Catalogue no. 82-003-X
ISSN 1209-1367

Health Reports

Using personal health insurance numbers to link the Canadian Cancer Registry and the Discharge Abstract Database

by Dianne Zakaria, Richard Trudeau, Claudia Sanmartin, Patricia Murison, Gisèle Carrière, Maureen MacIntyre, Donna Turner, Brandon Wagar, Mary Jane King, Kim Vriends, Ryan Woods, Gina Lockwood and Rabiâ Louchini

Release date: June 17, 2015

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

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

Depository Services Program

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

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" > "The agency" > "Providing services to Canadians."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{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
*	significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Using personal health insurance numbers to link the Canadian Cancer Registry and the Discharge Abstract Database

by Dianne Zakaria, Richard Trudeau, Claudia Sanmartin, Patricia Murison, Gisèle Carrière, Maureen MacIntyre, Donna Turner, Brandon Wagar, Mary Jane King, Kim Vriends, Ryan Woods, Gina Lockwood and Rabiâ Louchini

Abstract

Background: Linking cancer registry and administrative data can reveal health care use patterns among cancer patients. The Canadian Cancer Registry (CCR) contains personal health insurance numbers (HINs) that facilitate linkage to hospitalization information in the Discharge Abstract Database (DAD).

Data and methods: Valid HINs, captured in the CCR or obtained through probabilistic linkages to provincial health insurance registries, were used to deterministically link prostate, female breast, colorectal and lung cancers diagnosed from 2005 through 2008 with the DAD for fiscal years 2004/2005 to 2010/2011.

Results: At least 98% of tumours diagnosed from 2005 through 2008 had valid HINs in the CCR or obtained through probabilistic linkages. For provinces submitting day surgeries to the DAD, linkage rates to at least one DAD record were higher for female breast (95.6% to 98.1%), colorectal (96.9% to 98.7%) and lung cancers (92.8% to 96.3%) than for prostate cancers (77.2% to 91.6%). Among linked records, agreement was high for sex (99% or more) and complete date of birth (97% or more); the likelihood of a consistent diagnosis in the CCR and on at least one linked DAD record was higher for female breast (86.8% to 97.2%), colorectal (94.6% to 97.7%) and lung cancers (90.3% to 95.5%) than for prostate cancers (77.4% to 87.8%).

Interpretation: Deterministically linking the CCR and DAD using personal HINs is a feasible and valid approach to obtaining hospitalization information about cancer patients.

Keywords: Administrative databases, data linkage, health care, health insurance number

Data linkages enhance the usefulness of information in different sources (for example, administrative databases, censuses, surveys) and provide insight not available when data sources are used in isolation. Linking information from cancer registries to administrative health data offers opportunities to study health care use patterns (including treatment) of cancer patients.[1-3] The Canadian Cancer Registry (CCR),[4] which contains unique patient identifiers that facilitate record linkage, has been probabilistically linked to census and mortality data to examine cancer outcomes for key subgroups.[5]

This study investigates the feasibility and validity of using personal health insurance numbers (HINs) to deterministically link the CCR and the Discharge Abstract Database (DAD) to obtain hospitalization information about people with primary cancers. Because patient names are not captured by the DAD, the provincially assigned HINs are essential for linkage, and have been used previously for deterministic data linkages in Ontario and Manitoba.[6-8] The methods employed to link the CCR and the DAD for nine provinces are described and the quality of this deterministic linkage is evaluated. Details on linkage rates, agreement on demographic identifiers and clinical diagnoses, and out-of-province hospital admissions are presented for prostate, female breast, colorectal and lung cancers, which together account for more than half of primary cancers diagnosed annually.[9] Cancers diagnosed from 2005 through 2008 with a valid HIN were included in the linkage. Because the territories have small cancer counts, and because Quebec does not submit data to the DAD, cancers reported by these jurisdictions were excluded from the linkage.

Data and methods

Data sources

Since 1992, the CCR has collected demographic and clinical information about Canadian residents diagnosed with primary tumours.[4] The data in the CCR include a person's name (current surname, birth surname and given name), sex, and date of birth; postal code of residence and HIN at the time of diagnosis are included in the tumour record(s) for each person.

Statistics Canada uses internal record linkage to ensure that all information for an individual is attached to a single person-level identifier. If a provincial/territorial cancer registry submits a new person-record and record linkage shows that person to be registered in the CCR, the cancer registries involved are consulted to confirm that the new record pertains to someone already in the CCR. If this is the case, that record is assigned the same person-level identifier as the other records for that individual.

Because the HIN is unique to individuals living in a specific province/territory and a key identifier in hospital discharge

Authors: Dianne Zakaria (dianne.zakaria@phac-apsc.gc.ca) is with the Public Health Agency of Canada. Richard Trudeau and Patricia Murison are with the Health Statistics Division and Claudia Sanmartin and Gisèle Carrière are with the Health Analysis Division at Statistics Canada. Maureen MacIntyre is with Cancer Care Nova Scotia. Donna Turner is with CancerCare Manitoba and the University of Manitoba. Brandon Wagar is with the Canadian Institute for Health Information. Mary Jane King is with Cancer Care Ontario. Kim Vriends is with the Prince Edward Island Cancer Registry. Ryan Woods is with the British Columbia Cancer Registry. Gina Lockwood is with the Canadian Partnership Against Cancer. Rabiâ Louchini is with the Fichier des Tumeurs du Québec.

databases,[10] it is ideal for deterministic linkages. Further, because both residence and personal HIN at time of diagnosis are captured in the CCR and linked to a person-level identifier, people diagnosed with primary tumours in more than one province/territory will have more than one HIN that can be used to link to the DAD. That is, each person in the CCR will be linkable to hospital episodes related to all of his/her primary cancers and associated HINs in the CCR. However, if a person registered as having a tumour moves to another province/territory and receives a new HIN, subsequent hospital episodes using the new HIN would not be obtainable through linkage unless a new primary cancer was registered in the CCR under the new HIN.

