theory for record linkage. *Journal of the American Statistical Association* 1969; 64: 1183-210.

45. Rao JNK, Wu CFJ, Yue K. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 1992; 18(2): 209-17.

46. Rust KF, Rao JNK. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 1996; 5: 281-310.

47. Turcotte M, Schellenberg G. *A Portrait of Seniors in Canada* (Statistics Canada, Catalogue 89-519) Ottawa: Minister of Industry, 2007. Available at: www.statcan.ca/english/freepub/89-819-XIE/89-519-XIE2006001.pdf. Accessed January 4, 2008.

Appendix

Appendix Table A
Number and percentage of Canadian Community Health Survey respondents who agreed to have their survey responses linked with their administrative health records (HN) and who provided valid HN, by selected characteristics, Canada excluding Quebec, 2001

	Agreed to link		Agreed to link and HN valid	
	Number	%	Number	%
Total	**98,450**	**90.4**	**72,354**	**66.5**
Province/Territories				
Newfoundland and Labrador	3,533	91.3	2,933	75.8
Prince Edward Island	3,238	88.7	2,236	61.2
Nova Scotia	4,938	92.8	4,108	77.2
New Brunswick	4,634	92.8	3,746	75.0
Ontario	35,674	90.8	24,917	63.4
Manitoba	7,653	90.4	5,552	65.5
Saskatchewan	7,417	92.6	6,142	76.7
Alberta	12,757	88.2	9,155	63.3
British Columbia	16,493	90.1	11,990	65.5
Territories	2,113	83.9	1,575	62.6
Sex				
Female	52,865	90.5	40,334	69.1
Male	45,585	90.3	32,020	63.4
Age group				
12 to 24	19,246	91.8	13,538	64.5
25 to 34	14,482	90.9	10,119	63.5
35 to 44	18,892	90.2	13,883	66.3
45 to 54	16,036	89.6	11,905	66.5
55 to 64	11,493	90.0	8,695	68.1
65 to 74	9,778	90.3	7,684	71.0
12 to 74	89,927	90.5	65,824	66.3
75 or older	8,523	89.5	6,530	68.6

Source: 2000/2001 Canadian Community Health Survey.

Combining cycles of the Canadian Community Health Survey

by Steven Thomas and Brenda Wannell

Abstract

Background
A single cycle of the Canadian Community Health Survey (CCHS) may not meet researchers' analytical needs. This article presents methods of combining CCHS cycles and discusses issues to consider if these data are to be combined. An empirical example illustrates the proposed methods.

Data and methods
Two methods can be used to combine CCHS cycles: the separate approach and the pooled approach. With the separate approach, estimates are calculated for each cycle separately and then combined. The pooled approach combines data at the micro-data level, and the resulting dataset is treated as if it is a sample from one population.

Results
For the separate approach, it is recommended that the simple average of the estimates be used. For the pooled approach, it is recommended that weights be scaled by a constant factor where a period estimate covering the time periods of the individual cycles can be created. The choice of method depends on the aim of the analysis and the availability of data.

Interpretation
Combining cycles should be considered only if the most current period estimates do not suffice. Both methods will obscure cycle-to-cycle trends and will not reveal changing behaviours related to public health initiatives.

Keywords
data collection, data pooling, meta-analysis, statistical data interpretation

Authors
Steven Thomas (1-613-951-7300; Steven.Thomas@statcan.gc.ca) is with the Household Survey Methods Division, and Brenda Wannell (1-613-951-8554; Brenda.Wannell@statcan.ca) is with the Health Statistics Division at Statistics Canada, Ottawa, Ontario, K1A 0T6.

T he Canadian Community Health Survey (CCHS) consists of two cross-sectional sample surveys. The .1 cycle collects general health information from more than 120 health regions, while the .2 cycle focuses on specific health topics and collects data for estimation at the provincial level.

Despite large sample sizes, a single CCHS cycle may not meet users' needs. For instance, researchers may be interested in studying a rare population defined by detailed geography or by relatively rare socio-demographic or health characteristics. Because a single cycle may yield few observations for such a population, combining cycles may be considered. For example, this option was used by Tremblay et al.[1] in an examination of the relationship between body mass index and ethnicity, and by Tjepkema[2] in a study of health care use among gay, lesbian and bisexual Canadians.

