

NELS 3x3¹ Data Dictionary

As of February 2012

Prepared by
Martha Cox, Analyst

In collaboration with
Lynn Lethbridge, NELS ICE Analyst
lynn.lethbridge@dal.ca

and

Ron Dewar, Surveillance and Epidemiology Unit (SEU),
Cancer Care Nova Scotia (CCNS), Ron.Dewar@ccns.nshealth.ca

For

Grace Johnston, Professor, Dalhousie University and SEU CCNS Epidemiologist

The 3x3 NELS research project was supported by NELS with funding from the Canadian Institutes of Health Research through a strategic initiative grant for “Interdisciplinary Capacity Enhancement (ICE) Reducing Health Disparities and Promoting Equity for Vulnerable Populations”.

¹ For the 3x3 NELS study protocol and findings from this study, see <http://www.dal.ca/sites/nels/research/nels/3x3.html>

Table of Contents

Introduction	3
Data Set Descriptions	4
Crosswalk	4
Vital Statistics	5
Cape Breton Palliative Care	8
Colchester – East Hants Palliative Care	11
Capital Health Palliative Care	13
CCNS Registry	15
<i>Consultations Data</i>	15
<i>Disease Data</i>	15
<i>Radiotherapy Data</i>	16
Cardiovascular Registry	17
Diabetes Registry	19
Appendices	21
Appendix 1: ICD codes used for categorizing diagnoses and cause of death	21
Appendix 2: Text strings used to collapse diagnoses as free text into categories	22
Appendix 3: Text strings used to collapse physician specialties into categories	23

Introduction

The NELS 3x3 Database consists of eleven separate SAS data sets.

- Three of these contain data from palliative care programs in Capital, Cape Breton, and Colchester-East Hants District Health Authorities. These data sets are named *capital* [sic], *cape_breton*, and *colchester*, respectively.
- Five data sets contain data from provincial disease surveillance programs. Three are from Cancer Care Nova Scotia (CCNS): *ccns_consults*, *ccns_disease*, and *ccns_rt*. Cardiovascular Health Nova Scotia (CVHNS) provided the *cvorig* data set, and *origdiab* contains data from the Diabetes Care Program of Nova Scotia (DCPNS).
- In addition, data was received from Nova Scotia Vital Statistics and stored as data set *vs*. Later updates to this data are found in the data set *new_causes*. The latter will not be described here. Instead, descriptions of the *vs* data set reflect revisions incorporating *new_causes*.
- Finally, the linkage analyst in CCNS/SEU (Ron Dewar) has provided a crosswalk data set, *vs_id*, containing the patient identifiers from Vital Stats, CVHNS, and DCPNS. This is used to match patients' records across these files. Data from CCNS already contains the Vital Stats patient identifier.

Each of these data sets is described in detail on the following pages as of February 2012. Martha Cox prepared this data dictionary based on the raw 3x3 NELS study analysis data provided to her by Ron Dewar and the cardiovascular and diabetes programs. Martha worked along with Lynn Lethbridge who prepared subsets of these data for analysis using SAS 9.2 for production of posters, reports, PPT presentations and papers for publication. The files described herein were linked using a unique study numbers for each study subject created for this 3x3 NELS study analysis. The 3x3 NELS study protocol and findings from this study, see <http://www.dal.ca/sites/nels/research/nels/3x3.html>. There were no identifiable personal identifiers in this 3x3 NELS study analysis dataset. Ron was the probabilistic linkage analyst and he was the only person with the identifier information, and he did not have access to the full analysis dataset. Research Ethics Board approval for this work was provided by the Capital Health, Cape Breton and Colchester East Hants district health authorities. Data access agreements were arranged with each of the seven data providers.

Data Set Descriptions

Crosswalk to link 3x3 NELS study data provided by the provincial cardiovascular and diabetes programs with the linked data provided by Ron Dewar which included 3x3 NELS study data from the cancer registry, Nova Scotia Vital Statistics, and Palliative Care Program data from three district health authorities: Capital Health, Cape Breton, and Colchester East Hants

Cross walk file to link VS to Diabetes and CV

Data set name = **vs_id**

Total Observations: 31,020

Variable	Description	Type/ Length	Values and their meanings ²	Frequency (%)
vs_id	Study ID variable for data linkage to all other datasets	Num 8	Range: 1000003 - 1121458 missing	31,020 (100.00) 0 (0.00)
cardio_id	Study ID variable for data linkage to VS data	Num 8	Range: 3000003 - 3060856 missing	24,909 (80.30) 6,111 (19.70)
diabetes_id	Study ID variable for data linkage to VS data	Num 8	Range: 4000010 - 4065536 missing	10,471 (33.76) 20,549 (66.24)

² Missing values for these variables simply means that there are more patients in Vital Statistics than in the Registry, not that there are Registry patients with missing ID numbers.

