

## Minimum elements for reporting a multi-jurisdiction feasibility assessment of algorithms based on routinely collected health data: Health Data Research Network Canada recommendations

Naomi C. Hamm<sup>1</sup>, Sharon Bartholomew<sup>2</sup>, Yinshan Zhao<sup>3</sup>, Sandra Peterson<sup>4</sup>, Saeed Al-Azazi<sup>5</sup>, Kimberlyn McGrail<sup>4</sup>, and Lisa M. Lix<sup>1,\*</sup>

### Submission History

Submitted:	12/06/2024
Accepted:	21/10/2024
Published:	28/01/2025

<sup>1</sup>Department of Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada

<sup>2</sup>Centre for Surveillance and Applied Research, Public Health Agency of Canada, Ottawa, Canada

<sup>3</sup>Population Data BC, University of British Columbia, Vancouver, Canada

<sup>4</sup>Centre for Health Services and Policy Research, University of British Columbia, Vancouver, Canada

<sup>5</sup>George and Fay Yee Centre for Healthcare Innovation, University of Manitoba, Winnipeg, Canada

### Abstract

#### Background

Research and surveillance using routinely collected health data rely on algorithms or definitions to ascertain disease cases or health measures. Whenever algorithm validation studies are not possible due to the unavailability of a reference standard, algorithm feasibility studies can be used to create and assess algorithms for use in more than one population or jurisdiction. Publication of the methods used to conduct feasibility studies is critical for reproducibility and transparency. Existing guidelines applicable to feasibility studies include the STRENGTHENING the REPORTING of OBSERVATIONAL studies in EPIDEMIOLOGY (STROBE) and REPORTING of STUDIES CONDUCTED using OBSERVATIONAL ROUTINELY COLLECTED HEALTH DATA (RECORD) guidelines. These guidelines may benefit from additional elements that capture aspects particular to multi-jurisdiction algorithm feasibility studies and ensure their reproducibility. The aim of this paper is to identify the minimum elements for reporting feasibility studies to ensure reproducibility and transparency.

#### Methods

A subcommittee of four individuals with expertise in routinely collected health data, multi-jurisdiction health research, and algorithm development and implementation was formed from Health Data Research Network (HDRN) Canada's Algorithms and Harmonized Data Working Group (AHD-WG). The subcommittee reviewed items within the STROBE and RECORD guidelines and evaluated these items against published feasibility studies. Items to ensure transparent reporting of feasibility studies not contained within STROBE or RECORD guidelines were identified through consensus by subcommittee members using the Nominal Group Technique. The AHD-WG reviewed and approved these additional recommended elements.

#### Results

Eleven new recommended elements were identified: one element for the title and abstract, one for the introduction, five for the methods, and four for the results sections. Recommended elements primarily addressed reporting jurisdictional data variabilities, data harmonization methods, and algorithm implementation techniques.

#### Significance

Implementation of these recommended elements, alongside the RECORD guidelines, is intended to encourage consistent publication of methods that support reproducibility, as well as increase comparability of algorithms and their use in national and international studies.

#### Keywords

Algorithms; feasibility studies; administrative data; transparency

\*Corresponding Author:

Email Address: [lisa.lix@umanitoba.ca](mailto:lisa.lix@umanitoba.ca) (Lisa M. Lix)

## Introduction

Research and surveillance studies that use routinely collected health data often rely on established algorithms (i.e., case or concept definitions) to identify diseases or conditions [1, 2] or to measure health characteristics (e.g., co-morbidity indices) [3] and health determinants (e.g., smoking status, health service use) [4]. Ideally, these algorithms are validated against a reference standard to assess their accuracy; multi-jurisdiction validation studies are essential to determine generalizability [5]. In the absence of a reference standard, a feasibility study is useful to assess whether an algorithm, either existing or newly developed, is reasonable for use in more than one population or jurisdiction. Feasibility studies assess algorithm face validity, the availability of data elements, and potential utility of the algorithm across different settings [6]. Results from feasibility studies provide important information for researchers who may wish to use these algorithms or concepts with confidence that they meet scientific standards. This is particularly true for population studies that span multiple jurisdictions [7]. Despite this, publication of feasibility studies in scientific literature is limited compared to that of validation studies.

Transparent reporting of feasibility studies is an important facilitator for algorithm reuse and contributes to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles of open data and open science [8]. Publication of feasibility studies ensures they are subject to peer review and that the process of algorithm development and evaluation is described in sufficient detail to ensure it can be replicated and/or reproduced in other contexts. Guidelines for reporting studies that use routinely collected health data include the REporting of studies Conducted using Observational Routinely collected health Data (RECORD) guidelines [9], which provides a checklist of items for transparent reporting observational research using routinely collected health data. The RECORD statement/reporting guidelines is an extension of the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines [10]. However, as feasibility studies focus specifically on applying algorithms and assessing their face validity across multiple jurisdictions, additional elements may need to be reported to ensure full reproducibility and transparency. Clear guidelines on reporting feasibility studies encourages both researchers and journals to publish feasibility studies and facilitate greater trust in the resulting publications.

