A Canadian real-world evidence (RWE) study to evaluate provider perspectives on and impact of sentinel lymph node biopsy omission in early-stage breast cancer

A Data Management Plan created using DMP Assistant

Creator: Ipshita Prakash

Principal Investigator: Ipshita Prakash

Data Manager: Aynslie McIntyre, Pamela Brazeau-Porrello

Contributor: Aynslie McIntyre, Pamela Brazeau-Porrello

Affiliation: McGill University

Template: McGill Medicine and Quantitative Health Sciences

Project abstract:

Background: Recent randomized trials (SOUND, INSEMA) demonstrated non-inferiority of omitting sentinel lymph node biopsy (SLNB) in selected patients with early-stage breast cancer, prompting guideline changes. While SLNB omission reduces morbidity, it may lead to under-staging and unintended exclusion from adjuvant therapies such as axillary radiation and CDK4/6 inhibitors. The real-world impact of these changes in Canada is not well understood.

Objectives: We aim to (1) quantify the implications of SLNB omission on eligibility for adjuvant CDK4/6 inhibitor therapy and axillary radiation use, and (2) evaluate national trends and predictors of SLNB omission in routine practice.

Methods: This is a multi-centre, retrospective cohort study including patients with stage I–II, cT1–2 cN0 invasive breast cancer who underwent upfront surgery between 2015–2025 across Canadian academic and community hospitals. Patients will be stratified by SLNB status, with detailed clinicopathologic and treatment data abstracted from institutional electronic medical records into a standardized REDCap database. Analyses will include descriptive statistics, logistic regression to identify predictors of omission, and concordance with trial-defined eligibility.

Expected Results: We anticipate identifying the proportion of patients who would be excluded from CDK4/6 inhibitor therapy or axillary radiation if SLNB is omitted, the frequency of occult nodal disease, and the real-world rate of upstaging to ≥pN2. Trends in SLNB omission before and after 2023 will be described, with exploration of patient- and system-level predictors.

Conclusions: This study will provide the first pan-Canadian, real-world evaluation of SLNB omission following recent trial evidence. Findings will inform the balance between surgical de-escalation and appropriate access to adjuvant therapies, guiding safe and equitable implementation of evolving axillary management guidelines.

Identifier: 17225

Start date: 01-11-2025

End date: 31-10-2026

Last modified: 22-09-2025

A Canadian real-world evidence (RWE) study to evaluate provider perspectives on and impact of sentinel lymph node biopsy omission in early-stage breast cancer

Data Collection

What types of data will you be collecting? What file formats will you be generating during the data collection phase? What data collection instrument or scales will you use to collect the data? What software programs will you use to collect the data? (Please use the table provided in the text response as a guide for the information that should be included in a response to this prompt. Include a description to summarize the information in the table.).

Data types: This study will collect retrospective clinical data from electronic medical records (EMR) and institutional databases.
Data will include: -

- Demographic and baseline data (e.g., age, race/ethnicity, menopausal status, comorbidities/ECOG).
- Diagnostic and tumor characteristics (dates, tumor size, grade, histology, ER/PR/HER2 status, Ki-67, nodal status, AUS findings).
- Surgical and pathologic data (surgery type, axillary procedure, pathologic size/grade/nodal status).
- Genetic information (testing status, mutation status).
- Adjuvant treatment details (radiation, systemic therapy including endocrine therapy, chemotherapy, CDK4/6 inhibitors, PARP inhibitors).
- Follow-up data (date of last follow-up, vital status)

