Technical Appendix:

Lung Cancer and Equity: A Focus on Income and Geography

November 2020

Contents

Background information

For the first time, data are available to examine the extent of disparities in lung cancer among people with low income and those who live in rural and remote areas. This was accomplished through a record linkage project through Social Data Linkage Environment (SDLE) at Statistics Canada. The purpose of this record linkage project was to create a linked dataset that will allow researchers to analyze the socio-demographic, socio-economic, and health characteristics of persons on the Canadian Cancer Registry (CCR) and enable them to study relationships between these factors and patient outcomes.

Data sources – part of the record linkage project

- **Canadian Cancer Registry (CCR, 1992-2015):** The CCR is a dynamic database of all primary cancer cases submitted to Statistics Canada by the provincial and territorial cancer registries on all Canadian residents diagnosed with cancer from 1992 onward. It is a person-oriented longitudinal database by cancer incidence that captures basic demographic information and cancer characteristics. The CCR is considered to be a cohort file; all other datasets that's part of the record linkage project were linked to the CCR.
- **T1 Family file (T1FF, 1992-2015):** T1FF is a file that was created to allow the development and dissemination of annual small area socio-economic data for Canadians and their families. It is derived primarily from income tax returns which are provided to Statistics Canada by the Canada Revenue Agency. The T1FF contains information on sources of income (from T1 taxfilers or those receiving the Canada Child Tax Benefits) and some demographic indicators which are derived from both the taxfilers and non-taxfilers (e.g., non-tax filing spouse, children, etc.). The T1FFs linked to the CCR included the tax filing years from 1992 to 2015.
- **Canadian Vital Statistics Death Database (CVSD, 1992–2014):** The CVSD collects demographic and cause of death information annually from all provincial and territorial vital statistics registries on all deaths in Canada. Death data are received from the province or territory of occurrence of death event. The concerted effort among the jurisdictions and Statistics Canada to disseminate complete and accurate registration of all events in Canada enables the use of the CVSD as the confirmatory death database to be used in the death record linkage of the CCR. Records eligible for record linkage were death events that occurred from January 1, 1992 through December 31, 2014.
- **Discharge Abstract Database (DAD, fiscal year 1994/95–2015/16):** The DAD includes administrative, clinical and demographic information on hospital discharges from all provinces and territories, except Quebec. Over time, the DAD has also been used to capture data on day surgery, long-term care, rehabilitation and other types of care. For this record linkage, the DAD files covering fiscal years from 1994/95 to 2015/16 were linked to the CCR.

Other data sources – outside of the linkage record project

- **Canadian Community Health Survey (CCHS):** CCHS is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. It surveys a large sample of respondents and is designed to provide reliable estimates at the health region level.

Definition of income quintile and geography

The following definitions of family income quintile and geography were used in the report.

- **Income quintile:** refers to family income quintile. It is based on income adequacy deciles which is a derived variable that's part of the T1FF linked with CCR. The calculation of income adequacy

decile variable was done on the full T1FF and were based on total family after tax income adjusted for family size. For each of the age group (<=24, 25-64, 65+), the income adequacy deciles were then assigned within Census Metropolitan Area (CMA) or Census Agglomeration (CA), and outside of CMA/CA at jurisdictional level to take geographical variation for cost of living into consideration.

- **Geography:** an individual was categorized as living in either an urban or rural region of Canada. The geographic classification was based on a Statistical Area Classification type (SAC type). Generally, CMAs and CAs were considered as urban, the rest of areas in Canada was rural. The postal codes in the CCR were used to assign SAC type by running the Postal Code Conversion File Plus (PCCF+ version 6D).

Additional notes on age-standardized incidence and mortality rates:

Statistics Canada prepared the following denominator population files (through custom requests) to enable the calculations of age-standardized incidence and mortality rates.

- **Population file by family income quintile:** contains annual population counts by family income adequacy deciles, sex and 5-year age group based on T1FF and geographic attribute files from census. To ensure consistency, the same methodology that was used to create income adequacy decile in the T1FF linked with CCR was also used to generate income adequacy deciles for the population files. Most recent census year was used for each taxation year based on residence postal codes.
- **Population file by geography:** 2016 Census files were used to aggregate the population accounts by geography type, sex and 5-year age group for age-standardization of incidence rates. The geographical type of residence areas from the census files were identified by linking to geographic attribute files of the same census year at dissemination area (DA) level. In this report, two classifications of geography were used: urban and rural based on SAC type. Please refer to the table below on how urban and rural areas were defined.