For the present study, analyses were completed on a snapshot of the CCR (April 4, 2012), which included 2,301,833 people and 2,483,305 tumours diagnosed from 1992 through 2008.

The DAD is a national database (excluding Quebec[11]) that contains information on all separations (discharges, deaths, sign-outs and transfers excluding stillbirths and cadaveric donor cases) from acute care institutions in Canada. Since its inception in 1963, the DAD's coverage and content have varied substantially. As of fiscal year (FY) 2004/2005, facilities submitted directly or indirectly (via ministries of health) to the DAD using the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Canada* (ICD-10-CA)[12] to code diagnoses, and the *Canadian Classification of Interventions* (CCI)[13] to code diagnostic and therapeutic interventions (Text Table 1). In FY 2004/2005, the Canadian Institute for Health Information (CIHI) introduced the "analytical institution type" variable (for example, acute care, chronic care, rehabilitation) that is assigned to each separation record. The DAD does not have a person-level identifier, but it does capture HIN and the province/territory issuing the HIN.[10,16]

Canadian Cancer Registry and Discharge Abstract Database linkage

To identify tumours with linkage potential, the CCR HIN variable was assessed for its presence and compliance with basic HIN characteristics (length, alphanumeric format, embedded patterns, check digit). For provinces with a high percentage of HINs in the CCR that satisfied basic checks ("valid" HINs), a deterministic linkage to the DAD was performed based on HIN and province reporting the tumour. The province reporting the tumour was used as a proxy for province issuing the HIN.

For provinces with a low percentage of valid HINs in the CCR, and for which Statistics Canada had access to the provincial health insurance registry (Ontario and Manitoba), a two-step linkage was performed. Name, sex, date of birth and postal code of residence were used to probabilistically link CCR records to provincial health insurance registries to obtain HINs. HIN and province reporting the tumour were then used to deterministically link the CCR to the DAD. Because tumours are linked to persons in the CCR, it was possible to associate each tumour with all of an individual's linked DAD records.

Only linkages occurring during clinically relevant follow-up periods were considered. Linkage rates for non-invasive tumours were not examined. For colorectal, lung and female breast cancers, the follow-up period extended from 31 days before to 365 days after the CCR diagnosis date; for prostate cancer, the follow-up period extended from 31 days before to 730 days after the diagnosis date. To be retained as a linked record, the DAD date of admission had to occur within the follow-up period.

FYs 2004/2005 to 2010/2011 of the DAD were used for the linkage. During this period, neither Alberta nor Ontario submitted day surgeries to the DAD.[16] As well, in FY 2004/2005, two facilities in Nova Scotia did not submit day surgeries to the DAD; the number rose to three in FY 2005/2006, and to four in FY 2010/2011.[16,17] Day surgery data from Ontario, Alberta and the four Nova Scotia facilities are submitted to the National Ambulatory Care Reporting System (NACRS).[18]

Tumours with a valid HIN in the CCR or a HIN obtained through probabilistic linkage to provincial health insurance registries were compared with tumours without a valid HIN by sex, age at diagnosis, microscopic confirmation, and tumour behaviour (non-malignant/malignant). Linkage rates of malignant/invasive tumours with valid HINs to at least one DAD record during follow-up were examined by analytical institution type, calendar year of diagnosis, sex and age group at diagnosis. Cancers linking to at least one DAD record were compared with cancers not linking in terms of diagnostic confirmation (histology, cytology, clinical/imaging/unknown, and autopsy only/death certificate only) and number of days alive during follow-up. Because the data file used for this study had vital status confirmation complete to December 31, 2008, examination of the number of days alive was limited to cancer records that had a follow-up end date on or before December 31, 2008.

The statistical tests for categorical variables were Fisher's exact test and the chi-square test, and for continuous vari-

Text Table 1
Cancer definitions

Cancer	Topography	ICD-10-CA
Prostate	C61.9	C61
Breast	C50.0-C50.9	C50.0-C50.99
Colorectal	C18.0-C18.9,C19.9,C20.9,C26.0	C18.0-C18.9,C19, C20,C26.0
Lung	C34.0-C34.9	C34.0-C34.99

Notes: Cancers defined according to Surveillance Epidemiology and End Results Program site groupings; limited to malignant/invasive tumours; exclude histologies 9050-9055, 9140, 9590-9992.[14] Female breast cancer was identified using the sex documented in the Canadian Cancer Registry.
Sources: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Canada[12]; International Classification of Diseases for Oncology, Third Editon.[15]

Table 1
Primary tumours in Canadian Cancer Registry and percentage with valid health insurance number, by province (excluding Quebec), 1992 to 2008

