Research Data Request Form- Example Dataset Creation Plan

Please note that this is just an example meant to provide clarity on how to complete the Research Data Request Form. The information contained in this example does not correspond to a real study and is meant only to provide guidance on completing the form. Completion of the form is required in order for the Data Disclosure Team to process a request. Providing accurate and sufficiently detailed information can facilitate the request process more effectively.

While communication with the Data Disclosure Team is encouraged during the entire data request process, providing concise, detailed information upfront for all requirements in the Research Data Request Form will minimize any delays that may result from a need for more information on behalf of CCO prior to processing the request.

For more information on requesting data at CCO, please contact the Data Disclosure Team at Datarequest@cancercare.on.ca

A. CONTACT INFORMATION

Name of Principle Investigator	Dr. Nikola Kage
Role/Title	Senior Scientist
Name of Organization	Breast Cancer Education Institution of Northeast Canada
Address	1234 Toronto Avenue, Toronto ON, M1M 3F3
Phone	416-000-0000
Email	Doctor_NikolaKage@BCEINC.ca
Name of Primary Contact	Markie Marc
Role/Title	Project Coordinator
Name of Organization	Breast Cancer Education Institution of Northeast Canada
Address	1234 Toronto Avenue, Toronto ON, M1M 3F3
Phone	416-000-0000
Email	MarkieMarc@BCEINC.ca

NOTE: Please complete the Research Privacy Requirements Form with additional names of all Co-Investigator(s) and person(s) who will have access to requested data.

B. PROJECT DESCRIPTION

Project Title	Breast cancer survival in the Greater Toronto Area: Does income matter?
Research Purpose and Clinical Relevance	**Briefly describe the purpose for the research project, stating the research question or hypothesis to be examined and the clinical relevance of research findings.** The aim of this study is to estimate the relative survival differences by income for patients diagnosed with breast cancer in the GTA. Breast surgery utilization will also be explored to investigate potential differences in treatment strategies. Important sociodemographic characteristics will also be examined to further understand the impact that the social determinants of health can have on patient outcomes and treatment strategies for breast cancer. This study will shed light on breast cancer survival in the GTA. The findings can help inform clinicians and public health workers on which populations may require targeted education and intervention.

Research Proposal
Append a full REB-approved proposal describing the research project. The proposal should include the objectives, methodology, and the anticipated public and/or scientific benefit.

C. RESEARCH APPROVALS

Funding Organization	CIHR

Funding and Granting Information

Period of Grant: From **4/1/2016** To **3/31/2018**

Amount of Grant: $125,000

Amount available for data request: $50,900

Append the approved research budget to the Intake Form.

Ethics Approval Status

Identify all Research Ethics Boards (REBs) who reviewed the research proposal, the status of the application(s), and any decision from each. Copy the statement below as necessary to capture all REBs.

REB: University of Toronto, Health Sciences REB

Current Status: Approved

NOTE: The REB(s) providing approval must demonstrate compliance with PHIPA under O.Regs 329/04 s.15

D. DATASET CREATION PLAN

Is this a request for permission to use the data to which you already have access? Yes ☐ No ☒

Datasets Required

List all required data elements under each corresponding dataset in the table below. Reference the Frequently Asked Questions for details about available data elements.

Dataset	Data Element(s)	Year(s)	Rationale or Purpose of Use
Ontario Cancer Registry	1. Last contact date 2. Vital status	2006-present	Required for survival analysis.

CIHI Discharge Abstract Database	1. Main diagnosis 2. Secondary diagnosis 3. Main intervention 4. Secondary intervention 5. Admission date	2006-present	Variables required for Cox Proportional Hazards model and logistic regression model to control for potential confounders.
CIHI National Ambulatory Care Reporting System	1. Main diagnosis 2. Secondary diagnosis 3. Main intervention 4. Secondary intervention 5. Registration date	2006-present	Variables required for Cox Proportional Hazards model and logistic regression model to control for potential confounders.
Choose an item.	Click here to enter text.	Click here to enter text.	Click here to enter text.
Choose an item.	Click here to enter text.	Click here to enter text.	Click here to enter text.

Data Linkages

Complete table below if the researcher will be linking CCO data to other datasets.

Planned data linkages (e.g. list the databases that will be linked to CCO data)	**What variables will be used for the linkage?**
CCO data will be linked to hospital database containing sociodemographic data on cohort and other clinically relevant information from chart abstractions.	Health card number, first name, last name, date of birth, postal code.
Click here to enter text.	Click here to enter text.

