Assessing MOHLTC Research Protocols – Secondary Data Studies

By: Tommy Tam, MSc

Definition of secondary data study:
Secondary data studies use data source(s) collected by someone other than the user for health service planning, performance reporting and evaluation, clinical decision-making and research.

There are several organizations that serves as repositories for secondary data in Canada and some examples are the Institute for Clinical Evaluative Sciences (ICES); Manitoba Centre for Health Policy (MCHP); Canadian Institute for Health Information (CIHI); Ontario Ministry of Health and Long-Term Care (MOHLTC); and Statistics Canada.

Examples of administrate databases include hospital discharge records, physician billing claims, ambulatory care records, prescription drug records, the Canadian Community Health Surveys (CCHS), and the Health Profiles.

Some advantages and disadvantages of secondary data studies:
Advantages:

- Provide large study populations and longer observation periods, allowing for examination of specific subpopulations
- Relatively inexpensive and expedient approach for answering time-sensitive questions posed by decision makers
- Reduces cost of data collection and reduces the administrative burden on organizations
- Ability to link with other data sources/databases
- Well suited for creating detailed and longitudinal analyses on subpopulations and regions (e.g., population health and healthcare utilization)

Disadvantages:

- Retrospective approach
- "Statistical" usability of a data source because the secondary data study uses a data source for a purpose different than the one for which the data was originally collected
- Lack of control of the content in the data source
- Lack of quality control over the data source
- Possibility of having missing items or missing records
- Difference in concepts which might lead to bias problems, as well as coverage problems
- Timeliness of the data

Previous secondary data studies have examined:

- Population who use the health care system
- Utilization and quality of health care services
- General health status and outcomes

- System characteristics, such as organization, funding, costing and efficiency

Checklist for secondary data studies:

PERMISSION:

- Prior to conducting any secondary data studies, investigators must check if the desired data can be permitted for secondary use. Investigators may also be required to complete an application in order to obtain permission and access to the data for analysis.

INTRODUCTION:

- Background and purpose:
    - Background information provides the broader context for the study and its relevance to research, outcomes and emerging policy priorities
    - Purpose of the study should be clearly stated and its potential implications for patients, practitioners and policymakers should be discussed

- Research questions:
    - An explicit set of research questions should be created to address the study population, interventions, comparisons, outcomes, and study design

- Hypotheses:
    - Hypotheses should be clearly stated to avoid opportunities for unsystematic data exploration. Therefore, it is important that the proposals show evidence of a well-developed data analysis plan for hypothesis-testing studies

RESEARCH DESIGN:

- Rationale:
    - Each research design has particular strengths and weaknesses depending on the setting, research question, and available data. As a result, the investigators should provide a clear reasoning for the selection of the design and explain why it is the best approach to answer the research questions

- Data analysis plan:
    - Define sample/cohort selection
        - The process of deriving the final sample from the population, such as the inclusion and exclusion criteria applied, should be described
    - Identify and justify the administrative dataset(s) used
        - Investigators should state explicitly what administrative databases are used for the study and why they are used (e.g., CIHI's Discharge Abstract Database (DAD) offer information on hospital discharges)

        - If linkage between administrative datasets is required, the investigators should also describe the procedures and criteria for the linkage
    - Define the time frame
        - It is important that the data is able to identify the intervention and outcomes if they actually occurred. This is because some procedures may not be routinely captured in the data, which may result in a lack of sample size for meaningful results. Some questions to consider are:
            - What is the start/end date?
            - What is the follow-up date?
            - When does the observation window end?
            - What time range is the look-back window?
    - Define relevant variables and outcomes
        - The investigators should identify all variables hypothesized to influence the outcome of interest and include all available variables in the model. Sometimes, databases lack information on some of the variables that would be expected to influence the outcome measure of interest. Failure to account for the effects of all variables that have a significant influence on the outcome of interest can lead to confounding bias.
            - What is the main exposure/risk factor?
            - What are baseline characteristics?
            - What are outcome definitions?
    - Statistical models/methods
        - Investigators should explain the reasoning for choosing the models/statistical methods used because statistical methods are based on a variety of underlying assumptions. For instance, studies that combine data from several databases should describe what analyses to be conducted to account for hierarchical or clustered data
        - Investigators should also outline approaches to test statistical assumptions and examine the validity of these assumptions underlying their analysis
    - Comparison group (if necessary)
        - For studies examining inferences about the effects of a particular intervention, they should describe the process for identifying the comparison group and the characteristics of the comparison group as they relate to the intervention group
    - Influential cases
        - Approaches to examine the sensitivity of the results due to influential cases should be explained. Sometimes, the findings of secondary data studies can be very sensitive to influential cases. Therefore, investigators should outline procedures to measure the impact of influential cases and if necessary, approaches to deal with influential cases
    - Missing values