	Newfoundland and Labrador		Prince Edward Island		Nova Scotia		New Brunswick		Ontario		Manitoba		Saskatchewan		Alberta		British Columbia	
Year diagnosed	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1992 to 2008	**40,993**	**98.5**	**13,568**	**97.1**	**97,175**	**95.7**	**69,521**	**99.9**	**876,624**	**24.0**	**108,314**	**0.0**	**86,404**	**99.6**	**230,455**	**99.6**	**343,134**	**95.5**
1992	2,028	98.0	692	68.9	4,725	62.8	3,367	99.9	42,953	95.7	6,068	0.0	4,580	98.8	10,777	96.4	17,254	80.7
1993	2,192	98.5	714	84.5	5,104	73.1	3,562	99.9	43,690	96.3	6,253	0.0	4,816	99.5	11,090	97.2	18,045	81.9
1994	2,077	99.1	782	99.1	5,100	93.5	3,664	99.8	44,347	95.9	6,237	0.0	4,663	99.7	11,432	98.7	17,588	84.2
1995	2,345	98.9	712	98.5	4,941	97.2	3,446	99.9	44,108	95.6	6,135	0.0	4,584	99.8	11,546	99.8	17,734	87.6
1996	2,283	98.6	696	98.9	4,903	98.6	3,560	99.8	44,677	95.0	6,046	0.0	4,606	99.7	11,803	99.9	18,271	92.8
1997	2,285	98.8	711	100.0	5,056	98.4	3,775	99.8	47,129	0.0	6,226	0.0	4,799	99.7	12,444	99.8	19,032	94.9
1998	2,290	98.1	791	99.9	5,570	99.2	3,999	99.9	48,003	0.0	6,381	0.0	4,913	99.7	12,957	99.8	19,266	95.9
1999	2,297	99.1	812	99.9	5,775	99.0	4,048	99.9	49,493	0.0	6,392	0.0	5,094	99.6	13,634	99.9	20,328	99.5
2000	2,391	98.5	771	99.9	5,849	98.9	3,912	99.9	50,942	0.0	6,586	0.0	5,067	99.5	14,072	99.9	20,271	99.3
2001	2,499	98.4	796	98.9	6,024	99.3	4,144	99.9	52,308	0.0	6,590	0.0	5,294	99.6	14,732	99.9	20,786	99.6
2002	2,507	99.1	844	98.9	6,075	99.6	4,085	99.9	53,381	0.0	6,622	0.0	5,259	99.7	15,266	99.9	20,864	99.7
2003	2,447	99.2	909	99.1	6,019	99.4	4,190	99.8	54,000	0.0	6,458	0.0	5,251	99.7	14,310	100.0	21,224	99.6
2004	2,390	99.3	879	98.9	6,304	99.7	4,321	99.8	56,776	0.0	6,452	0.0	5,501	99.7	14,582	100.0	21,221	99.7
2005	2,484	99.4	830	99.4	6,424	99.7	4,701	99.8	59,310	0.0	6,334	0.0	5,367	99.5	14,579	100.0	22,006	99.9
2006	2,688	99.3	839	100.0	6,889	99.7	4,945	99.8	60,863	0.0	6,256	0.0	5,284	99.8	15,288	100.0	22,589	99.9
2007	2,849	98.4	911	99.9	6,276	99.7	4,824	99.9	62,927	0.1	6,569	0.0	5,619	99.8	15,731	100.0	23,443	99.9
2008	2,941	95.1	879	100.0	6,141	99.6	4,978	99.9	61,717	0.0	6,709	0.0	5,707	99.7	16,212	100.0	23,212	99.9

N = total number of tumours registered in Canadian Cancer Registry
% = percentage of registered tumours with valid health insurance number
Source: Canadian Cancer Registry data file, April 4, 2012.

ables, the two-sample t-test ($\alpha = 0.05$, two-tailed). To assess the validity of the linkages, agreement on sex, date of birth and diagnosis was examined, and out-of-province hospital admissions were calculated.

Statistics Canada ensures respondent privacy during the linkage process and subsequent use of linked files. Only employees directly involved in the process have access to the unique identifying information required for linkage (such as names and health insurance numbers) and do not access health-related information. When the data linkage is completed, an analytical file is created from which identifying information is removed. This de-identified file is accessed by analysts for validation and analysis.

Table 2
Primary tumours in Canadian Cancer Registry and percentage probabilistically linked to provincial health insurance number, Ontario and Manitoba, 1992 to 2008

	Ontario		Manitoba	
Year diagnosed	N	%	N	%
1992 to 2008	**876,624**	**97.9**	**108,314**	**95.8**
1992	42,953	96.8	6,068	88.2
1993	43,690	96.9	6,253	93.1
1994	44,347	97.1	6,237	94.6
1995	44,108	97.4	6,135	94.2
1996	44,677	97.4	6,046	93.5
1997	47,129	97.8	6,226	94.7
1998	48,003	97.8	6,381	95.2
1999	49,493	98.0	6,392	95.1
2000	50,942	98,1	6,586	96.0
2001	52,308	98.4	6,590	96.2
2002	53,381	98.1	6,622	96.6
2003	54,000	98.1	6,458	97.1
2004	56,776	98.1	6,452	97.8
2005	59,310	98.3	6,334	98.4
2006	60,863	98.2	6,256	98.8
2007	62,927	98.3	6,569	98.9
2008	61,717	98.5	6,709	99.2

N = total number of tumours registered in Canadian Cancer Registry
% = percentage of registered tumours linking to provincial health insurance number
Sources: Canadian Cancer Registry data file, April 4, 2012; Ontario and Manitoba health insurance registries.