Geography	SAC type	Description
Urban	1	Census subdivision within census metropolitan area
	2	Census subdivision within census agglomeration with at least one census tract
	3	Census subdivision within census agglomeration having no census tracts
Rural	4	Census subdivision outside of census metropolitan area and census agglomeration having strong metropolitan influence
	5	Census subdivision outside of census metropolitan area and census agglomeration having moderate metropolitan influence
	6	Census subdivision outside of census metropolitan area and census agglomeration having weak metropolitan influence
	7	Census subdivision outside of census metropolitan area and census agglomeration having no metropolitan influence
	8	Census subdivision within the territories, outside of census agglomerations

General methodological notes:

- The CCR used for record linkage followed the International Agency for Research on Cancer (IARC) rules. The cancer incidence sites/types were classified by the World Health Organization, International Classification of Diseases for Oncology, Third Edition (ICD-O-3).

- Lung cancer was defined as ICD-O3 codes of C34 with behavior code of 3. Cases with histology types 9590-9992 (leukemia, lymphoma and multiple myeloma), 9050-9055 (mesothelioma) and 9140 (Kaposi sarcoma) were excluded.

- Lung cancer cases were staged as 0 through IV, occult, stage unknown, unstageable/not applicable based on the 7th edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual. When staging information based on AJCC 7th edition was not available, AJCC 6th edition was used. Please refer to inclusion criteria of each indicator for which stages were included in the analysis.

- All data included in this report were retrieved from the Research Data Centre at the University of Toronto during fall 2019; exceptions are noted in the data specification table under *Notes.*

Age-standardized incidence rates

Description	Age-standardized incidence rates per 100,000 population for lung cancer
Measurement timeframe	Diagnosis years from 2013 to 2015
Data Source(s)	Canadian Cancer Registry (CCR) linked with T1 Family File (T1FF) Population files (For more details, please see *Background information* section*)*
Stratification Variables	• Family income quintile: 2 years prior to diagnosis year • Geography: urban, rural
Exclusion	• Those without relevant family income quintile information (for results stratified by income quintile) • Cases where geography could not be assigned due to missing or incomplete information (for results stratified by geography)
Denominator	Canadian population estimates for those 30 years or older by family income quintile and geography
Numerator	Number of new invasive lung cancer cases where age at time of diagnosis was 30 years or older
Notes	1) QC's most recent data from CCR are from 2010 and therefore, QC was excluded from this analysis. 2) For more information around the population files used for this indicator, please refer to *Background information* section. 3) Incidence rates were age-standardized to the Canadian 2011 population using direct method.

Stage distribution

Definition	Stage distribution of lung cancer patients
Rationale	Stage at diagnosis is an important determinant of cancer survival and can provide helpful information about the efficacy of cancer screening programs. Understanding if/how trends in stage at diagnosis vary across patient populations is important for identifying potential inequities that exist.
Cohort	Individuals diagnosed with lung cancer from 2013 to 2015 and met the following criteria: • 30 years or older at the time of diagnosis • Family income data 2 years prior to diagnosis are available • Valid staging information is also available (I, II, III, IV) • Geography information is available and therefore assigned
Data Source(s)	Canadian Cancer Registry (CCR) linked to T1 Family File (T1FF)
Stratification Variables	• Family Income quintile: 2 years prior to diagnosis year • Geography: urban, rural
Measure	Stage distribution by family income quintile and geography.
Notes	QC's most recent data from CCR are from 2010 and do not include staging information and therefore, QC was excluded from this analysis.

Stage Distribution for lung, colorectal, prostate and breast cancer:

Stage distribution for lung, colorectal, prostate and breast cancer from 2012 to 2016 were retrieved from Statistics Canada's website (CANSIM 13-10-0761-01) in spring 2020.
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310076101