Cohort Detail

Complete this section if the cohort will be provided to CCO.

What variables will be provided to CCO to perform the linkage?
Health card number, first name, last name, date of birth, postal code

Study Design

Please specify the study design for this request (e.g., cohort study, case-control, data-cut).

Longitudinal survival analysis, cohort study.

Inclusion Criteria

Specify all inclusion criteria for data extract. If subjects need to be identified by CCO, outline how they should be identified (e.g. index event, timeframe for entering study).

1. Cohort provided to CCO will contain desired study population:
    a. Female
    b. Age 30-65 at diagnosis
    c. Diagnosed with first and only breast cancer case between 2006-2010 2. CCO will obtain data for only patients residing in:
    a. Central LHIN
    b. Central East LHIN
    c. Central West LHIN
    d. Mississauga Halton LHIN
    e. Toronto Central LHIN

Exclusion Criteria

Specify all exclusion criteria for data extract.

1. **Exclude CIHI DAD and NACRS hospitalization records where the main diagnosis (ICD) are:** a. C70.0 – C70.9
    b. C71.0 – C71.9
    c. C72.0 – C72.9
2. Exclude patients who died within 30 days of breast cancer diagnosis.
3. Exclude patients where the LHIN of residence cannot be determined.
4. Exclude patients where the neighbourhood income quintile could not be determined.
5. Exclude patients where the last contact date is within 1 year of diagnosis.

Study Size

Enumerate all groups involved in study (e.g. exposed, unexposed, cases, controls). If the number of study cases differs from the number of cases in the cohort applicable to this data request, please indicate so.

1,978 patients provided. Number of exposed and unexposed unknown, to be determined by CCO.

Follow Up Period

If follow-up of subjects is required, state the follow up period.

Records from 2006 to most recent available.

Main Exposure or Intervention

If there is a main exposure, state how it should be defined (e.g. ICD-10 code, Neighbourhood income quintile).

Use QAIPPE from PCCF+ to identify exposed and unexposed cohorts. Exposed cohort to be defined as individuals with QAIPPE from medium-high quintile to highest quintile (HIGH). Unexposed cohort to be defined as individuals with QAIPPE from lowest quintile to middle quintile (LOW).

Exposure variable: INCOME (HIGH, LOW)

Primary Outcome

If there is a primary outcome of interest, state how it should be defined (e.g. ICD-10 code). If there are additional outcomes of interests, state how they should be defined.

Primary outcome is vitality status coded as alive (A) or dead (D). Vitality variable: VITALITY_STATUS (A, D)

Secondary outcome is an occurrence of breast surgery from CIHI DAD or NACRS hospitalization records under main or secondary intervention to be coded as yes (Y) or no (N). The following CCI codes should be used:

1YK87LA, 1YK89LA, 1YK90LAXXE, 1YL87LA, 1YL89LA, 1YM87DA, 1YM87GB, 1YM87LA, 1YM87LAXXA, 1YM87UT, 1YM88LAPM, 1YM88LAPME, 1YM88LAPMG, 1YM88LATP, 1YM88LAXXE, 1YM88LAXXF, 1YM88LAXXG, 1YM89LA, 1YM89LAXXA, 1YM90LAPM, 1YM90LAPME, 1YM90LAQF, 1YM90LAQFE, 1YM90LAQFF, 1YM90LATP, 1YM90LATPF, 1YM90LATPG, 1YM90LAXXE, 1YM90LAXXF, 1YM90LAXXG, 1YM91LA, 1YM91TR, 1YM91WP, 1YM91LATP, 1YM91LAXXA, 1YM91LAXXE, 1YM91TRXXA, 1YM91TRXXE, 1YM92LAPME, 1YM92LAPMF, 1YM92LAPMG, 1YM92LAQFE, 1YM92LAQFG, 1YM92LATPE, 1YM92LATPF, 1YM92LATPG, 1YM92LAXXE, 1YM92LAXXF, 1YM92LAXXG, 1YM92TRXXF, 1MD87LA, 1MD89LA, 1MD89LAXXA, 1MD89LAXXE, 1MD89LAXXF

Breast surgery variable: B_SURGERY (Y, N)

Output Variables

If output should be formatted in a particular way, define how this should be done (e.g. age groups: 21-30, 31-40, 41-50, 51-54, 54+).