Results

Health Insurance Number validity check

The HIN check revealed that deterministic linkage using HINs in the CCR would not be feasible for Ontario and Manitoba (Table 1). Among the remaining provinces, the percentage of valid HINs was 99% or more for 2005 to 2008 (except for Newfoundland and Labrador in 2007 and 2008—98.4% and 95.1%, respectively; almost all the "invalid" HINs were actually missing).

For Ontario and Manitoba, 97.9% and 95.8% of tumours probabilistically linked to a HIN in the respective health insurance registries (Table 2). For the 2005-to-2008 period, the percentage of tumours with a valid HIN exceeded 98% in both provinces. Of tumours probabilistically linked, agreement on sex between the CCR and the provincial health insurance registry was 99.9% for Ontario and 100.0% for Manitoba; agreement on complete date of birth was 98.7% for Ontario and 71.9% for Manitoba. Agreement on date of birth in Manitoba rose from 51.1% in 1992 to 93.6% in 2008. Examination of the Manitoba discrepancies revealed that health registry dates of birth disproportionately used the first day of the month.

Several differences emerged between tumours with and without a valid HIN. Tumours with a valid HIN were more likely to belong to males (50.3% versus 46.1%, $p < 0.0001$), to be malignant (92.9% versus 83.7%, $p < 0.0001$), and to be microscopically confirmed (85.7% versus 74.6%, $p < 0.0001$). Among invasive tumours, those with a valid HIN were more likely to be microscopically confirmed (85.4% versus 70.2%, $p < 0.0001$).

CCR-DAD linkage rate during follow-up

Prostate cancer had the lowest linkage rate to the DAD. For provinces submitting day surgeries to the DAD, the percentage of prostate cancers linking to at least one DAD record ranged from 77.2% to 91.6% (Table 3). When the analysis was limited to DAD records submitted by acute care institutions, the percentage linking was lower, ranging from 58.1% to 65.4%. Linkage rates tended to decline with advancing age at diagnosis until 80, and then increased (Table 4). The method of diagnostic confirmation did not differ substantially between prostate cancers that linked and those that did not (Table 5). The average number of days alive during follow-up was actually greater for prostate cancers not linking to the DAD (741.2 versus 716.1 days, $p < 0.0001$).

Table 3
Number of cancers and percentage linking with at least one Discharge Abstract Database record during follow-up, by type of cancer, province (excluding Quebec) and analytical institution type, 2005 to 2008

Province and analytical institutional type	Type of cancer			
	Prostate	Female breast	Colorectal	Lung
Newfoundland and Labrador				
N	1,556	1,289	1,867	1,306
%	91.6	98.1	98.7	94.4
% AC	63.8	83.5	95.2	85.2
Prince Edward Island				
N	536	370	417	465
%	83.4	96.8	98.6	96.3
% AC	64.4	93.2	94.2	86.2
Nova Scotia				
N	3,270	2,903	3,231	3,457
%	87.2	96.2	97.4	94.0
% AC	64.9	79.2	92.6	84.5
New Brunswick				
N	2,952	2,148	2,106	2,802
%	79.5	97.2	97.0	94.9
% AC	58.7	77.6	95.8	91.4
Ontario				
N	36,600	31,288	29,684	30,052
%	62.6	56.9	93.5	87.4
% AC	62.6	56.8	93.5	87.3
Manitoba				
N	2,561	3,124	3,139	3,289
%	77.2	95.6	97.8	94.1
% AC	64.3	71.6	92.9	85.4
Saskatchewan				
N	3,231	2,667	2,729	2,803
%	86.9	97.5	98.0	94.5
% AC	60.9	91.8	94.4	89.9
Alberta				
N	8,282	7,646	6,635	7,052
%	65.7	86.4	94.4	84.8
% AC	65.4	86.3	94.2	83.8
British Columbia				
N	12,328	11,131	10,260	11,037
%	80.9	95.6	96.9	92.8
% AC	58.1	76.2	92.9	81.8

N = number of registered cancers with valid health insurance number
% = percentage of cancers linked to at least one Discharge Abstract Database record during follow-up
% AC = percentage of cancers linked to at least one Discharge Abstract Database record, submitted by an acute care institution, during follow-up
Notes: Follow-up period for colorectal, lung and female breast cancer extended from 31 days before to 365 days after Canadian Cancer Registry date of diagnosis; for prostate cancer, follow-up extended from 31 days before to 730 days after date of diagnosis.
Sources: Canadian Cancer Registry data file, April 4, 2012; Discharge Abstract Database, fiscal years 2004/2005 to 2010/2011.

Female breast cancer linkage rates varied by analytical institution type, calendar year, and age at diagnosis. Among provinces submitting day surgeries to the DAD, the percentage of female breast cancers linking to at least one DAD record ranged narrowly from 95.6% to 98.1%. When the analysis was limited to DAD records submitted by acute care institutions, the percentage linking varied widely from 56.8% to 93.2% (Table 3). For provinces with both acute care and day surgery captured in the DAD, overall breast cancer linkage rates remained relatively stable over time, but the linkage rate to acute care records declined, suggesting a shift of procedures toward day surgery (data not shown). Apart from Ontario, linkage rates tended to be lower in the 80 or older age range (Table 4). Compared with breast cancers not linking to a DAD record, those that linked were more likely to be histologically confirmed (98.3% versus 93.3%, $p < 0.0001$); the average number of days alive during follow-up was almost the same (383.9 versus 385.1 days, $p = 0.0626$) (Table 5).