The possibility of combining cycles exists because data for the same characteristics have generally been collected in all .1 cycles, and some of the same information is collected in .2 cycles. Nonetheless, as the CCHS has evolved, differences have emerged from cycle to cycle that may mean combining cycles is not feasible, or if still possible, may affect the results, depending on the analytical objectives of the study.

This article explains methods of combining CCHS cycles and offers guidelines for interpreting the results. Although the information pertains specifically to the CCHS, many of the issues have broader applicability. A case study illustrates the methods and shows that satisfactory estimates can be produced from combined cycles.

Starting in 2007, the CCHS implemented continuous collection with the intention of producing annual files as well as two-year combined files. This introduces different "period estimates," which will be the topic of a related article. This article focuses on the methodology and considerations for combining *past* cycles of the CCHS.

An evolving survey

The CCHS was not designed as a rolling sample,[3,4] expressly constructed to allow the different samples collected over time to be combined. Consequently, combining should be undertaken only after it has been determined that the estimates from a single cycle do not

meet analytical needs, and also, that the combined results will be relevant and interpretable.

Since its inception in 2000/2001, the CCHS has evolved. Consequently, the estimates derived from different cycles may not be comparable. To determine if combining cycles is feasible, changes in questionnaire content, survey coverage, geography, and mode of collection must be considered.

Changes in content

The CCHS questionnaire has undergone continual modification, including the introduction of new modules and removal of old ones. When content modifications are substantial, variable names usually change. Nonetheless, the same variable name does not necessarily indicate that exactly the same question was asked, so the wording of questions should be verified before cycles are combined. Users can consult CCHS documentation, notably the data dictionaries and questionnaires available from Statistics Canada's website (surveys and statistical programs within Definitions, Data Sources and Methods at http://www.statcan.ca/english/concepts/index.htm). Revisions to question wording, module structure, and response categories may mean that combining is not appropriate.

Changes in coverage

The populations targeted by certain modules of the CCHS questionnaire may differ from cycle to cycle. The most obvious example is the optional content that health regions/provinces can choose. As a result, the modules administered to the residents of a particular area in one cycle may be asked of the residents of an entirely different area in the next.

Another possibility is a change in the target population of a module. For instance, in cycle 1.1, the sexual behaviour module was asked of people aged 15 to 59, but in cycle 2.1, the target age group was narrowed to 15 to 49.

Changes in geography

The data file for each CCHS cycle contains geography coding and identifiers for the health regions as they were when the data were disseminated. However, health regions can change from one cycle to another. While these may be as minor as changes in names or codes, it is also possible for boundaries to be redrawn. If this has occurred, the files must be updated to a common geography (usually the most recent) before cycles can be combined. More information about boundary changes is available in the Internet publication, *Health Indicators* (health regions and peer group section, health region changes subsection) at http://www.statcan.ca/bsolc/english/bsolc?catno=82-221-XIE. If updated health region boundaries are required, correspondence files providing the relationship between Dissemination Areas (DA) or Enumeration Areas (EA) and the health regions for a given reference period are available in the Internet publication, *Health Regions: Boundaries and correspondence with census geography*, at http://www.statcan.ca/bsolc/english/bsolc?catno=82-402-X&CHROPG=1.

Changes in mode

The "mode effect" is the impact the method of collection has on the way respondents answer survey questions. CCHS interviews are conducted both by telephone and in person. The information that respondents provide can differ depending on the mode used for their interview. A 2004 study[5] found that several CCHS variables are susceptible to the mode effect, including, but not limited to, height and weight, physical activity, contact with doctors, and unmet health care needs.

To secure consistent estimates, efforts are made to maintain the same mix of telephone and personal interviews from one cycle to the next. However, large supplementary additions to the survey (buy-in samples) can affect the telephone/personal interview balance, because these supplementary interviews

are usually conducted by telephone. For cycle 1.1, the proportion of telephone interviews was quite low, a factor that should be recognized when considering combining that cycle with others.