All data are structured, tabular, numeric or categorical variables, extracted manually into a database. No imaging (e.g., DICOM), -omics, audio, or video data will be collected. No raw genomic sequencing data will be generated; only categorical results of clinically indicated genetic testing.
Collection times: Data are collected retrospectively at a single point in time (chart abstraction from existing records). Longitudinal variables (e.g., treatment course, survival, follow-up) will be abstracted retrospectively but not collected prospectively.
File formats: The primary research dataset will be stored in CSV (.csv) format to ensure open, non-proprietary long-term accessibility. Working datasets may also be maintained in Excel (.xlsx) and REDCap exports (.csv, .xlsx) during collection and cleaning. Statistical analysis files may include R (.R, .Rmd, .Rproj) or SPSS (.sav) files depending on analytic platform. All final analytic datasets for preservation will be in CSV format.
Instrument or scale: The ECOG Performance Status will be collected as documented in charts (Oken et al., 1982). Clinical and pathologic staging will follow AJCC 8th edition criteria. Histologic subtype and biomarker data (ER, PR, HER2, Ki-67) will be abstracted directly from pathology reports, according to institutional standards. No novel surveys, psychological scales, or laboratory instruments are being administered as part of this study.
Software programme: REDCap (v13.x or institutional current version) will be used for secure data capture. Microsoft Excel (Office 365) may be used for initial data handling. R Statistical Software (version 4.3.x) will be the primary platform for statistical analyses. SPSS (version 29, if applicable) may be used by collaborators.

Does this project involve the use or analysis of secondary data? What is the data-sharing arrangement for these data?

This study will not reuse any publicly available or external secondary datasets. All data will be abstracted directly from patient medical records and institutional databases at under the authority of the Research Ethics Board (REB) approval.

What conventions and procedures will you use to structure, name, and version-control your files to help you and others better understand how your data are organized?

To ensure clarity, consistency, and reproducibility, all files will follow a standardized naming convention. File names will use only alphanumeric characters, underscores, and dashes, and will not exceed 32 characters. The convention incorporates the date, project acronym, file type/topic, and version.
Convention: [YYYYMMDD]_[ProjectAcronym]_[CreatorInitials]_[FileTopic]_[FileType]_v[###]
[YYYYMMDD] – Date of file creation.
[ProjectAcronym] – [SLNBOmission]
[CreatorInitials] – Initials of the individual who created or modified the file
[FileTopic] – Brief description of the file contents (e.g., “dataset”, “code”, “analysis”, “table1”).
[FileType] – Indicates whether the file is raw, cleaned, or processed (e.g., raw, clean, final).
[v###] – Version number recorded in the format v001, v002, etc.
Each time a dataset or code file is updated, the version number will be incremented (v001, v002, etc.). Older versions will be archived in a “/previous_versions” subfolder to ensure traceability.

Data Analysis

What tools, devices, platforms, and/or software packages will be used to analyze and manipulate data during the project? Does your data need to be copied to a new media or cloud platform, or converted to a different file format when you analyze your datasets?

The REDCap platform will be used to collect and store data securely, and Microsoft Excel will be used for basic data cleanup and organization. The R statistical software will primarily be used for data analysis and, in some cases, SPSS depending on collaborator preference. Data will be kept in common formats like CSV or Excel, and results such as tables or figures will be saved in Word or PDF. All data will stay on secure institutional servers.

What quality assurance measures will be implemented to ensure the accuracy and integrity of the data?

To ensure accuracy and integrity, data will be abstracted using a standardized data collection form with clearly defined variables. A subset of records (approximately 10%) will be double-checked by a second reviewer to confirm consistency and reduce errors. Any discrepancies will be discussed and corrected, with updates documented in a changelog. During data cleaning, checks will be performed for missing values, implausible entries (e.g., negative time intervals), and consistency across variables (e.g., diagnosis date preceding surgery date). Data queries will be resolved by returning to the original medical record. For transparency in analysis, all statistical code will be saved and version-controlled, and outputs will be cross-checked against source data to confirm validity.

Documentation and Metadata

What information about your research would someone outside of your research team need to know to reuse or interpret your data?

Because this study involves patient-level clinical data, the raw dataset will not be shared outside the research team due to privacy restrictions. However, to support transparency and allow others to interpret the findings, we will make accompanying materials available where appropriate. These may include the study protocol, a detailed codebook outlining variable definitions and coding schemes, and a description of the data abstraction process. Aggregate results (e.g., summary tables, figures) will be shared in publications and presentations. If needed, information about the software and statistical methods used, as well as annotated analysis scripts, could be provided to allow others to understand and reproduce the analytic approach, even though the underlying patient-level data remain confidential.