Colorectal cancer linkage rates varied little by analytical institution type, calendar year, sex and age at diagnosis. Compared with colorectal cancers not linking to a DAD record, those that linked were more likely to be histologically confirmed (95.2% versus 74.2%, $p < 0.0001$) and had more days alive during follow-up (345.2 versus 294.8, $p < 0.0001$) (Table 5). The difference in average days alive was attributable to the high percentage of autopsy-only/death-certificate-only cases among colorectal cancers that did not link; for all other diagnostic confirmation categories, average days alive were greater for colorectal cancers not linking (data not shown). In the CCR, the date of death and the date of diagnosis are the same for autopsy-only/death-certificate-only cases, which limits the number of days alive during follow-up to 31.

Lung cancer linkage rates varied by analytical institution type and age at diagnosis. Among provinces submitting day surgeries to the DAD, the percentage of lung cancers linking to at least one DAD record ranged from 92.8% to 96.3%. When the analysis was limited to records from acute care institutions, the percentage linking ranged from 81.8% to 91.4% (Table 3). Linkage rates tended to decline in the oldest age group (Table 4). Compared with lung cancers not linking to a DAD record, those that linked were more likely to be histologically confirmed (62.9% versus 45.3%, $p < 0.0001$), and had fewer days alive during follow-up (239.0 versus 264.4, $p < 0.0001$), despite having a smaller percentage of autopsy-only/death-certificate-only cases

Table 4
Percentage of cancers linking with at least one Discharge Abstract Database (DAD) record during follow-up, by age group at diagnosis, type of cancer and province (excluding Quebec), 2005 to 2008

Type of cancer and province	Age group at diagnosis						
	Total	0 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 or older
	Percentage linking with at least one DAD record						
Prostate							
Newfoundland and Labrador	91.6	x	x	96.8	92.2	86.9	89.3
Prince Edward Island	83.4	. .	93.3	87.3	82.3	80.5	84.1
Nova Scotia	87.2	x	x	90.5	88.4	83.0	86.6
New Brunswick	79.5	. .	85.7	87.0	81.1	72.4	78.0
Ontario	62.6	46.7	76.5	72.6	66.6	48.0	68.6
Manitoba	77.2	x	x	87.4	79.4	69.0	77.2
Saskatchewan	86.9	. .	87.5	91.2	87.8	85.2	84.1
Alberta	65.7	x	x	73.1	66.3	55.9	68.4
British Columbia	80.9	66.7	90.2	85.9	84.0	75.6	79.7
Female breast							
Newfoundland and Labrador	98.1	100.0	98.0	99.1	98.5	98.7	92.5
Prince Edward Island	96.8	100.0	97.9	97.5	99.1	95.9	89.1
Nova Scotia	96.2	98.9	97.4	98.0	98.2	96.1	88.7
New Brunswick	97.2	100.0	99.1	98.6	99.4	96.9	89.9
Ontario	56.9	61.4	54.2	52.6	54.0	61.0	67.2
Manitoba	95.6	97.7	98.7	98.4	98.5	96.0	84.1
Saskatchewan	97.5	98.7	98.0	99.2	98.6	99.6	90.4
Alberta	86.4	87.1	87.0	86.2	85.9	89.4	82.3
British Columbia	95.6	94.2	95.6	96.8	97.4	96.7	89.2
Colorectal							
Newfoundland and Labrador	98.7	100.0	97.8	98.2	99.4	99.1	97.3
Prince Edward Island	98.6	100.0	100.0	98.2	100.0	97.9	98.0
Nova Scotia	97.4	100.0	98.6	99.1	99.5	98.1	93.3
New Brunswick	97.0	95.7	97.9	98.6	96.0	97.8	95.7
Ontario	93.5	90.9	90.1	92.1	93.3	95.3	93.6
Manitoba	97.8	98.2	97.6	99.6	99.1	98.2	95.3
Saskatchewan	98.0	96.7	98.4	99.4	99.1	99.4	94.9
Alberta	94.4	87.8	93.5	93.9	94.8	95.8	93.5
British Columbia	96.9	97.1	97.8	98.2	98.9	98.4	92.3
Lung							
Newfoundland and Labrador	94.4	x	x	98.6	96.9	92.1	86.5
Prince Edward Island	96.3	. .	100.0	100.0	97.5	96.9	90.2
Nova Scotia	94.0	92.9	98.4	97.6	96.3	94.4	86.1
New Brunswick	94.9	100.0	99.1	96.8	95.7	95.1	91.1
Ontario	87.4	90.3	88.0	87.2	88.3	88.0	84.7
Manitoba	94.1	90.9	96.9	97.2	96.1	94.3	89.3
Saskatchewan	94.5	100.0	98.9	98.5	97.5	95.6	86.4
Alberta	84.8	85.7	83.5	86.1	84.7	86.1	81.9
British Columbia	92.8	100.0	96.4	95.9	95.0	94.0	85.7

. . not available
x suppressed to meet confidentiality requirements of *Statistics Act*
Notes: Follow-up period for colorectal, lung and female breast cancer extended from 31 days before to 365 days after Canadian Cancer Registry date of diagnosis; for prostate cancer, follow-up extended from 31 days before to 730 days after date of diagnosis.
Sources: Canadian Cancer Registry data file, April 4, 2012; Discharge Abstract Database, fiscal years 2004/2005 to 2010/2011.