Combining different surveys

For the reasons outlined above, the results of *different* cross-sectional health surveys may not be comparable, and in most situations, should not be combined. Therefore, it is recommended that the regional component of the CCHS (.1 cycles) not be combined with the provincial components (.2 cycles – Mental Health (2002) and Nutrition (2004)).

An evolving population

The feasibility of combining CCHS cycles derives from the fact that if random samples are taken from a population, the accumulated samples can be considered as one large random sample from the same population. However, if the population changes significantly between cycles, the samples cannot be treated as though they came from the same population. In the case of the CCHS, the samples for the successive cycles are drawn from an evolving population. Consequently, the combined sample is not necessarily representative of any of the populations represented by one cycle alone, but rather, the combined population.

Differences that emerge from cycle to cycle may stem from the reasons mentioned above—changes in the questionnaire, coverage and collection mode—or from sampling variability. However, changes from one cycle to another may reflect actual changes in the parameter under study. In such situations, combining cycles is still possible, but interpretation of the results requires an understanding of the effect of the time periods covered by the combined sample estimate. It is also important to be aware that, when combined in a single estimate, such trends will be obscured.

Methods for combining

Methods for combining data from different surveys can be divided into two broad categories: the separate approach and the pooled approach. The separate approach employs composite estimation techniques, whereby estimates are calculated for each survey separately and then combined. The pooled approach combines sample data at the micro-data level, and the resulting dataset is treated as if it is a sample from one population.

The separate approach

The separate approach creates an average of estimates calculated from the different CCHS cycles. The advantage is that, with some assumptions, the combined result is easy to interpret. As well, an average can be calculated from existing tables, which makes the approach appealing to users of the Public Use Microdata Files (PUMFs) and to those who rely on existing tables of estimates.

The disadvantage of the separate approach is that it can be cumbersome. If the required estimates are not published or do not include the variances, estimates must be calculated from each survey separately before being integrated. PUMF users will be limited by the information contained on the PUMF, and users relying on tables will have to gain access to the microdata. If many estimates are needed, the process is time-consuming.

In the case of the CCHS, estimates of a population parameter θ (which can be any statistic such as a mean, total or ratio) can be calculated separately for each cycle, $\hat{\theta}_1$, $\hat{\theta}_2$, ..., $\hat{\theta}_k$, where k is the number of cycles available. A simple average can then be calculated as:

$$\hat{\theta}_c^{avg} = \frac{\sum_{i-1}^{k}\hat{\theta}_i}{k}$$

For variance to be estimated relatively easily, the samples must be independent, which is true for most CCHS cycles. The exceptions are 2.1 and 2.2, where cycle 2.1 respondents were used as a

frame for cycle 2.2. Therefore, cycles 2.1 and 2.2 cannot be easily combined with the separate approach.

Based on the assumption of independence between the cycles, an estimate of the variance of the simple average of the three .1 cycles can be calculated as:

$$\hat{V}(\hat{\theta}_c^{avg}) = \hat{V}\left(\frac{\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3}{3}\right)$$

$$= \frac{1}{9}\left[\hat{V}(\hat{\theta}_1) + \hat{V}(\hat{\theta}_2) + \hat{V}(\hat{\theta}_3)\right]$$

It is evident that the estimated variance of the average of the three cycles is roughly one-third of the estimated variance of an estimate from one cycle alone. Standard errors can be calculated by taking the square-root of the variance, and estimates of the CV can be calculated as:

$$CV(\hat{\theta}_c^{avg}) = \frac{\sqrt{\hat{V}(\hat{\theta}_c^{avg})}}{\hat{\theta}_c^{avg}}$$

In some instances, it may be desirable to estimate a weighted average rather than the simple average, with more weight given to one estimate than another. Assuming that a researcher is interested in estimating for the same parameter θ as described with the simple average, separate estimates $\hat{\theta}_1$, $\hat{\theta}_2$, ..., $\hat{\theta}_k$ can be calculated, and a composite estimate or a weighted average can then be calculated as:

$$\theta_c = \sum_{i-1}^{k}\alpha_i\hat{\theta}_i$$

where $\sum_{i-1}^{k}\alpha_i = 1$.

If each estimate $\hat{\theta}_i$ is an unbiased estimate of θ, then $\hat{\theta}_c$ will also be unbiased, for any choices of α_i. That is, if each cycle correctly estimates the same constant statistic for the same population, the combined result will correctly estimate the same statistic.