Calculate and categorize age at hospitalization using patient birthdate and CIHI DAD admission date.
Age variable: AGE (30-39, 40-49, 50-59, 60-65)

Analytical Plan

Describe the proposed analysis using CCO data.

Kaplan-Meier and Cox Proportional Hazards models will be used to examine longitudinal patient survival assessing the association of neighbourhood income and vitality status. Diagnostic and intervention codes from CIHI DAD and NACRS along with sociodemographic and clinical data from hospital database will be used to control for potential confounders.

Logistic regression will be used to estimate the odds ratio for neighbourhood income quintile and breast surgery occurrence. Diagnostic and intervention codes from CIHI DAD and NACRS along with sociodemographic and clinical data from hospital database will be used to control for potential confounders.

The impact of sociodemographic factors will also be assessed for their association with mortality and breast surgery occurrence.

Other Considerations

If there are other important considerations that need to be captured, indicate them here.

If there is a large range in the types of diagnostic or intervention codes in CIHI DAD or CIHI NACRS, codes may have to be categorized. This can be discussed as the data extraction process occurs.

Preferred Format

Specify the preferred format of the completed data (e.g., SAS file, Excel). Where possible, attach a template.

SAS file where the record granularity is at the hospitalization-level (i.e. do not transpose data to patient level).

Research Privacy Requirements

E. CO-INVESTIGATOR(S) AND PERSONS WHO MAY HAVE ACCESS TO REQUESTED DATA

Co-Investigators
Complete additional copies as required.

To add additional CO-I information, select the table and click the + sign on the right side of the table.

Name	Dr. Larrie Devid
Role/Title	Oncologist
Name of Organization	Toronto Central Hospital
Email	LarrieDevid@TCHospital.ca
Qualifications/Subject Matter Expertise	Breast cancer expertise and biostatistician.
Why is access required for this person?	**If access to CCO data is not required, enter N/A** Provide clinical insight to help guide analysis and interpretation of results.

Persons Who May Have Access to the Data
Complete additional copies as required.
To add additional CO-I information, select the table and click the + sign on the right side of the table.

Name	**Tamara Latford**
Role/Title	Research Assistant
Name of Organization	Breast Cancer Education Institution of Northeast Canada
Email	Tamara_Latford@BCEINC.ca
Qualifications/Subject Matter Expertise	PhD Candidate, Epidemiology
Why is access required for this person?	Provide operational Support

F.TIMELINE, DATA RETENTION, DISPOSAL OR RETURN OF DATA

Project Timeline

Date by which data is required: 7/1/2016

Project Start Date: 7/1/2016

Planned Project Completion Date: 3/31/2017

Data Retention and Destruction

Describe the safeguards that will be in place to protect the confidentiality and security of the data provided by CCO:

All records will be stored on an encrypted server that will be password protected. No paper records of data will be made. Premise is secured by electronic access card. Hospital has strong safeguards in place to protect patient privacy and confidentiality.

How long will the information provided by CCO be retained in an identifiable format? (I.e., how long will direct identifiers be stored alongside the information?) 3/31/2017

Date when access to record level data provided by CCO will no longer be required (I.e., when do you plan to return or destroy the data):
3/31/2017

Provide justification for the length of time that the data provided by CCO will be retained:

Time is required to complete all statistical analyses.

The information provided by CCO will be: ☐ **Returned:** Click here to enter a date.
Or ☒ **Destroyed: 3/31/2017**

PHIPA and CCO require that data be transferred, returned or destroyed in a secure manner. Should data be returned to CCO, please contact CCO prior to return in order to confirm the appropriate method(s) for the return of data. Please see CCO's Information Security Policy, as well as the Information & Privacy Commissioner, [4]Ontario Fact Sheet # 10: Secure Destruction of Personal Information and Best Practices for the Secure Destruction of Personal Health Information for guidance on the secure destruction of data

NOTE:

- *Records of personal health information disclosed by CCO for research purposes must not be retained for a period longer than set out in the approved research plan. Researchers must return or destroy all data provided by CCO within 60 days of the date listed above.*
- *Assertions of the destruction of information will require that researchers supply CCO with a Certificate of Destruction setting out the date, time and location of the secure destruction and the method of secure destruction employed as well as details of the items destroyed. The Certificate of Destruction will bear the signature of the persons who securely destroyed the information. A Certificate of Destruction from a third party service provider will be acceptable if it contains all of this information.*

G.BENEFITS, HARM AND CONFLICT OF INTEREST

Alternatives

Is it possible to perform the research project without using Personal Health Information (PHI) or record level data? Yes ☐ No ☒

If No, explain why the research cannot be accomplished without the use of PHI or record level data (i.e. what alternative methods were considered)
PHI is required to obtain survival data on patients in cohort and to obtain CIHI DAD and NACRS diagnostic and intervention data for statistical analyses.