(Table 5). The average number of days alive for lung cancers that did not link significantly exceeded the average for cancers that did link for each category of diagnostic confirmation except autopsy-only/death-certificate-only, where average days alive were equal at 31 (data not shown).

Validity of linked records

More than 99% of DAD records that linked to cancers agreed with the patient sex reported in the CCR, and apart from two estimates for Prince Edward Island, more than 97% agreed with the complete date of birth in the CCR (Table 6). For Prince Edward Island, the 42 prostate cancer discrepancies involved 16 people, and the 47 lung cancer discrepancies, 19 people. Common discrepancies were one-day and one-year differences, and transposed month and day.

Generally, prostate cancers were most likely, and female breast cancers were least likely, to link to an out-of-province hospital admission. For all cancers, Prince Edward Island had the highest percentage linking to at least one out-of-province admission; Ontario had the lowest (Table 7).

The likelihood of a CCR record linking to at least one DAD record with a consistent diagnosis varied by cancer (Table 7). Prostate cancers were least likely (77.4% to 87.8%) and colorectal

Table 5
Method of diagnostic confirmation and mean days alive, by type of cancer and linkage status, 2005 to 2008

Type of cancer, method of diagnostic confirmation and mean days alive	No link to Discharge Abstract Database record during follow-up			Linked to at least one Discharge Abstract Database record during follow-up			
	N	% or mean	SD	N	% or mean	SD	p-value
Prostate							
Total	21,134	100		50,182	100		
Method of diagnostic confirmation (%)							<0.0001
Histology	20,115	95.2	. . .	48,080	95.8	. . .	
Cytology	20	0.1	. . .	64	0.1	. . .	
Clinical/Unknown	747	3.5	. . .	1,923	3.8	. . .	
Autopsy only/Death certificate only	252	1.2	. . .	115	0.2	. . .	
Mean days alive	10,269	741.2	111.8	24,708	716.1	146.3	<0.0001
Female breast							
Total	15,428	100	. . .	47,138	100	. . .	
Method of diagnostic confirmation (%)							<0.0001
Histology	14,396	93.3	. . .	46,337	98.3	. . .	
Cytology	120	0.8	. . .	215	0.5	. . .	
Clinical/Unknown	697	4.5	. . .	542	1.1	. . .	
Autopsy only/Death certificate only	215	1.4	. . .	44	0.1	. . .	
Mean days alive	11,124	385.1	58.8	35,565	383.9	56.1	0.0626
Colorectal							
Total	2,907	100	. . .	57,161	100	. . .	
Method of diagnostic confirmation (%)							<0.0001
Histology	2,158	74.2	. . .	54,414	95.2	. . .	
Cytology	18	0.6	. . .	235	0.4	. . .	
Clinical/Unknown	271	9.3	. . .	2,326	4.1	. . .	
Autopsy only/Death certificate only	460	15.8	. . .	186	0.3	. . .	
Mean days alive	2,124	294.8	156	42,395	345.2	110.2	<0.0001
Lung							
Total	6,451	100	. . .	55,812	100	. . .	
Method of diagnostic confirmation (%)							
Histology	2,924	45.3	. . .	35,131	62.9	. . .	<0.0001
Cytology	1,646	25.5	. . .	12,191	21.8	. . .	
Clinical/Unknown	1,068	16.6	. . .	8,046	14.4	. . .	
Autopsy only/Death certificate only	813	12.6	. . .	444	0.8	. . .	
Mean days alive	4,887	264.4	156	41,686	239	144.6	<0.0001

N = number of registered cancers with valid health insurance number
. . . not applicable
Notes: Follow-up period for colorectal, lung and female breast cancers extended from 31 days before to 365 days after Canadian Cancer Registry date of diagnosis; for prostate cancer, follow-up extended from 31 days before to 730 days after date of diagnosis. Examination of number of days alive during follow-up was limited to cancer records that underwent vital status confirmation at Statistics Canada and had follow-up end date on or before December 31, 2008. Clinical confirmation includes laboratory tests, marker studies, direct visualization without microscopic confirmation, imaging and physical exam.
Sources: Canadian Cancer Registry data file, April 4, 2012; Discharge Abstract Database, fiscal years 2004/2005 to 2010/2011.

What is already known on this subject?

- Data linkages increase the utility of information in different sources and offer analytical possibilities beyond what is available when data sources are used in isolation.
- Probabilistic linkages involving the Canadian Cancer Registry (CCR) have been performed based on personal information such as name, date of birth, sex, and postal code.
- The Discharge Abstract Database (DAD) contains health insurance numbers (HINs) that can be used to deterministically link to the CCR, but the feasibility of such a linkage is unknown.

What does this study add?

- Deterministically linking the CCR and the DAD using personal HINs is feasible and a valid way to obtain hospitalization information about people with primary cancers.
- Linkage rates reflected differences across provinces in records submitted to the DAD and in clinical practice for the four most commonly diagnosed cancers: prostate, female breast, colorectal and lung cancer.
- Among linked records, agreement on basic identifiers was high, and the occurrence of consistent diagnoses and out-of-province hospitalizations was in line with expectations.