*

Nova Scotia Vital Statistics data fields for 3x3 NELS study analysis

Data set name = **vs**

Total Observations (Deaths 1995-2009): 121,458

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
vs_id	Study ID variable to link to VS	Num 8	Range: 1000001 – 1121460 missing	121,458 (100.00) 0 (0.00)
death	Date of death	Date (Num 8)	Range: 01JAN1995 - 31DEC2009 missing	121,458 (100.00) 0 (0.00)
birth	Date of birth	Date (Num 8)	Range: 19SEP1813 - 13DEC2009 missing	121,458 (99.86) 176 (0.14)
Birth_source	Dataset source for birth date (for those not VS)	Char 4	CCNS.....From Cancer Registry missing.....From Vital Stats	10 (0.01) 121,448 (99.99)
sex	Sex	Char 2	F.....Female M.....Male U.....Unknown	60,117 (49.50) 61,319 (50.49) 22 (0.02)
pcode	Residential postal code	Char 8	<i>Values are 3-8 alpha-numeric characters, which can be categorized³ as :</i> Probably good ^{4,5} Invalid characters Probably US zip codes Incomplete or too long missing	115,157 (94.81) 103 (0.08) 11 (0.01) 9 (0.01) 6,178 (5.09)
Pcode_source	Dataset source for postal code (for those not VS)	Char 4	CCNS.....From Cancer Registry postFrom a PCP, Cardio, or Diabetes missing.....From Vital Stats	3,877 (3.19) 1,458 (1.20) 116,123 (95.61)
Death_pcode	Death location postal code	Char 6	<i>Values are 6 alpha-numeric characters, which can be categorized² as :</i> Probably good ³ Invalid characters, looks like a PC Invalid, looks like part of city name missing	40,615 (33.44) 106 (0.09) 739 (0.61) 79,998 (65.86)

³ Rules for testing postal codes:

- First character must be a letter, other than D, F, I, O, Q, U, W and Z.
- Characters 3 and 5 must be a letter, other than D, F, I, O, Q, and U.
- Characters 2, 4 and 6 must be a numeric value (0-9).

⁴ “Probably good” means that the value has 6 characters which satisfy the Canada Post rules for each individual character, i.e., that position contains only the letters or numbers allowed. A given value still may not be a valid postal code.

⁵ Includes 7 observations which begin with ‘X’. Although they may be valid Northwest Territories postal codes, in practice in Nova Scotia, “X9X9X9” is often using for missing values.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
County	County of Resident	Char 16	ANNAPOLIS COUNTY ANTIGONISH COUNT CAPE BRETON COUN COLCHESTER COUNT CUMBERLAND COUNT DIGBY COUNTY GUYSBOROUGH COUN HALIFAX COUNTY HANTS COUNTY INVERNESS COUNTY KINGS COUNTY LUNENBURG COUNTY PICTOU COUNTY QUEENS COUNTY RICHMOND COUNTY SHELBURNE COUNTY VICTORIA COUNTY YARMOUTH COUNTY missing	3,950 (3.25) 2,356 (1.94) 17,662 (14.54) 6,723 (5.54) 5,515 (4.54) 3,432 (2.83) 1,710 (1.41) 37,058 (30.51) 4,929 (4.06) 3,017 (2.48) 6,968 (5.74) 7,273 (5.99) 7,327 (6.03) 2,119 (1.74) 1,665 (1.37) 2,307 (1.90) 1,381 (1.14) 4,064 (3.35) 2,002 (1.65)
Place_of_residence ⁶	Nursing home residence indicator	Char 16	Nursing home Others missing	27,986 (23.04) 80,290 (66.11) 13,182 (10.85)
Place_of_death ⁷	Nursing home indicator for location of death	Char 16	Hospital Nursing home Others missing	75,370 (62.05) 20,957 (17.25) 25,044 (20.62) 87 (0.07)
Cause_of_death	All causes of death	Char 32	<i>Each value consists of a list of ICD-9 codes separated by commas. These can be roughly collapsed⁸ as follows:</i> Cancer Congestive Heart Failure Acute MI Chronic Ischemic Heart Disease Unstable Angina Diabetes Other missing <i>Note: Patients can be in more than one category.</i>	12,632 (10.40) 5,287 (4.35) 4,460 (3.67) 7,571 (6.23) 0 (0.00) 3,492 (2.88) 95,367 (78.52) 307 (0.25)
is_hcn	HCN comes from VS	Num 8	1Yes 0No	100,959 (83.12) 20,499 (16.88)

⁶ Created by CCNS/SEU comparing the full residential address provided by Vital Stats to a list of addresses for senior care facilities.

⁷ Created by CCNS/SEU comparing the full address of place of death provided by Vital Stats to a list of addresses for senior care facilities.

⁸ See Appendix 1 for the list of ICD codes used for these categories.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
hcn_source	Dataset source for HCN (for those not from VS)	Char 4	CBPC..... From Cape Breton PCP CCNS..... From Cancer Registry CEH..... From Colchester PCP CHPC..... From Capital PCP Card..... From Cardio Registry missing..... From Vital Stats	43 (0.04) 7,088 (5.84) 13 (0.01) 154 (0.13) 1,453 (1.20) 112,707 (92.80)
is_nursing ⁹	Nursing home indicator	Num 8	1.....Yes 0.....No	1,547 (1.27) 119,911 (98.73)
Extended_care ¹⁰	Alternate nursing home flag.	Char 3	Yes missing	8,983 (7.40) 112,475 (92.60)

⁹ Created by CCNS/SEU using the “Nursing Home” field provided by Vital Stats. This field was introduced in mid-2008 and contained the name of the facility. The data is more complete for 2009. SEU has converted this to a YES/NO value.

¹⁰ Field provided by Vital Stats as a YES/NO value. Values are based on full addresses, but not sure of the source of the list. Some values are present for 1995-2002; more complete for 2003+.