Surgical resection rates

Definition	Percentage of patients with non-small cell lung cancer who received a surgical resection within 180 days of their diagnosis date
Rationale	Surgical resections are an important and effective cancer treatment method. Exploring surgical treatment trends across different patient populations provides insight into if/how treatment trajectories vary for different patient populations and could indicate differences in decision-making or access across patients.
Cohort	Individuals diagnosed with stage I or II non-small cell lung cancer between 2012 and 2014 and met the following criteria: • 30 years or older at the time of diagnosis • Family income data 2 years prior to diagnosis are available • Geography information is available and therefore assigned
Data Source(s)	Canadian Cancer Registry (CCR) linked with T1 Family File (T1FF) and Discharge Abstract Database (DAD)
Stratification Variables	• Family income quintile: 2 years prior to diagnosis • Geography: urban, rural
Denominator	Non-small cell lung cancer patients whose stage at diagnosis was I or II
Numerator	Of the denominator, those who received a surgical resection in an acute care hospital within 180 days of their diagnosis date
Notes	1) QC's most recent data from CCR are from 2010 and do not include staging information and therefore, QC was excluded from this analysis. Also, DAD do not include records from QC. 2) The lung cancer cases with the following histology codes were excluded from this analysis (not considered to be non-small cell lung cancer): "8041" "8042" "8043" "8044" "8045") 3) The following Canadian Classification of Health Interventions (CCI) codes were used to identify hospital discharges with lung cancer resection: "1GR87DA", "1GR87PN", "1GT87DA", "1GR87NW", "1GR87QB", "1GT87NW", "1GT87QB","1GR89DA", "1GR89NW", "1GR89QB", "1GR91NW", "1GR91NWXXA", "1GR91NWXXG", "1GR91NWXXN", "1GR91QB", "1GR91QBXXA", "1GR91QBXXF", "1GR91QBXXG", "1GR91QBXXN", "1GR91QBXXQ","1GT89NW", "1GT89QB", "1GT91NW", "1GT91NWXXF", "1GT91NWXXG", "1GT91NWXXN", "1GT91NWXXQ", "1GT91QB", "1GT91QBXXF", "1GT91QBXXG", "1GT91QBXXN", "1GT91QBXXQ", "1GT89DA"

Cause-specific observed survival

Definition	Age-standardized cause-specific observed 3-year survival of lung cancer patients
Rationale	Several factors can influence the likelihood of an individual surviving cancer; understanding if/how trends in survival vary by different factors and across populations can help identify potential disparities that exist.
Cohort	Individuals diagnosed with lung cancer in 2010 and 2011 who were 30 to 99 years old at the time of diagnosis The following cases were excluded from analysis: • Death certificate only case (DCO) • Autopsy case with zero-day survival • Stages other than I through IV at diagnosis • Family income data 2 years prior to diagnosis are unavailable • Geography could not be assigned due to missing or incomplete information For individuals with multiple lung cancer records during the measurement timeframe, the following rules were applied: • Any cases beyond 31 days of the first case were removed. • If there were multiple records within 30 days of the first record, a record with the most advanced stage was kept. If stage at diagnosis were the same across multiple records, an earliest record was kept.
Data source(s)	Canadian Cancer Registry (CCR) linked with T1 Family File (T1FF) and Canadian Vital Statistics Death Database (CVSD)
Stratification variables	• Stage within family income quintile • Geography: urban, rural

Notes	1) A unique ID was assigned at national level in the linked CCR so that multiple records of a patient across jurisdictions could be identified. In this analysis, only one unique record for a patient at the national level was included.
	2) This analysis calculates cause-specific survival probabilities. For each patient, last follow-up-date (December 31st, 2014) and cause of death were used to assign their vital and censored status:
	    • Death status was assigned if a patient died of lung cancer by the last follow-up date
	    • A patient was considered to be censored if cause of death was not lung cancer or if the patient was alive by the last follow-up date
	3) The survival time was measured in days; 365.24 days was considered as one year. The survival intervals were set as follow: 1 month for the first 6 months followed by 3-month intervals until 3 years. For each survival interval, age-specific survivals were estimated based on the following groups: 30-44, 45-54, 55-64, 65-74 and 75+.
	4) The age-standardization was performed using direct method based on the age-specific survivals where International Cancer Survival Standards (ICCS) type 1

	weights were adopted to accommodate the age range used (30-44, 45-54, 55-64, and 75+). Confidence interval was computed using Taylor series approximation.

Daily and occasional smokers

Definition	The percentage of the population aged 12 years and older who reported smoking daily or occasionally
Rationale	Reporting on tobacco use at the population level allows for the assessment of tobacco prevention and cessation strategies.
Measurement timeframe	2015 to 2016 calendar years combined
Data source(s)	Canadian Community Health Survey
Stratification variables	• Household income quintile at national level • Geography: urban, rural
Denominator	Individuals aged 12 years and older Individuals whose responses were "don't know" or "refusal to answer" to the question that asked about their smoking status were excluded from the denominator.
Numerator	Of the denominator, those who reported that they smoked daily or occasionally
Notes	1) The following question was used to determine smoking status: At the present time, do you smoke cigarettes every day, occasionally or not at all? 2) CCHS data are based on a representative sample which is then extrapolated to the overall population. Data retrieval date: Summer 2018