        - Investigators should identify approaches to deal with missing values (e.g., complete-case analysis, multiple imputation) and examine the accuracy of the imputation methods for the missing data

- Design limitations:
    - Potential design limitations should be identified and addressed by the investigators. It is important that investigators describe potential biases and offer approaches to overcome these limitations

STRENGTHENING THE REPORTING OF OBSERVATIONAL STUDIES IN EPIDEMIOLOGY (STROBE) STATEMENT:

- The STROBE Statement was developed as a guideline for reporting observational studies of three main study designs: cohort, case-control, and cross-sectional studies. The STROBE Statement is comprised of 22 items, which relate to the title, abstract, introduction, methods, results and discussion sections of articles.

- The STROBE Statement is referred to in the Uniform Requirements for Manuscripts Submitted to Biomedical Journals by the International Committee of Medical Journal Editors.
    - The BMJ Open, British Medical Journal and Lancet are among many journals that refer to the STROBE Statement in their Instructions for manuscript submissions.

DATA QUALITY:

- Data quality has been defined as "the whole of planned and systematic procedures that take place before, during and after data collection to guarantee the quality of data in a database...for its intended use" (Iron & Manuel, 2007, pg. 2)

Data quality Indicators and definitions for administrative data (Iron & Manuel, 2007):

1. ***Are the data correct?***
    - Accuracy: Do the data reflect the truth?

2. ***Are the data complete?***
    - Completeness: Do the data include all records that are collected?
    - Comprehensiveness and coverage: Do the data cover 100% of the intended population?

3. ***Are the data reliable?***
    - Reliability: Are the data reproducible?
    - Validity: Do the data make sense?

4. ***Are the data useable?***

- Anonymity: Do the data adhere to jurisdictional privacy laws, procedures and practices?
- Linkability: Can the data be connected to other data to reflect health care system complexity?
- Timeliness: Is there a short lag between data collection and use?
- Usability: Are the data organized, accessible and provided in a format that can be easily used?
- Temporal consistency: Are the data elements standardized to evaluate change over time?

RESEARCH ETHICS:

- Investigators must outline the process of obtaining research ethics approval for the use of the secondary data (if applicable).

TEAM CHARACTERISTICS/EXPERTISE:

- Investigators with knowledge of the administrative datasets used and prior experience with conducting secondary data studies
- Biostatistician with a strong background in statistics, medical and health research and can provide methodological and statistical expertise

References:

- Daas et al. (2009). Checklist for quality evaluation of administrative data sources. Online. Statistics Netherlands. Available at: http://www.cbs.nl/nr/rdonlyres/0dbc2574-cdae-4a6d-a68a-88458cf05fb2/0/200942x10pub.pdf
- Iron, K & Manuel, D. (2007). Quality Assessment of Administrative Data (QuAAD): An Opportunity for Enhancing Ontario's Health Data. Online. ICES. Available at: http://www.ices.on.ca/Publications/Atlases-and-Reports/2007/Quality-assessment-of-administrative-data
- Motheral et al. (2003). A checklist for retrospective database studies – Report of the ISPOR Task Force on Retrospective Databases, Value Health 6:90-7.
- http://www.statcan.gc.ca/pub/12-539-x/2009001/administrative-administratives-eng.htm
- Ayanian, J.Z. (1999). Using administrative data to assess health care outcomes. European Heart Journal 20, 1689–169. Available at: http://eurheartj.oxfordjournals.org/content/20/23/1689.full.pdf
- Smith et al. (2011). Conducting High-Value Secondary Dataset Analysis: An Introductory Guide and Resources. J Gen Intern Med 26(8):920–9.
- Holosko, M.J. & Thyer, B.A. (2011). Pocket Glossary for Commonly Used Research Terms. SAGE.

- von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Ann Intern Med. 147(8):573-577.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case

Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders

		(b) Indicate number of participants with missing data for each variable of interest
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time
		Case-control study—Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Reference: von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Ann Intern Med. 147(8):573-577.