Cape Breton Palliative Care Program data fields for 3x3 NELS study analysis

Data set name = **cape_breton**

Total individuals linked to Vital Statistics: 5,955 (total observations: 5,955)

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
vs_id	Study ID variable to link to VS	Num 8	Range: 1000024 - 1121460 missing	5,955 (100.00) 0 (0.00)
pcode	Residential postal code	Char 255	<i>Values are 0-6 alpha-numeric characters, which can be categorized as¹¹:</i> Probably good ^{12,13} Incomplete or too long Invalid characters missing	5,822 (97.77) 8 (0.13) 24 (0.40) 101 (1.70)
diagnoses	Diagnoses at enrollment	Char 64	<i>Almost 1500 unique, non-missing values, entered as free text and listing one or more conditions. These can be roughly collapsed¹⁴ as follows:</i> Cancer COPD Congestive Heart Failure Stroke Renal Failure Alzheimers/Dementia Diabetes Acute MI Other <i>Note: Patients can be in more than one category.</i>	4,664 (78.32) 287 (4.82) 200 (3.36) 157 (2.64) 154 (2.59) 130 (2.18) 14 (0.24) 7 (0.12) 465 (7.81)
First_refer	Date of first referral	Date (Num 8)	Range: 22DEC1992 - 31DEC2011 missing	5,932 (99.61) 23 (0.39)

¹¹ Rules for testing postal codes:

-- First character must be a letter, other than D, F, I, O, Q, U, W and Z.

-- Characters 3 and 5 must be a letter, other than D, F, I, O, Q, and U.

-- Characters 2, 4 and 6 must be a numeric value (0-9).

¹² "Probably good" means that the value has 6 characters which satisfy the Canada Post rules for each individual character, i.e., that position contains only the letters or numbers allowed. A given value still may not be a valid postal code.

¹³ Includes 7 observations which begin with 'X'. Although they may be valid Northwest Territories postal codes, in practice in Nova Scotia, "X9X9X9" is often using for missing values.

¹⁴ The free text values were assigned to these categories by searching for key words and phrases. See Appendix 2 for a list of the text strings used for each category.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
Place_of_referral	Place of referral	Char 255	<i>Multiple values which collapse to:</i> Home (HOME, Home) Hospital (BADDECK HOSPITAL, HOSPITAL, HOSPITLA, Hospital, ONTARIO HOSPITAL) Nursing home (NSG HOME, NSHF-LTC, NURSING HOME) missing	2,561 (43.01) 3,166 (53.17) 220 (3.69) 8 (0.13)
status	Specialty of referring physiciain	Char 255	<i>202 unique, non-missing values, entered as free text, which can be roughly collapsed¹⁵ as follows:</i> Physician: Fam, Peds, or Geriatrics Physician: Emergency Medicine Physician: Surgeon Physician: Medical Oncology Physician: Radiation Oncology Physician: Palliative Medicine Physician: Other Specialist Physician: Unspecified Nurse Family or friend Self, other program, or invalid value missing	2,899 (48.68) 60 (1.01) 387 (6.50) 1,136 (19.08) 143 (2.40) 21 (0.35) 618 (10.38) 449 (7.54) 94 (1.58) 85 (1.43) 47 (0.80) 16 (0.27)
Followed_by_ CB_Cancer	Patient id followed by CB cancer centre	Num 8	1 Yes 0 No missing	1,320 (22.17) 4,627 (77.70) 8 (0.13)
Discharge_date	Date of patient discharged or died	Date (Num 8)	Range: 01NOV1994 - 01JAN2010 missing	5,947 (99.87) 8 (0.13)
Closed	Patient's chart has been closed	Num 8	1 Yes 0 No missing	720 (12.09) 5,227 (87.77) 8 (0.13)
reason	Reason for closing chart	Char 255	Closed Condition Stable DEATH Family Refused Service Refused Service Relocated missing	xxx ¹⁶ 450 (7.56) xxx 7 (0.12) 41 (0.69) 217 (3.64) 5,236 (87.93)

¹⁵ The free text values were assigned to these categories by searching for key words and phrases. See Appendix 3 for a list of the text strings used for each category.

¹⁶ Values are masked for cells with less than 5.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
Location	Location of death	Char 255	<i>113 unique, non-missing values, which roughly collapse as:</i> At home In hospital Nursing home missing	1,282 (21.53) 3,513 (58.99) 436 (7.32) 724 (12.16)
Days_in_service	Number of days followed by PCS	Char 255	Range: 0 - 1337 missing	1,515 (25.44) 4,440 (74.56)
Institution	Institution patient is referred from	Char 255	Hospital Nursing home missing	2,258 (37.92) 1,130 (18.98) 2,567 (43.11)
death	Indicated patient died	Num 8	1 Yes 0 No missing	5,492 (92.23) 463 (7.77) 0 (0.00)

Colchester – East Hants Palliative Care Program data for 3x3 NELS study analysis

Data set name = **colchester**

Total individuals linked to Vital Statistics: 2,401 (total observations: 2,401)