Table 6
Agreement on sex and date of birth between Discharge Abstract Database (DAD) and Canadian Cancer Registry (CCR) for DAD records linking during follow-up, by type of cancer and province (excluding Quebec), 2005 to 2008

Type of cancer and linkage characteristics	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia
Prostate									
Number of linked DAD records	4,032	1,003	7,727	5,779	35,771	4,403	7,789	9,359	22,391
% agreeing with CCR sex	100.0	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0
% agreeing with CCR date of birth	98.6	95.8	98.9	98.7	99.6	97.3	98.2	98.8	99.0
Female breast									
Number of linked DAD records	3,085	747	6,613	4,312	25,303	6,241	6,565	9,641	21,514
% agreeing with CCR sex	100.0	99.9	99.9	99.8	100.0	100.0	100.0	100.0	99.9
% agreeing with CCR date of birth	97.8	98.9	99.0	99.4	99.7	98.7	99.1	99.5	99.3
Colorectal									
Number of linked DAD records	7,800	1,240	9,579	4,631	49,317	9,752	8,940	12,478	29,306
% agreeing with CCR sex	99.9	99.9	100.0	99.3	100.0	99.9	100.0	100.0	99.8
% agreeing with CCR date of birth	98.6	97.2	99.0	99.4	99.7	98.1	97.6	99.3	99.1
Lung									
Number of linked DAD records	3,455	1,290	7,656	7,351	50,168	7,391	7,374	12,191	24,416
% agreeing with CCR sex	99.7	99.8	99.9	99.9	100.0	99.9	100.0	100.0	99.8
% agreeing with CCR date of birth	97.5	96.4	98.8	98.8	99.6	97.2	97.6	99.0	99.0

Note: Follow-up period for colorectal, lung and female breast cancers extended from 31 days before to 365 days after CCR date of diagnosis; for prostate cancer, follow-up extended from 31 days before to 730 days after date of diagnosis.
Sources: Canadian Cancer Registry data file, April 4, 2012; Discharge Abstract Database, fiscal years 2004/2005 to 2010/2011.

Table 7
Out-of-province admissions and agreement on diagnosis for Canadian Cancer Registry (CCR) cancers linking to Discharge Abstract Database (DAD) during follow-up, by province (excluding Quebec), 2005 to 2008

Type of cancer and linkage characteristics	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia
Prostate									
Number of cancers linking to at least one DAD record	1,425	447	2,852	2,347	22,911	1,978	2,807	5,442	9,973
% of linked cancers with at least one out-of-province admission	4.5	27.7	7.5	1.5	0.3	1.6	9.4	1.5	1.2
% of linked cancers with at least one DAD record documenting cancer diagnosis	83.7	81.2	87.8	81.6	83.1	79.1	85.2	84.9	77.4
Female breast									
Number of cancers linking to at least one DAD record	1,264	358	2,792	2,088	17,800	2,985	2,601	6,608	10,642
% of linked cancers with at least one out-of-province admission	0.5	2.8	1.7	0.8	0.4	0.7	2.2	0.9	0.6
% of linked cancers with at least one DAD record documenting cancer diagnosis	95.7	97.2	96.6	94.9	86.8	95.9	96.6	93.3	95.8
Colorectal									
Number of cancers linking to at least one DAD record	1,842	411	3,146	2,042	27,767	3,069	2,674	6,265	9,945
% of linked cancers with at least one out-of-province admission	1.4	10.2	0.8	3.1	0.3	1.5	3.5	1.3	0.9
% of linked cancers with at least one DAD record documenting cancer diagnosis	96.6	96.6	96.9	94.6	96.1	97.2	97.7	96.6	97.1
Lung									
Number of cancers linking to at least one DAD record	1,233	448	3,251	2,659	26,251	3,096	2,648	5,981	10,245
% of linked cancers with at least one out-of-province admission	1.1	27.0	3.1	1.0	0.4	1.0	2.6	1.5	0.9
% of linked cancers with at least one DAD record documenting cancer diagnosis	91.0	95.5	93.7	90.3	93.6	93.5	92.9	93.4	91.5

Notes: Follow-up period for colorectal, lung and female breast cancers extended from 31 days before to 365 days after CCR date of diagnosis; for prostate cancer, follow-up extended from 31 days before to 730 days after date of diagnosis.
Sources: Canadian Cancer Registry data file, April 4, 2012; Discharge Abstract Database, fiscal years 2004/2005 to 2010/2011.

cancers were most likely (94.6% to 97.7%) to link to at least one DAD record with a consistent cancer diagnosis.

Discussion

Deterministic linkage to the DAD is feasible for 8 of the 10 provinces in the CCR, because a high percentage of registered tumours have a valid HIN. For Ontario and Manitoba, direct linkages using HINs are not feasible, but a probabilistic linkage to the provincial health insurance registries obtained HINs for more than 98% of tumours in the CCR from 2005 through 2008, which made deterministic linkage to the DAD feasible. Nonetheless, because this study examined the validity of HINs in the CCR, results for Ontario and Manitoba should not be interpreted as the validity of HINs in the respective cancer registries.

The agreement on sex and date of birth among linked CCR-DAD records and patterns in the results provide construct validity for deterministic linkages using the HIN.

First, cancers for which surgery is the preferred treatment (female breast, colorectal and lung)[18] had high linkage rates and were the most likely to have a cancer diagnosis consistent with the CCR on at least one linked DAD record. Conversely, despite a follow-up period that was twice as long, the lowest linkage rate was for prostate cancer, optimal treatment of which is debated, taking account of life expectancy at diagnosis, the likelihood of the cancer causing problems, and side effects of treatment.[19] Prostate cancers were the least likely to have a consistent cancer diagnosis on at least one linked DAD record.