Depending on the analysis, there are several choices for α_i. Some choices include an increasing weight function with more weight given to more current cycles, or a weight function based on variances, which results in a more

efficient estimate of the population parameter (that is, lower variance). More information about these methods is available in Chu, Brick and Kalton[6] and Korn and Graubard.[7]

Once the composite estimate has been computed using the appropriate value of i, an estimate of the variance can be calculated as a function of the original variances, and standard errors and CVs can be estimated. Assuming the cycles are independent, the variance can be estimated as:

$$\hat{V}(\hat{\theta}_c) = \hat{V}\left(\sum_{i-1}^{k}\alpha_i\hat{\theta}_i\right)$$

$$= \sum_{i=1}^{k}\alpha_i^2\hat{V}(\hat{\theta}_i)$$

For the separate approach to yield an unbiased estimate of a population parameter, the estimates being combined must each be unbiased estimates of the same population parameter. As noted earlier, this is problematic for the CCHS, the purpose of which is to measure the characteristics of an evolving population at different points in time. Because the assumption of a constant statistic is questionable, thereby making a weighted average difficult to interpret, it is recommended that users interested in the separate approach employ the simple average, where this assumption is not required, and the result is easier to interpret.

The pooled approach

The pooled approach consists of combining different CCHS cycles at the micro-data level to obtain a dataset that can be analysed as a single sample from a population. The pooled approach is an attractive option because of the power of the increased sample size, and because, once combined, it is not necessary to return to the individual datasets.

The disadvantages are that more technical expertise in the manipulation of data files is required, and it is not an option for users who do not have access to the microdata files. PUMF users are able to calculate an estimate using the pooled approach, but cannot

calculate the variance because CV tables are not available for the combined data file.

In its most basic form, pooling involves taking the individual data files with the corresponding weights and using a simple merge or set statement in SAS to create one data file. At the same time, the bootstrap weight files must be combined for variance estimation. Once these files are created, the resulting data file and bootstrap weight file can be treated as if it was one sample from one population. Estimates of rates and proportions, as well as statistical models, can be created with the files and any statistical program capable of estimating variances using the bootstrap method, such as Statistics Canada's Bootvar program.

The approach described above may not be appropriate for estimating totals. For example, to estimate the number of diabetes cases from two independent surveys of a common population, it is not possible to sum the sample weights from both surveys for respondents with diabetes—this would overestimate the total by a factor of two.[7] An option is to rescale the original sampling weights w_i by the factor α_i to represent the population of interest, as was done with the separate approach.

There are several choices of α_i.[8] Because the assumption that each CCHS cycle can be used to estimate the same population parameter is questionable, it is recommended that weights be scaled by a constant factor, $\alpha_i = 1/k$. If two cycles are combined, this means that $\alpha=0.5$; in the case of three cycles, $\alpha=0.33$. The resulting estimates can be interpreted as representing the characteristics of the average population (or a period estimate), which covers the combined time periods of the individual cycles. This does not require the assumption that each cycle estimates the same parameter.

It is not always necessary to adjust the weights when pooling the data. When weights are adjusted, the assumption is that they are being adjusted to properly represent a population. The problem is that when weights from different time periods are combined, the resulting weights do not represent the current population, but rather, an average population that does not exist. Consequently, creating totals with a combined file may not be appropriate, whether or not the weights are adjusted. On the other hand, ratios, proportions and means can be regarded as useful statistics when considered as period estimates. For these types of statistics, the results using the original weights or the weights that have been adjusted using a common $\alpha_i = 1/k$ will give the same result. This also holds for regression parameters, where weights are used in the model in order to take the survey design into account rather than to make estimates for some finite population.

One of the main applications of the pooled approach is in complex analysis using regression models.[1,2] With the increased sample size available from combined data, more detailed regression models can be studied. As well, the cycle/time effect can be considered in the model, and if significant, controlled. Other factors such as the mode effect can also be considered/controlled in such models, thereby making it possible to combine results from different cycles that would otherwise not be comparable.