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
vs_id	Study ID variable to link to VS	Num 8	Range: 1000056 - 1121451 missing	2,401 (100.00) 0 (0.00)
first_visit	Date of first visit	Date (Num 8)	Range: 20MAR1997 - 18FEB2010 missing	2,388 (99.46) 13 (0.54)
Postal_code	Residential postal code	Char 7	<i>Values are 6 alpha-numeric characters, which can be categorized as:</i> Probably good ^{17,18} Invalid ¹⁹ or missing	1,602 (66.72) 799 (33.27)
County	Residential county	Char 20	Colchester Cumberland East Hants Halifax Pictou Invalid or missing	1,923 (80.09) 12 (0.50) 444 (18.49) 10 (0.42) 9 (0.37) 3 (0.12)
Resident_type	Residence type	Char 50	Apartment/Condominium Boarding Room Home for Special Care House Nursing Home Senior's Housing Other missing	189 (7.87) 12 (0.50) 36 (1.50) 1,508 (62.81) 104 (4.33) 55 (2.29) 17 (0.71) 480 (19.99)
discharge	Discharge date	Date (Num 8)	Range: 26AUG2002 - 01OCT2009 missing	160 (6.66) 2,241 (93.34)
PCP_death_date	PCP date of death	Date (Num 8)	Range: 12APR2002 - 11MAR2010 missing	2,241 (93.34) 160 (6.66)
With_mets	Patient has cancer with metastasis	Num 8	1 Yes 0 No	541 (22.53) 1,860 (77.47)

¹⁷ “Probably good” means that the value has 6 characters which satisfy the Canada Post rules for each individual character, i.e., that position contains only the letters or numbers allowed. A given value still may not be a valid postal code.

¹⁸ Includes 7 observations which begin with ‘X’. Although they may be valid Northwest Territories postal codes, in practice in Nova Scotia, “X9X9X9” is often using for missing values.

¹⁹ Rules for testing postal codes:

- First character must be a letter, other than D, F, I, O, Q, U, W and Z.
- Characters 3 and 5 must be a letter, other than D, F, I, O, Q, and U.
- Characters 2, 4 and 6 must be a numeric value (0-9).

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
Reason_for_Discharge	Reason for Discharge	Char 50	DBBS..... died before being seen Deceased... deceased NP not palliative OOD Reloc..... relocated YES..... missing	16 (0.67) 421 (17.53) xxx ²⁰ 13 (0.54) xxx 30 (1.25) 1,913 (79.68)
diagnoses ²¹	Diagnoses at PCP enrollment	Char 64	207 unique, non-missing values, entered as free text and listing one or more conditions, separated by commas. These roughly collapse ²² as follows: Cancer Congestive Heart Failure COPD Renal Failure Stroke Alzheimers/Dementia Other <i>Note: Patients can be in more than one category.</i>	1,641 (68.35) 36 (1.50) 102 (4.25) 86 (3.58) 64 (2.67) 55 (2.29) 283 (11.78)
death_place ²³	Place of death	Char 16	Home Hopsital [sic] ukn ²⁴	722 (30.07) 1,518 (63.22) 161 (6.71)
first_refer	Date of first referral	Date (Num 8)	Range: 20MAR1997 - 18FEB2010 missing	2,401 (100.00) 0 (0.00)
Admission_date	Admission date	Date (Num 8)	Range: 08JUL2000 - 10FEB2010 missing	1,915 (79.76) 486 (20.24)
Type	Type of admission to hospital	Char 50	--- Direct ER Admit ER only missing	515 (21.45) 480 (19.99) 571 (23.78) 464 (19.33) 371 (15.45)
Place_discharge	Place client was discharged to	Char 50	HOME or FAMILY other missing	44 (1.84) 7 (0.29) 2,350 (97.88)

²⁰ Values are masked for cells with less than 5.

²¹ Values were converted to uppercase before analysis.

²² The free text values were assigned to these categories by searching for key words and phrases. See Appendix 2 for a list of the text strings used for each category.

²³ Values were converted to uppercase before analysis.

²⁴ Missings coded as 'ukn' by the linkage analyst.

Capital Health Palliative Care Program data fields for 3x3 NELS study analysis

Data set name = **capital** [sic]

Total individuals linked to Vital Statistics: 12,976 (total observations: 12,976)

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
vs_id	Study ID variable to link to VS	Num 8	Range: 1000004 - 1121456 missing	12,976 (100.00) 0 (0.00)
first_visit	Date of first visit	Date (Num 8)	Range: 18JUN1984 - 30DEC2009 missing	12,976 (100.00) 0 (0.00)
Postal_code	Residential postal code	Char 7	Values are 5-7 alpha-numeric characters, which can be categorized as: Probably good ^{25,26} Invalid characters ²⁷ Incomplete, too long, or missing ²⁸	12,568 (96.86) 22 (0.17) 386 (2.97)
Primary_diagnosis	Diagnosis at enrollment; ICD-9 codings with decimals	Char 7	413 unique, non-missing values, entered as a single ICD-9 code. These can be roughly collapsed ²⁹ as follows: Cancer Congestive Heart Failure Acute MI Chronic Ischemic Heart Disease Diabetes Other ³⁰ Note: Patients can be in only one category.	10,935 (84.27) 250 (1.93) 27 (0.21) 26 (0.20) 15 (0.12) 1,723 (13.28)

²⁵ “Probably good” means that the value has 6 characters which satisfy the Canada Post rules for each individual character, i.e., that position contains only the letters or numbers allowed. A given value still may not be a valid postal code.

²⁶ Includes 7 observations which begin with ‘X’. Although they may be valid Northwest Territories postal codes, in practice in Nova Scotia, “X9X9X9” is often using for missing values.

²⁷ Rules for testing postal codes:

-- First character must be a letter, other than D, F, I, O, Q, U, W and Z.

-- Characters 3 and 5 must be a letter, other than D, F, I, O, Q, and U.