If you will not be obtaining patient consent, please provide an explanation as to why consent for the disclosure of PHI is not being sought from the individuals to whom the PHI relates:
Project poses minimal to no risk to patients due to the nature of the work. All results from project will be present in aggregate form with cell counts 5 or less being suppressed. No contact with patients will be made.

If PHI is to be linked to other data (as indicated in Section D) include an explanation of why such linkage is necessary:
Linkage of CCO data to hospital database is required as important sociodemographic and clinical information is present in the hospital database. After completion of the project, no information from CCO will be retained.

Plan for Dissemination of Results

Describe the levels at which results will be reported, noting the smallest reporting unit (i.e. confirm that cell sizes 5 or less will not be reported).

Aggregate descriptive statistics and relative measures (hazard ratio, odds ratio) will be presented. Cell size 5 or less will not be reported.

Include description of how this data will be disseminated once the research project is complete (e.g., publications, presentations, etc.):
Project findings will be presented at scientific conferences and published in scientific journals.

Potential Harm

Describe any way the use of personal health information in this research project might harm patients (e.g. privacy breach leading to potential identification of patient, stigmatisation): It is possible that patients from this study could be identified if there is a security breach.

Describe how the PI intends to address the potential for harm outlined above:
Once CCO data has been linked to hospital data, the analysis will only take place on data that has been stripped of the following identifiers: Health card number, first name, last name, date of birth, postal code. This should minimize the risk of a breach since members of the project will be spending most of time working on data without the previously mentioned identifiers.

Conflict of Interest

Will any of the individuals involved in the research project's interest in the disclosure of the requested PHI or in the performance of the research likely result in an actual or perceived conflict of interest with other duties of the researchers?

Yes ☐ No ☒

Acknowledgements by Principal Investigator

The Researcher is requesting record level data from CCO. The Researcher understands and acknowledges that records requested may contain confidential personal health information (PHI) about individuals, including potentially identifiable information such as diagnoses dates, and names of physicians or hospitals, or may otherwise be in a form where individuals may be identifiable. If access to these records is approved, the Researcher must abide by the provisions of CCO's Research Data Disclosure Agreement (RDDA).

The Principal Investigator (PI) acknowledges and understands that the records requested may contain identifiable, record-level personal health information (PHI). If this information is released to the PI, the PI must abide by the provisions of CCO's Research Data Disclosure Agreement (RDDA). If and when this request is approved by CCO, the PI and all those who will have access to the data will sign the required Non-disclosure/Confidentiality Agreement for Researchers before the data is provided by CCO. The PI will also provide a purchase order for the amount to be specified by CCO and pay the invoice promptly. In situations where the PI or others who will have access to data are students, the students' academic supervisor or advisor is also required to sign the Non-disclosure/Confidentiality Agreement.

1. The PI agrees to ensure that cell sizes less than or equal to 5 will not be reported without prior written approval from CCO.
2. The PI agrees to only conduct data linkages in accordance with the approved Research Proposal.
3. The PI agrees that the retention period for data received from CCO indicated in section F is consistent with the retention period set out in the approved Research Proposal.
4. The PI agrees to ensure security and protection of identifiable record level data in accordance with best practices, including the Information & Privacy Commissioner, Ontario *Fact Sheet # 16: Health-Care Requirement for Strong Encryption, Fact Sheet # 12: Encrypting Personal Health Information on Mobile Devices* and *Fact Sheet #14: Wireless Communication Technologies: Safeguarding Privacy & Security* (see FAQs for more information on IPC Fact Sheets).
5. The PI agrees to ensure that data returned or destroyed be done in a secure manner in accordance with the Information & Privacy Commissioner, Ontario *Fact Sheet # 10: Secure Destruction of Personal Information* and *Best Practices for the Secure Destruction of Personal Health Information.*

The Principal Investigator certifies that the information reported in this form and the appended Research Project Proposal is accurate and agrees to comply with the terms and conditions contained in this form.

Name of Principal Investigator	Dr. Nikola Kage

Title	Senior Scientist
Signature	
Date	Click here to enter a date.