Health Data Research Network (HDRN) Canada is a non-profit corporation founded in 2020 with the aim of facilitating the use of multi-jurisdictional electronic health and social data to drive improvements in health and health equity [11]. Strategic goals of HDRN Canada include diversifying and harmonizing data and supporting innovative and privacy-sensitive data use. The Algorithms and Harmonized Data Working Group (AHD-WG), comprised of individuals from HDRN Canada's 14 member organizations, oversees the development of data resources and their metadata, as well as measures of population health, health service use, and determinants of health that can be used for research across Canadian provinces and territories. Goals of AHD-WG include establishing a process for multi-jurisdiction algorithm validation as well as developing and validating algorithms to be shared across jurisdictions. Conducting and reporting

studies that evaluate the feasibility of applying algorithms across jurisdictions is an important step to achieve these goals. Our aim is to provide guidance for feasibility study reporting in support of reproducibility and transparency of algorithm measurement across different contexts.

This paper includes a checklist of the recommended minimum elements for reporting algorithm feasibility studies to facilitate their reproducibility and transparency. These elements are meant to be used in addition to the STROBE and RECORD reporting guidelines.

## Methods

### Subcommittee formation and process

A subcommittee consisting of four individuals from HDRN's AHD-WG was formed (NCH, SB, YZ, and SP). Individuals were selected via volunteer. Members of the subcommittee had expertise in administrative health data, electronic medical record data, chronic disease surveillance, multi-jurisdiction health research, and algorithm development and implementation. Subcommittee members had experience in software programming, quantitative data analysis, and conducting methods research using secondary health data.

Consensus on minimum element identification and item categorization (see *Literature Review* and *Item Categorization* subsections) was reached using the Nominal Group Technique (NGT) [12]. Subcommittee members were located across Canada; therefore, to accommodate the different locations and time zones, the NGT method was employed using shared documents and one teleconference meeting over a three-month period. Minimum element assessments and item categorizations were generated independently by each subcommittee member and shared via written documents. Subcommittee members could then comment, clarify, or discuss ideas using comments within a shared document. Based on the comments, ideas were re-assessed as needed until 100% consensus was reached. If consensus could not be reached within the three-month work period, the majority opinion was used as the final decision.

### Literature review

The committee reviewed published algorithm feasibility studies contained within HDRN's Algorithms Inventory [13, 14]. The Algorithms Inventory is an existing repository created through a systematic review process that aims to identify algorithms for administrative health data that were developed and/or validated in two or more Canadian provinces/territories. The algorithms are extracted from published studies; they encompass concepts of population health, health services, and determinants of health. At the time of this work, the Algorithms Inventory contained studies published between 1989 and 2021; the majority were published in 2010 or later. Details of the search strategy, screening, and quality control of the Algorithms Inventory can be found in the supplemental material. Of the 58 feasibility studies contained in the inventory, about 20 articles that were listed under the health domain (i.e., applied algorithms for health conditions) and were open access were randomly selected for review. If

members of the subcommittee knew of articles that provided a strong example of transparency, these were also included in the review. Grey literature was considered if it was recommended by subcommittee members to provide examples of transparent reporting or justification of recommended minimum elements. Studies were first reviewed to identify elements that were commonly reported. Then, studies were assessed to determine which elements were needed to ensure reproducibility and transparency. Elements were assessed as “minimum” or “not minimum”. A minimum element was defined as an element where the study was not reproducible in its absence.

## Item categorization

Minimum elements were compared to the items contained in the STROBE and RECORD guidelines (see Table 1) and three different item categories were created:

- 1) Sufficient items, where the existing STROBE and RECORD items sufficiently covered the details required for transparent reporting of feasibility studies;
- 2) Additional context items, where the existing STROBE and RECORD items covered the details required for transparent reporting; however, additional context was required to apply the item to feasibility studies;
- 3) New items, where the minimum element was not included in the STROBE or RECORD items. Explanations were provided along with references to justify inclusion of new items.

## Algorithms and harmonized data working group review

Subcommittee members compiled the final list of items (sufficient, additional context, and new) and explanations for new items into a report and presentation. The presentation was given to members of HDRN Canada’s AHD-WG and members were invited to comment or provide feedback during a WG meeting. The written report was also distributed to the AHD-WG and members were given three weeks to provide any additional comments. AHD-WG members were asked to review the content, as well as assess for readability and interpretation. Comments were incorporated into the report by NCH with other subcommittee members providing input. The final report was approved by all members of HDRN Canada’s AHD-WG.