-- Characters 2, 4 and 6 must be a numeric value (0-9).

²⁸ There are 36 observations which have ‘missing’ entries (i.e., starting with ‘X’).

²⁹ See Appendix 1 for the list of ICD codes used for these categories.

³⁰ There are 107 observations which contain a missing entry of ‘999.’

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
Comorbidity	Comorbidities	Char 255	<p>1998 unique, non-missing values, entered as one or more ICD-9 codes, separated by semi-colons. These can be roughly collapsed³¹ as follows:</p> <p>Cancer Congestive Heart Failure Acute MI Chronic Ischemic Heart Disease Diabetes Other</p> <p><i>Note: Patients can be in more than one category.</i></p>	<p>5,467 (42.13) 299 (2.30) 96 (0.74) 317 (2.44) 648 (4.99) 667 (5.14)</p>
Discharge	Date of discharge	Date (Num 8)	Range: 12MAR1987 - 07FEB2010 missing	<p>3,818 (29.42) 9,158 (70.58)</p>
first_refer	Date of first referral	Date (Num 8)	Range: 01MAY2003 - 30DEC2009 missing	<p>5,449 (41.99) 7,527 (58.01)</p>

³¹ See Appendix 1 for the list of ICD codes used for these categories.

Cancer Care Nova Scotia Registry data fields for 3x3 NELS study analysis

Consultations Data

Data set name = **ccns_consults**

Total individuals linked to Vital Statistics: 21,492

(As individuals can have more than 1 observation, there are 35,060 total observations.)

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
vs_id	Study ID variable to link to VS	Num 8	Range: 1000019 - 1121459 missing	35,060 (100.00) 0 (0.00)
Consult	Date of consultation	Date (Num 8)	Range: 04JAN1994 - 22DEC2009 missing	35,060 (100.00) 0 (0.00)
Department	Department consulted	Char 4	MEDO Medical Oncology RAD..... Radiation Oncology	16,222 (46.27) 18,838 (53.73)

Disease Data

Data set name = **ccns_disease**

Total individuals linked to Vital Statistics: 52,989

(As individuals can have more than 1 observation, there are 66,368 observations.)

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
vs_id	Study ID variable to link to VS	Num 8	Range: 1000001 - 1121460 missing	66,368 (100.00) 0 (0.00)
disease_no	Index of primary cancers for this patient	Num 8	Range: 1 - 41 missing	66,085 (99.57) 283 (0.43)
Site	Type of cancer	Char 6	<i>667 unique, non-missing values, entered as a single code using the coding system identified in the site_rev variable.</i> value present missing	66,085 (99.57) 283 (0.43)
site_rev ³²	Coding system revision	Char 3	2 ICD-O-2 3 ICD-O-3 8 ICD-8 O ICD-O (v1) missing	28,071 (42.30) 22,842 (34.42) 2,743 (4.13) 12,429 (18.73) 283 (0.43)

³² ICD-O (v1) and ICD-8 codes usually begin with a number and look similar to ICD-9 codes.
ICD-O-2 and ICD-O-3 codes usually begin with a letter and look similar to ICD-10 codes.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
Histo	Histology Code	Char 6	<i>1291 unique, non-missing values, entered as a single code using the coding system identified in the site_rev variable.</i> value present missing ³³	64,368 (97.71) 1,517 (2.29)
Diag	First diagnosis date	Date (Num 8)	Range: 01SEP1941 - 31DEC2009 missing	66,085 (99.57) 283 (0.43)

Radiotherapy Data

Data set name = **ccns_rt**

Total individuals linked to Vital Statistics: 14,798

(As individuals can have more than 1 observation, there are 23,966 observations)

One observation is a treatment plan.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
vs_id	Study ID variable to link to VS	Num 8	Range: 1000026 - 1121459 missing	23,966 (100.00) 0 (0.00)
intent	Reason for radiation	Char 1	A Adjuvant C Cure P Palliative R Radical X Not Provided missing	698 (2.91) 425 (1.77) 15,343 (64.02) 5,099 (21.28) 2,396 (10.00) 5 (0.02)
Total_visits	Total visits for particular radiation treatment	Num 8	Range: 1 - 700 missing	23,023 (96.07) 943 (3.93)
Total_dose	Total radiation dose	Num 6	Range: 2 – 30,001 missing	23,777 (99.21) 189 (0.79)
start	Start date of radiation	Date (Num 8)	Range: 04JAN1993 - 22DEC2009 missing	23,966 (100.00) 0 (0.00)

³³ Records using ICD-8 (*site_rev=8*) will probably not have histology codes recorded.

