Enhancing measurement of disparities in cancer through linkage of the Canadian Cancer Registry to hospitalization, income tax, immigration and census long form data: A first in Canada

A Collaboration Between the Canadian Partnership Against Cancer and Statistics Canada

Presenters: Yves Decady statistics Canada
Sharon Fung Canadian Partnership Against Cancer

October 18, 2018
Outline

• Background
• Methods
  • The Social Data Linkage Environment (SDLE)
  • Linkage process
  • Limitations
• Results & Linkage Possibilities
  • Benefits of data linkages
  • Research applications
  • Dissemination plans for the future
• Conclusion
Background
Equity and Canada’s Health Care System

- Two pillars of Canada’s health care system are universality and accessibility.

- Our system has a duty to ensure equitable access to quality care for all residents of Canada, regardless of income, place of residence, immigrant status, education, age, gender, sexual orientation, race or ethnicity.
Equity in Canada’s Cancer Control Systems

Quality refers to the effectiveness of care—high-quality care is evidence-based and improves health outcomes.

Equity refers to the absence of sociodemographic barriers, such as socio-economic status, place of residence and immigrant status, in accessing effective cancer control.

Seamlessness refers to an integrated, person-centred cancer control system that allows patients and their families to easily understand, access and navigate the system.

Sustainability refers to meeting the health care needs of the population in a way that optimizes the balance between resource use and patient outcomes.
Previous Challenges with Measuring Disparities

• Using ecological data, we have previously measured differences in access to and utilization of cancer care services by:
  • Income
  • Immigrant status
  • Geography

• However, ecological data lacks the individual level information required to identify the extent of these disparities
New Opportunities Through Data Linkage

• Statistics Canada’s innovative work with data linkage using record-level data from the Canadian Cancer Registry will enhance previous work on disparities

• Allows us to explore relationships between socioeconomic and sociodemographic factors and cancer diagnosis, treatment patterns and patient outcomes at the patient-level
Methods
The Social Data Linkage Environment

• Promotes the innovative use of existing administrative and survey data to address important research questions and inform socioeconomic policy

• The personal identifiers in all data sources included in the Social Data Linkage Environment can be matched to find the same person across multiple data sources
The Canadian Cancer Registry (CCR) is a dynamic, person-oriented database that contains a rich source of information on cancer characteristics.

- Has limited socioeconomic and cancer treatment information.
- CCR can be enriched using record linkage to datasets that contain sociodemographic, socioeconomic and treatment information.
### Overview of Linked Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discharge Abstract Database (DAD)^D</strong></td>
<td>Clinical and demographic information on hospitalizations</td>
</tr>
<tr>
<td>1994/95-2015/16</td>
<td></td>
</tr>
<tr>
<td>**National Ambulatory Care Reporting System</td>
<td>All hospital-based and community-based ambulatory care visits</td>
</tr>
<tr>
<td><em>(NACRS)^D</em></td>
<td></td>
</tr>
<tr>
<td>2002/03-2015/16</td>
<td></td>
</tr>
<tr>
<td><strong>T1 Family File (T1FF)^D,P</strong></td>
<td>Individual and family income</td>
</tr>
<tr>
<td>1992-2015</td>
<td></td>
</tr>
<tr>
<td><strong>Canadian Vital Statistics – Death Database</strong></td>
<td>Death information</td>
</tr>
<tr>
<td><em>(CVSD)^P</em></td>
<td></td>
</tr>
<tr>
<td>1992-2014</td>
<td></td>
</tr>
<tr>
<td><strong>Longitudinal Immigration Database (IMDB)^P</strong></td>
<td>Immigrant status, class and category</td>
</tr>
<tr>
<td>1992-2014</td>
<td></td>
</tr>
<tr>
<td><strong>Census of Population Long-Form^P</strong></td>
<td>Education, ethnicity, language spoken, etc.</td>
</tr>
<tr>
<td>2016</td>
<td></td>
</tr>
</tbody>
</table>

Linked

D = Deterministic record linkage
P = Probabilistic record linkage
Overview of the Linkage Process

Within the SDLE,

1. A unique record identifier is assigned to each record of a dataset followed by partitioning the dataset into two:
   (i) a dataset with only the record and personal identifiers
   (ii) an analytical dataset with no personal identifiers

2. Datasets with the record and personal identifiers are separately linked to the SDLE’s longitudinal database of basic personal identifiers that are linked to the SDLE identifiers (this is also when the SDLE is updated).