Second, Ontario, the province with the lowest overall linkage rate for female breast cancer, is also the most likely to perform mastectomies as day surgeries, which are not submitted to the DAD.[20]

Last, the percentage of linked cancers with at least one out-of-province admission varied by cancer type and by provincial population, suggesting that certain treatments may not be available nearby within some provinces.

Limitations

The main limitations of this research are differences across provinces in records submitted to the DAD and different clinical practice patterns in performing interventions on an inpatient, day surgery, or outpatient basis. As demonstrated with female breast cancer, even when the institution type is limited to acute care, linkage rates vary substantially. Another limitation is the exclusion of Quebec[11] and the territories from the linkage.

Conclusions

Personal HINs can be used to link the CCR and the DAD to obtain hospitalization information about people with primary cancers. Provincial variations in linkage rates of the four most commonly diagnosed cancers reflect differences in records submitted to the DAD and in clinical practice. Among linked records, agreement on basic identifiers was high. As more interventions are performed on a day surgery/outpatient basis and as more provinces/territories submit such records to the NACRS, combining data from multiple sources (for example, DAD, NACRS, physician billing databases) will be important in studying the health care experiences of people with cancer. Finally, if information about date of death or method of diagnostic confirmation is available, researchers may consider adjusting the follow-up period for cancers diagnosed at death to one year before the diagnosis date (date of death) to increase the potential for linkage. ■

References

1. Pinfold SP, Goel V, Sawka C. Quality of hospital discharge and physician data for type of breast cancer surgery. *Medical Care* 2000; 38(1): 99-107.
2. Turner D, Hildebrand KJ, Fradette K, Latosinsky S. Same question, different data source, different answers? Data source agreement for surgical procedures on women with breast cancer. *Healthcare Policy* 2007; 3(1): 46-54.
3. McBride M, Rogers PC, Sheps SB, et al. Childhood, adolescent and young adult cancer survivors research program of British Columbia: Objectives, study design and cohort characteristics. *Pediatric Blood and Cancer* 2010; 55(2): 324-30.
4. Statistics Canada. Canadian Cancer Registry (CCR) [cited April 25, 2014]. Available at: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3207&Item_Id=1633&lang=en
5. Peters PA, Tjepkema M, Wilkins R, et al. Data resource profile: 1991 Canadian Census Cohort. *International Journal of Epidemiology* 2013; 42(5): 1319-26.
6. Iron K, Zagorski BM, Sykora K, Manuel DG. *Living and Dying in Ontario: An Opportunity for Improved Health Information*. ICES Investigative Report. Toronto: Institute for Clinical Evaluative Sciences, 2008.
7. Fransoo R, Martens P, Prior H, et al. *Adult Obesity in Manitoba: Prevalence, Associations, and Outcomes*. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2011.
8. Manual DG, Perez R, Bennett C, et al. Seven more years: the impact of smoking, alcohol, diet, physical activity and stress on health and life expectancy in Ontario. An *ICES/PHO Report*. Toronto: Clinical Evaluative Sciences and Public Health Ontario, 2012.
9. Statistics Canada. Table 103-0550 – New cases for ICD-O-3 primary sites of cancer (based on July 2011 CCR tabulation file), by age group and sex, Canada, provinces and territories, annual, CANSIM (database) [cited April 25, 2014]. Available at: http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=1030550&paSer=&pattern=&stByVal=1&p1=1&p2=-1&tabMode=dataTable&csid=
10. Canadian Institute for Health Information. *Discharge Abstract Database* [cited December 1, 2013]. Available at: http://www.cihi.ca/CIHI-ext-portal/internet/en/document/types+of+care/hospital+care/acute+care/dad_metadata#
11. Canadian Institute for Health Information. *Data Quality Documentation, Hospital Morbidity Database-Multi-year Information.* Ottawa: Canadian Institute for Health Information, 2013.
12. Canadian Institute for Health Information. *International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Canada*. Ottawa: Canadian Institute for Health Information, 2003 and 2006.
13. Canadian Institute for Health Information. *Canadian Classification of Health Interventions*. Ottawa: Canadian Institute of Health Information, 2003 and 2006.
14. National Cancer Institute Surveillance, Epidemiology and End Results Program. Site Recode ICD-O-3/WHO 2008 Definition [cited April 25, 2014]. Available at: http://seer.cancer.gov/siterecode/icdo3_dwhoheme/
15. Fritz A, Percy C, Jack A, et al., eds. *International Classification of Diseases for Oncology, Third Edition, First Revision*. Geneva, Switzerland: World Health Organization, 2013.
16. Canadian Institute for Health Information. *Data Quality Documentation for External Users: Discharge Abstract Database, 2010-2011*. Ottawa: Canadian Institute for Health Information, 2011.
17. Canadian Institute for Health Information. *Executive Summary: Data Quality Documentation, Discharge Abstract Database, 2006-2007*. Ottawa: Canadian Institute for Health Information, 2007.
18. Canadian Institute for Health Information. *Data Quality Documentation, Discharge Abstract Database-Multi-year Information.* Ottawa: Canadian Institute for Health Information, 2012.
19. National Comprehensive Cancer Network (NCCN). *NCCN Guidelines* [cited April 25, 2014]. Available at: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp
20. Canadian Partnership against Cancer. *The 2014 Cancer System Performance Report.* Toronto: Canadian Partnership against Cancer, 2014.