Provincial Cardiovascular Registry data fields for 3x3 NELS study analysis

Data set name = **cvorig**

Total individuals linked to Vital Statistics: 24,909 (total observations: 48,158)

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
ID	Study ID variable to link to VS. Same as "cardio_id" in vs_id dataset	Num 8	Range: 3000003 - 3060856 missing	48,158 (100.00) 0 (0.00)
DOD	Date of death	Date (Num 8)	Range: 01APR1997 - 31DEC2009 missing	48,158 (100.00) 0 (0.00)
HospAdmitDate	Date of admission	Date (Num 8)	Range: 15OCT1997 - 29DEC2009 missing	48,158 (100.00) 0 (0.00)
DischDate	Date of discharge	Date (Num 8)	Range: 16OCT1997 - 31DEC2009 missing	48,158 (100.00) 0 (0.00)
AdmDXAMI* ³⁴	Admitted with acute myocardial infarction diagnosis	Char 1	Y..... Yes missing No	9,159 (19.02) 38,999 (80.98)
AdmDXCHF*	Admitted with congestive heart failure diagnosis	Char 1	Y..... Yes missing No	21,340 (44.31) 26,818 (55.69)
AdmDXUA*	Admitted with unstable diagnosis	Char 1	Y..... Yes missing No	8,621 (17.90) 39,537 (82.10)
AdmDXIHD*	Admitted with ischemic heart disease	Char 1	Y..... Yes missing No	3,292 (6.84) 44,866 (93.16)
AdmDXAF*	Admitted with atrial fibrillation	Char 1	Y..... Yes missing No	4,449 (9.24) 43,709 (90.76)
DisDXAMI*	Discharged with acute myocardial infarction diagnosis	Char 1	Y..... Yes missing No	13,058 (27.11) 35,100 (72.89)
DisDXCHF*	Discharged with congestive heart failure diagnosis	Char 1	Y..... Yes missing No	26,816 (55.68) 21,342 (44.32)
DisDXUA*	Discharged with unstable diagnosis	Char 1	Y..... Yes missing No	8,337 (17.31) 39,821 (82.69)
DisDXIHD*	Discharged with ischemic heart disease	Char 1	Y..... Yes missing No	8,752 (18.17) 39,406 (81.83)
DisDXAF*	Discharged with atrial fibrillation	Char 1	Y..... Yes missing No	6,775 (14.07) 41,383 (85.93)
AdmChronic	Admitted from a chronic care facility (nursing home)	Char 1	Y..... Yes missing No	3,548 (7.37) 44,610 (92.63)

*³⁴ Most individuals get into the registry with an admission or discharge for one of these conditions. A small percentage (1.31%) are admitted for a variety of 'other' reasons which are not included in this dataset.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
DischChronic	Discharged to a chronic care facility (nursing home)	Char 1	Y..... Yes missing No	3,866 (8.03) 44,292 (91.97)
HL	Diagnosed with hyperlipidemia prior to admission	Char 1	Y..... Yes missing No	16,831 (34.95) 31,327 (65.05)
Diabetes	Diagnosed with diabetes during admission	Char 1	Y..... Yes missing No	17,968 (37.31) 30,190 (62.69)
HT	Diagnosed with hypertension during admission	Char 1	Y..... Yes missing No	27,438 (56.97) 20,720 (43.03)
PrevCHF	Previously had congestive heart failure	Char 1	Y..... Yes missing No	19,634 (40.77) 28,524 (59.23)
PrevMI	Previously had myocardial infarction	Char 1	Y..... Yes missing No	18,707 (38.85) 29,451 (61.15)
PrevArrest	Previously had a cardiac arrest	Char 1	Y..... Yes missing No	787 (1.63) 47,371 (98.37)
PrevAdmUA	Previous history of unstable angina	Char 1	Y..... Yes missing No	7,243 (15.04) 40,915 (84.96)
PrevCOPD	Previous history of COPD	Char 1	Y..... Yes missing No	13,319 (27.66) 34,839 (72.34)
PrevAsthma	Previous history of asthma	Char 1	Y..... Yes missing No	2,671 (5.55) 45,487 (94.45)
PrevRenalFailure	Previous history of renal failure	Char 1	Y..... Yes missing No	6,797 (14.11) 41,361 (85.89)
PrevStroke_TIA	Previously had a stroke or transient ischemic attack	Char 1	Y..... Yes missing No	8,059 (16.73) 40,099 (83.27)
PrevMalignancy	Previous malignancy	Char 1	Y..... Yes missing No	5,266 (10.93) 42,892 (89.07)
PrevPVD	Previous history of peripheral vascular disease	Char 1	Y..... Yes missing No	6,465 (13.42) 41,693 (86.58)

Provincial Diabetes Registry data fields for 3x3 NELS study analysis

Data set name = **origdiab**

Total individuals linked to Vital Statistics: 10,471 (total observations: 10,470)

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
study_id	Study ID variable to link to VS. Same as "diabetes_id" in vs_id dataset	Char 7	Range: 4000010- 4065536 missing	10,470 (100.00) 0 (0.00)
dod	Date of death	Date (Num 8)	Range: 09JAN1995 - 31DEC2009 missing	10,470 (100.00) 0 (0.00)
DMtype_first	Type of diabetes	Num 8	1T1 2T2 3PT1 4PT2 9O 99missing	422 (4.03) 9,730 (92.93) xxx ³⁵ xxx 50 (0.48) 265 (2.53)
ddiag_first	Date of diagnosis	Date (Num 8)	Range: 01SEP1922 - 18MAY2009 missing	8,478 (80.97) 1,992 (19.03)
dref_first	Date of referral	Date (Num 8)	Range: 10JAN1960 - 01OCT2009 missing	4,734 (45.21) 5,736 (54.79)
DMdvisit_first	Date of visit	Date (Num 8)	Range: 05OCT1902 - 03JUL2009 missing	10,205 (97.47) 265 (2.53)
DMdvisit_last	Date of visit	Date (Num 8)	Range: 10OCT1972 - 15DEC2009 missing	10,205 (97.47) 265 (2.53)
treat_last	Diabetes treatment	Num 8	1Diet 2Insulin 3OAA 4IN/OAA 9missing	2,523 (24.10) 2,154 (20.57) 4,623 (44.15) 564 (5.39) 606 (5.79)
mp_none_last	Medical problems: None	Char 1	T Yes F No missing	147 (1.40) 9,967 (95.20) 356 (3.40)
mp_incompl_last	Medical problems: Incomplete	Char 1	T Yes F No missing	2,057 (19.65) 8,083 (77.20) 330 (3.15)
mp_thyroid_last	Medical problems: Thyroid	Char 1	T Yes F No missing	819 (7.82) 9,308 (88.90) 343 (3.28)
mp_hypert_last	Medical problems: Hypertension	Char 1	T Yes F No missing	4,940 (47.18) 5,235 (50.00) 295 (2.82)
mp_dyslip_last	Medical problems: Dyslipidemia	Char 1	T Yes F No missing	3,369 (32.18) 6,775 (64.71) 326 (3.11)