Comparing approaches

The separate approach and the pooled approach do not always yield the same estimate. As an illustration, the separate approach of taking a simple average of two ratios, *a/b* and *c/d,* is not equal to the pooled approach, where a period estimate is calculated. This is because, generally speaking

$$\left(a/b + c/d\right) \neq (a+c)/(b+d)$$

Therefore, while both methods are valid, the choice depends on the goal of the analysis. Using a Canada estimate as an example, some researchers may choose to study the average of provincial estimates, which gives equal weight to each province (separate approach), while others are interested in the national estimate (pooled approach), which is influenced more by larger provinces.

For ratios such as a proportion, the two approaches will generally yield the same results as long as the parameter being estimated remains constant between the two occurrences, or the populations remain unchanged. For statistics such as regression parameters, it may be preferable to use a pooled approach to calculate the parameters instead of taking an average of the regression parameters calculated from the different cycles.

The Durham project

In 2007, the Durham (Ontario) health unit proposed producing a report on the health of Durham's adolescents, using combined CCHS data. The *Adolescent Health Snapshot* would target the 12-to-19 age group, and when possible, ages 12 to 14 and 15 to 19 separately. Based on combined CCHS data, Durham rates would be compared with provincial rates to reveal differences that were not evident from one cycle alone.

The variables of interest (typically, low-prevalence characteristics) were:
- daily smokers
- daily and occasional smokers
- current alcohol drinkers
- heavy drinkers
- sexual activity
- level of physical activity
- physical inactivity
- fruit and vegetable consumption
- use of protective gear (helmets while biking)
- overweight and obesity (youth body mass index - BMI)

After initial analysis to ensure that comparable data were available from more than one CCHS cycle, two variables were dropped:
- protective equipment, because of questionnaire changes across cycles
- BMI, because the derived variable created for cycle 3.1 was not available for cycles 1.1 and 2.1.

Several other potential variables could not be included because they had not

been selected consistently as optional content by Durham region: suicidal thoughts, food insecurity, and illicit drug use. (An ancillary benefit of this project was that the value of combining cycles became evident and may influence regions' selection of optional content in the future.)

The *daily smokers* variable illustrates the process of combining cycles. For any analysis, it is recommended that there be at least 10 observations with the characteristic under study before an estimate is calculated. Even with combined data, analysis of the 12-to-14 age group was not possible because of the limited sample size and the small number of respondents who were daily smokers. However, it was possible to examine daily smoking among 15-to 19-year-olds in the Durham region.

Preliminary analysis of the entire 12-to-19 age group consisted of calculating the estimates for each cycle alone. It was clear that for the total age group combining cycles was not necessary: the estimates of daily smokers from each cycle were publishable, with coefficients of variation below the recommended 33% cutoff (Table 1). It was also clear that the proportion of daily smokers in the 12-to-19 age group fell sharply from just over 12% in cycle 1.1 to around 7% in cycles 2.1 and 3.1. Therefore, it would have been erroneous to conclude that the rates were the same from cycle to cycle, rendering some of the methods of combining outlined above inappropriate. As well, the drop in the smoking rate suggested that it may not have been appropriate to

combine data from cycle 1.1 with the other cycles. If the decline reflects a major policy initiative, it would be preferable to analyze combined data for only those periods (cycles 2.1 and 3.1) when the policy was in place.

The separate approach of calculating a simple average and the pooled approach of calculating a period estimate were both used to combine all three cycles of data. The combined data masked changes in behaviour, notably the sharp decline in teen smoking. As well, the resulting estimate is confusing, since it differs from the latest published rates. This illustrates that the estimates must be interpreted as the average over the period rather than as an estimate of the current smoking rate.

With the *separate approach*, estimates of the percentage of daily smokers were averaged:

(12.37% + 6.91% + 7.26%) / 3 = 8.67%.

To estimate the variance, the estimated variances for each cycle were calculated. The estimated variance for cycle 1.1 was calculated by:

Estimated Variance = $(CV*Estimate)^2$
= $(.2233*.1238)^2 = 0.0008$

Similar estimates were calculated for cycles 2.1 and 3.1: .0004 and .0005, respectively. These variance estimates were then used to estimate the variance of the combined estimate with

Estimated Combined Variance = (0.0008 + 0.0004 + 0.0005) / 9 = 0.0002

A CV for the combined estimate was calculated by

Combined CV = sqrt(.0002) / .0867 = 16.3%,

which was an improvement over the CVs for one cycle alone and is acceptable for release under the publication guidelines.