Are there metadata standards that you could use to describe your data?

This project does not generate data that requires metadata standards. Basic metadata like the data dictionary and codebook will be provided.

Storage and Backup

What are the storage requirements needed for your data?

Data will be housed on a secure computer housed in E-718 of the Segal Cancer Centre within a password-protected file until all data collection and analysis is complete. The total expected storage need is approximately 150–450 MB (≤ 0.5 GB). We will provision a conservative ceiling of 1 GB on secure institutional servers. No specialized high-capacity storage will be required.

Where will your data be stored during the *data collection* phase? Where will your data be stored during the *data analysis* phase, if different?

During the data collection phase, all information will be entered directly into REDCap. Data will remain encrypted and backed up regularly by institutional IT services.
During the data analysis phase, data will be housed on a secure computer housed in E-718 of the Segal Cancer Centre within a password-protected file until all data collection and analysis is complete. After obtaining all relevant data variables, the datafile will be anonymized. An ID number will be assigned to each patient. A document linking their ID number and MRN number will be kept separately in a password protected file behind the CCOMTL firewall, stored securely by the principal investigator.

Data Sharing and Preservation

***After* the project is complete, will your data be migrated to preservation formats?**

Yes. After project completion, the final de-identified analytic dataset will be migrated and preserved in an open, non-proprietary format (CSV) to ensure long-term accessibility and interoperability. Supporting documentation (e.g., codebook, data dictionary, and analysis scripts) will be saved in TXT, CSV, or PDF formats. Working files in proprietary formats (e.g., Excel, SPSS) will be retained only as needed for reference, with the CSV version serving as the preservation copy. All files will remain stored on secure institutional servers in accordance with institutional retention policies.

***After* the project is complete, where will the data be stored? If applicable, where will you share your data after the project is complete?**

After the project is complete, all study data will remain stored on secure institutional servers at the Jewish General Hospital, in accordance with CIUSSS CCO-Mtl policies and Research Ethics Board (REB) requirements. The data will be retained for 7 years following project completion, after which it will be securely destroyed (electronic files permanently deleted, paper forms shredded).
Because the dataset contains patient-level clinical information, it cannot be publicly shared or deposited in open repositories. Instead, aggregate results (e.g., tables, figures, and summary statistics) will be disseminated through peer-reviewed publications, conference presentations, and academic reports. Supporting documentation (e.g., study protocol, analysis scripts) may be made available upon reasonable request, provided this does not compromise patient privacy.

If you have collected restricted data, what steps would someone requesting your data need to follow to access it?

N/A

What procedures are in place to destroy the data after the retention period is complete?

See above.

Responsibilities and Resources

Who is the main contact and steward for the data collected in this study?

The main contact for the data collected is the Principal Investigator, Dr. Ipshita Prakash MD, MSc, FRCSC. Dr. Prakash is a breast surgical oncologist at the Jewish General Hospital's Segal Cancer Center and is an Assistant Professor of Surgery & Oncology at McGill University.
E-mail: ipshita.prakash@mcgill.ca

Who will have access to your data throughout the project?

Pamela Brazeau-Porrello MD, MSc (c)
Pamela is a master's student, who will be working on this project as part of her MSc in Surgical & Interventional Sciences at McGill University.

pamela.brazeau-porrello@mail.mcgill.ca
Aynslie McIntyre MD (c), MSc
Aynslie is a medical student, who will be working with Pamela on the data abstraction and cleanup for this project.
aynslie.mcintyre@mail.mcgill.ca

Will any new members be added, or responsibilities be transferred, over the course of the study?

If new collaborators or team members are added to the project to help with data abstraction, cleaning, statistical analysis or manuscript preparation, they will first be submitted for approval to the Research Ethics Board (REB) to ensure compliance with institutional and privacy requirements. Only after REB approval will they be granted access to study data. All new team members will be required to complete institutional privacy and confidentiality training, and their level of data access will be restricted to the minimum necessary for their role.