³⁵ Values are masked for cells with less than 5.

Variable	Description	Type/ Length	Values and their meanings	Frequency (%)
mp_overw_last	Medical problems: Overweight (BMI>27)	Char 1	T Yes F..... No missing	3,398 (32.45) 6,748 (64.45) 324 (3.09)
mp_retin_last	Medical problems: Retinopathy	Char 1	T Yes F..... No missing	482 (4.60) 9,634 (92.02) 354 (3.38)
mp_nephro_last	Medical problems: Nephropathy	Char 1	T Yes F..... No missing	669 (6.39) 9,450 (90.26) 351 (3.35)
mp_neuro_last	Medical problems: Neuropathy	Char 1	T Yes F..... No missing	534 (5.10) 9,582 (91.52) 354 (3.38)
mp_cad_last	Medical problems: Coronary Artery Disease (CAD)	Char 1	T Yes F..... No missing	2,222 (21.22) 7,915 (75.60) 333 (3.18)
mp_cvd_last	Medical problems: Cerebrovascular Disease (CVD)	Char 1	T Yes F..... No missing	407 (3.89) 9,685 (92.50) 378 (3.61)
mp_pvd_last	Medical problems: Peripheral Vascular Disease	Char 1	T Yes F..... No missing	701 (6.70) 9,425 (90.02) 344 (3.29)
mp_other_last	Medical problems: Other	Char 1	T Yes F..... No missing	2,526 (24.13) 7,609 (72.67) 335 (3.20)
mp_other_m_last	Description of other medical problems	Char 254	207 unique, non-missing values, entered as free text and listing one or more conditions, separated by semi-colons. These can be roughly collapsed ³⁶ as follows: Cancer Congestive Heart Failure Acute MI COPD Renal Failure Stroke Alzheimers/Dementia Other <i>Note: Patients can be in more than one category.</i>	302 (2.88) 35 (0.33) 12 (0.11) 75 (0.72) 11 (0.11) 42 (0.40) 20 (0.19) 595 (5.68)

³⁶ The free text values were assigned to these categories by searching for key words and phrases. See Appendix 2 for a list of the text strings used for each category.

Appendices

Appendix 1: ICD codes used for categorizing diagnoses and cause of death

Program	Disease	ICD-9	ICD-10
CARDIOVASCULAR			
	Congestive Heart Failure	428.0	I50.0
	Acute MI	410	I21-I22
	Chronic Ischemic Heart Disease	412-414, 429.2	I25
	Unstable Angina	411.1	I20.0
DIABETES			
	Diabetes	250	E10-E14
CANCER			
	Cancer	140-208	C00-C97
RENAL			
	Renal Failure	584-586, 588, 589	N17-N19

Appendix 2: Text strings used to collapse diagnoses as free text into categories

Cancer

- 'CA'
- '?CA'
- 'CA.'
- 'CA-'
- 'CA/'
- 'CANCER'
- 'ASTROCYTOMA'
- 'CARCINOID'
- 'CARCINOMA'
- 'CARDOMA'
- 'CHONDROSARCOMA'
- 'CHORDOMA'
- 'GASTRINOMA'
- 'GIOBLASTOMA'
- 'GLIOBLASTOMA'
- 'GLIOMA'
- 'HEPATOMA'
- 'HODGKIN'
- 'LEUKEMIA'
- 'LYMPHOMA'
- 'MALIGNANCY'
- 'MALIGNANT'
- 'MASS'
- 'MELANOMA'
- 'MESOTHELIOMA'
- 'METS'
- '/METS'
- 'METASTATIC'
- 'METASTIC'
- 'MYELOMA'
- 'MYLEOMA'
- 'NEUROBLASTOMA'
- 'PRIMARY'
- 'SARCOMA'
- 'TUMOR'
- 'AML'

Congestive Heart Failure

- 'CHF'
- 'CHG'
- 'CONGESTIVE HEART FAILURE'

Acute MI

- 'M.I.'
- 'MI'
- 'MI,'
- 'MI;'
- 'MYOCARDIAL INFARCTION'

COPD

- 'COPD'

Diabetes

- 'DIABETES'

Renal Failure

- 'RENAL FAILURE'
- 'RENAL RAILURE'

Stroke

- 'CVA'
- 'C.V.A.'
- 'STROKE'
- 'TIA'
- "TIA'S"
- 'INTRACEREBRAL BLEED'
- 'INTRACRANIAL BLEED'
- 'CEREBRAL BLEED'

Alzheimers/Dementia

- 'ALZHEIMER'
- 'AZHEIMER'
- 'DEMENTIA'
- 'CREUTZFELDT-JACOB'
- 'CONFUSION'
- 'MENTALLY INCOMPETENT'

Other

[Any "word" in the value that has not been assigned to one of the categories above.]