With the *pooled approach*, a period estimate was calculated:

(7,577 + 4,598 + 5,110) / (61,220 + 66,523 + 70,380) = 17,285 / 198,123 = 8.72%.

There was a small difference between the simple average and the period estimate, mainly due to changes in the population size and the smoking rate.

Weights could have been adjusted for the pooled approach by dividing the original weights by 3, but the result would have been the same:

5,761 / 66,041 = 8.72%.

However, in the case of totals, the estimated population was 198,123 with the unadjusted weights, which was roughly three times the estimate for each cycle. The pooled estimate with the adjusted weights was 66,041, which was the average of the population counts for each cycle.

To estimate the variances with the pooled approach, Bootvar was used to calculate the estimates using the bootstrap method. The variance estimate for the pooled estimate was .0002, with a corresponding CV of 15.3%. As was shown with the separate approach, this is an improvement over the estimates when each cycle is treated independently.

Finally, a comparison of pooled Durham rates with the provincial rate was expected to reveal statistically significant differences because of the improved precision of the increased sample size. This was generally not the case. Differences between Ontario

Table 1
Estimates of daily smokers aged 12 to 19, Canadian Community Health Survey, cycles 1.1 to 3.1, Durham Health Region

	Cycle 1.1			Cycle 2.1			Cycle 3.1		
	Sample count	Estimate	Coefficient of variation	Sample count	Estimate	Coefficient of variation	Sample count	Estimate	Coefficient of variation
Total aged 12 to 19	187	61,220	...	210	66,523	...	214	70,380	...
Daily smokers	27	7,577	22.33%	18	4,598	29.30%	16	5,110	30.26%
Proportion	...	12.38%	22.33%	...	6.91%	29.30%	...	7.26%	30.26%

... not applicable
Source: 2000/2001 Canadian Community Health Survey, cycle 1.1; 2003 Canadian Community Health Survey, cycle 2.1; 2005 Canadian Community Health Survey, cycle 3.1.

and Durham were so small that they could not be detected, even with the larger sample sizes.

Conclusion

Combining CCHS cycles yields larger sample sizes for analysis, and the resulting estimates are of higher quality than those from one cycle alone. Nonetheless, it cannot be assumed that the resulting estimates represent the same population, or that the population characteristics are the same as those that would emerge from one cycle alone, even though the same question was asked from one cycle to another. Over time, the individuals who constitute the population and their characteristics evolve. Estimates based on combined cycles describe an "artificial" population made up of different populations surveyed at different times. Therefore, researchers should consider the implications for their analyses before combining cycles. ■

References

1. Tremblay M, Pérez C, Ardern C, et al. Obesity, overweight and ethnicity. *Health Reports* (Statistics Canada, Catalogue 82-003) 2005; 16(4): 23-34.

2. Tjepkema M. Health care use among gay, lesbian and bisexual Canadians. *Health Reports (*Statistics Canada, Catalogue 82-003) 2008; 19(1): 1-12.

3. Kish L. Cumulating/combining population surveys. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 1999; 25(2): 129-38.

4. Alexander CH. Still rolling: Leslie Kish's "Rolling Samples" and the American Community Health Survey. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 2002; 28(1): 35-41.

5. St-Pierre M, Béland Y. Mode effects in the Canadian Community Health Survey: a comparison of CAPI and CATI. *2004 Proceedings of the American Statistical Association Meeting*, *Survey Research Methods*. Toronto: American Statistical Association, 2004.

6. Chu A, Brick JM, Kalton G. Weights for combining surveys across time or space. *Bulletin of the International Statistical Institute: 52nd Session, Contributed Papers, Book 2*. 1999: 103-4.

7. Korn EL, Graubard BI. *Analysis of Health Surveys*. New York: Wiley, 1999.

8. Friedman EM, Jang D, Williams VT. Combined Estimates from Four Quarterly Survey Data Sets. *2002 Proceedings from the Joint Statistical Meetings – Section on Survey Research Methods*. 2002: 1064-69.