3. The SDLE identifiers are paired to the record identifiers of the datasets from Step 2.
Within the SDLE (cont’d),

4. Records from separate datasets where a match is found between the SDLE identifiers are linked. Once linked, a unique identifier is assigned to each record to signify that they are linked (i.e., same identifier on each record from each separate dataset).

5. The SDLE and personal identifiers are removed and the record identifiers are used to merge back to the analytical datasets to create linked datasets.

Outside the SDLE,

6. Validation of the linked datasets:
   a. Bias assessment of linked and not-linked records
   b. Linkage rates
   c. Fitness for use
   d. Documentation of data limitations

7. Data access through the Research Data Centres network.
Data Limitations

- Cancer incidence data from Quebec are not available for diagnosis years 2011 and onwards
- There is no hospitalization data (DAD records) from Quebec
  - Therefore the treatment of residents from other provinces or territories received in Quebec will not be captured
  - This is important for bordering provinces, especially patients travelling to Quebec from the northern regions for treatment
- The NACRS only covers participating jurisdictions
- Many of the datasets in this linkage include death information (e.g., date of death) and the date of death values may not agree across datasets
- Limited chemotherapy and radiation therapy treatment data availability in the DAD and the NACRS
- The hospitalization information linked to a cancer case may not be related to the cancer diagnosis
Results & Linkage Possibilities
## CCR Cohort Linkage Status

<table>
<thead>
<tr>
<th>Linked dataset</th>
<th>Maximum number of CCR patients linked&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Linkage Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower bound</td>
</tr>
<tr>
<td>DAD</td>
<td>2,289,995</td>
<td>60.08%</td>
</tr>
<tr>
<td>NACRS</td>
<td>1,459,455</td>
<td>44.01%</td>
</tr>
<tr>
<td>T1FF</td>
<td>3,049,595</td>
<td>75.87%</td>
</tr>
<tr>
<td>CVSD/T1PMF&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1,549,220</td>
<td>99.24%&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>1</sup>Rounded to nearest 5

<sup>2</sup>The death record linkage was based on the CVSD and the T1PMF (1992-2014)

<sup>3</sup>Based on the total number of deceased patients according to the CCR (1992-2014)
How similar is the age at cancer diagnosis distribution for males in the CCR cohort compared to those of the linked data: CCR-DAD, CCR-NACRS and CCR-T1FF?
How similar is the age at cancer diagnosis distribution for females in the CCR cohort compared to those of the linked data: CCR-DAD, CCR-NACRS and CCR-T1FF?
Datasets that have been linked to the Canadian Cancer Registry:

- Discharge Abstract Database
- National Ambulatory Care Reporting System
- T1 Family File
- Canadian Vital Statistics (Death Database)
- Longitudinal Immigration Database
Benefits of Data Linkages

• For the **first time** in Canada, record-level linkage of the Canadian Cancer Registry with key datasets containing sociodemographic and socioeconomic information will be available for **exploratory analysis**

• **Leveraging existing data** to address research questions rather than creating new datasets and collecting additional information from Canadians
Research Applications of Linked Data

Ability to study persons diagnosed with cancer and their:

- Incidence, mortality and survival
- Multi-morbidity
- Frequency of hospitalization and emergency department visits
- Surgical treatment patterns and wait times
- End of life care
- ... and more

And examine how these findings are impacted by:

- Socioeconomic status
- Sociodemographic factors
- Type of cancer
Questions the Partnership is Interested in Exploring...

• Does income, immigrant status, education level, and/or place of residence of a patient impact their likelihood of:
  • Having more **unplanned hospitalizations**?
  • Receiving **evidence-based treatment**?
  • Dying in hospital?

• How does cancer impact a patient and their family’s **income and employment status** after diagnosis and treatment?
Next Steps

• Preliminary investigation is on-going
• Results will be disseminated throughout 2019 and 2020
• Linked data will be made available in all RDC facilities (starting Spring 2019)
  • Researchers will be able to apply to access data to answer cancer research questions
Conclusion

The generated evidence from this initiative provides an opportunity for:

- **The Partnership and cancer agencies** – creating an impetus for change to better address inequities in high-risk groups
- **Researchers** – broadening the scope of cancer research to look at the impact of social determinants of health
- **Patients** – experiencing a more equitable cancer control system that fulfills the promise of universality and accessibility