Appendix 3: Text strings used to collapse physician specialties into categories

Self

- 'SELF'

Family or friend

- 'DAUGHTER'
- 'NEICE'
- 'SON-IN-LAW'
- begins with 'FAMILY' and is not 'FAMILY PHYSICIAN'
- begins with 'FRIEND'

Physician: Family, Peds, or Geriatrics

- 'FAMILY PHYSICIAN'
- 'FMED'
- 'FMED (PEI)'
- 'FMED-RESIDENT'
- 'ONTARIO (FMED)'
- 'PHY - FM'
- 'PHY - GM'
- 'PHY. - FAM'
- 'PHY. - FM'
- 'PHY. - FM RES'
- 'PHY. - FMED'
- 'PHY. - GM'
- 'PHY. FM'
- 'PHY.- FM'
- 'PHYSICIAN - FM'
- 'PHYSICIAN - GM'
- 'GE'
- 'GMED'
- 'PAED'
- 'PAEDS'
- 'PEDIATRICIAN'
- 'PEDS'
- 'PHY. - PEDS'

Physician: Emergency Medicine

- 'EM'
- 'EMED'
- 'EMER'
- 'PHY. - ER'
- 'PHY. MED.(E.R.)'

Physician: Surgeon

- 'GSURG'
- 'ORTHOPEDIC SURGEON'
- 'OSURG'
- 'PHY. - GS'
- 'PHY. - GSURG'
- 'PHY. GS'
- 'PHYSICIAN (SURGEON)'
- 'PHYSICIAN - GS'
- 'SURG'
- 'SURGEON'
- 'Surg'
- 'TSURG'
- 'VSURG'

Physician: Medical Oncology

- 'MED ONCOLOGIST'
- 'MED. ONCOLOGIST'
- 'MED. ONCOLOGISTS'
- 'MED. ONGOLOGIST'
- 'MONC'
- 'ONCOLOGIST'
- 'ONCOLOGIST, MED.'
- 'ONCOLOGIST/PHYSICI'
- 'ONCOLOGY PHYSICIAN'
- 'PHY. - MED. ONC.'
- 'PHY. - MED.ONC.'
- 'PHY. - MONC'
- 'PHY. MED.ONC.'
- 'PHY.- MED.ONC.'
- 'PHYSICIAN (ONCOLOG'
- 'PHYSICIAN - MED.ON'

Physician: Radiation Oncology

- 'PHY. - RAD.ONC.'
- 'PHY. - RONC'
- 'RAD. ONCOLOGIST'
- 'RADIATION ONCOLOGI'
- 'RONC'

(continued)

Physician: Palliative Medicine

- 'PALL MED'
- 'PALL MED (HAL)'
- 'PALL/MED'
- 'PALLMED'
- 'PC PHYSICIAN'
- 'PHY. - P.C.'
- 'PHY. - P.C. MED.'
- 'PHY. - PAIN'
- 'PHY. - PC.MED.'
- 'PMED'

Physician: Other Specialist

- 'ANAES'
- 'ANESTHESIOLOGIST'
- 'ANESTHETIST'
- 'CARD'
- 'DERM'
- 'DERMATOLOGIST'
- 'ENDO'
- 'ENT PHYSICIAN'
- 'HAEM'
- 'HEAM'
- 'ID'
- 'IMED'
- 'IMED, HALIFAX'
- 'INTERNIST'
- 'NEPH'
- 'NEUR'
- 'NEURO'
- 'O & G'
- 'O&G'
- 'ORTH'
- 'OTO'
- 'PHYSICIAN - RESP'
- 'PHY. - O&G'
- 'PHY. - ENDO'
- 'PHY. - GYNE'
- 'PHY. - IMED'
- 'PHY. - NSURG'
- 'PHY. - OBS.'
- 'PHY. - ORTH'
- 'PHY. - ORTH.'
- 'PHY. - PM&R'
- 'PHY. - PSY.'
- 'PHY. - RESP'
- 'PHY. - RESP.'
- 'PHY. - RMED'

- 'PHY. - UROL'
- 'PSY'
- 'PSYCHIATRIST'
- 'RESP'
- 'RMED'
- 'UROL'
- 'UROLOGIST'

Physician: Unspecified

- 'PHSYCIAN'
- 'PHYSICIAN'
- 'PHYSICIAN'
- 'PHYSICIAN (NEPHEW)'
- 'PHYSICIAN MTU'

Nurse

- 'RN'
- 'NURSE'
- 'DVA'
- 'NURSING'
- 'UNIT MANGER, 4D'
- 'VON'

Other Program

- 'CARE COORD., HCNS'
- 'CARE COORDINATOR'
- 'CNA'
- 'COMMUNITY SERVICES'
- 'EDUCATION COORD.'
- 'H.C. ASSESSOR'
- 'HCNS, CARE COORD.'
- 'HOME CARE'
- 'HOME CARE ASSESSOR'
- 'P.C.PROGRAM'
- 'PC PROGRAM'
- 'PC SERVICE'
- 'PC SERVICE, QE11'
- 'QEII PC PROGRAM'
- 'SOCIAL WORK'
- 'SOCIAL WORKER'