## Results

Review of published feasibility studies alongside the current STROBE and RECORD guidelines found that the STROBE and RECORD guidelines were sufficient in many cases, particularly in the Title and Abstract, Introduction, and Discussion sections. Six additional context items were identified in the Introduction, Methods, and Discussion sections. Limitations of the STROBE and RECORD guidelines primarily pertained to methods for standardizing heterogenous data across different jurisdictions, algorithm details, and reporting overall and stratified algorithm results. Eleven new

items for reporting multi-jurisdictional feasibility studies were identified to address these limitations.

The new items for reporting algorithm feasibility studies recommended by HDRN Canada’s AHD-WG are included in Table 1. These items have been grouped with the STROBE and RECORD items under the relevant section (i.e., Title and Abstract, Introduction, Methods, Results, Discussion, and Other Information). A description and explanation for each new item is included. For additional context items, information that may be helpful for applying those items in the framework of algorithm feasibility studies is also provided.

## Title and abstract section

### New item 1

Item 1.a: Report the concepts or variables for which algorithm feasibility is assessed.

### Explanation

Identification of the concept or variable (e.g., dementia, healthcare service costs, diabetic related hospital visits) relevant to the algorithm(s) in the title or abstract will help other researchers to find the algorithm in published literature. If the feasibility of algorithms for multiple concepts or variables are being assessed, authors should clearly state each individual concept or variable.

## Introduction section: background rationale

### Additional context item 1

STROBE Item 2: Explain the scientific background and rationale for the investigation being reported.

### Explanation

The STROBE guidelines advise the authors to “explain the scientific background and rationale for the investigation being reported”. Authors should highlight the rationale for investigating the algorithm(s), the intended usage, if the algorithm(s) has been applied or validated in any setting previously, and relevant results and references for previous algorithm application or validation.

### New item 2

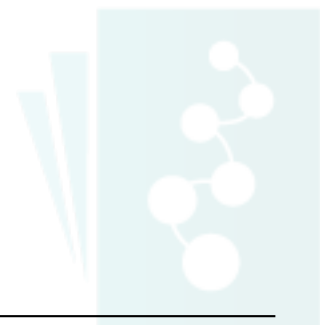
Item 2.a: Report the organization/system, project, or program behind the study, if applicable.

### Explanation

Surveillance systems and research projects or programs that conduct distributed or federated analyses often conduct algorithm feasibility studies first to ensure the variables of interest can be obtained across different jurisdictions or health databases [6, 15]. Examples include the Canadian Network for Observational Drug Effect Studies (CNODES) and the Public Health Agency of Canada (PHAC) [6, 15]. Authors should clearly state if the feasibility study is related to a research project or program, to facilitate transparency.

Table 1: STROBE items, RECORD items, with recommended additional context (in italics) and HDRN Canada’s recommended new items for reporting algorithm feasibility studies

	Item no.	STROBE items	RECORD* items	HDRN Canada’s recommended new items
<b>Title and abstract</b>				
	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	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.a Report the concepts or variables for which algorithm feasibility is assessed.
<b>Introduction</b>				
Background rationale	2	Explain the scientific background and rationale for the investigation being reported <i>(Highlight the rationale and intended usage, and, if applicable, provide results and references for previous algorithm application or validation)</i>		2.a Report the organization/ system, project, or program behind the study if applicable.
Objectives	3	State specific objectives, including any prespecified hypotheses		
<b>Methods</b>				
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 <i>(Highlight jurisdictions included and why they were selected)</i>		
Participants	6	(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> - 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	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. <i>(Highlight healthcare coverage requirements for inclusion in the study population)</i>	



Continued

Table 1: Continued

	Item no.	STROBE items	RECORD* items	HDRN Canada's recommended new items
		<p><i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p><i>(b) Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed</p> <p><i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>	<p>RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.</p> <p>RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.</p>	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	<p>RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.</p>	7.1.a Describe: (1) algorithm implementation details across jurisdictions, including type of data source, fields used from each data source, timeframe/ observation window, lookback window details; (2) any specifications such as diagnosis/ medication/procedure code(s), number of contacts required, or other definition details; and (3) any other specifications, such as fee/payment codes for a service (delivered or received) or location codes specifying a service location. Where the algorithm includes multiple healthcare contacts or other longitudinal considerations, state how the onset date/classification date is defined (i.e., date of first contact or contact where criteria are fulfilled).
Data sources/ measurement	8	<p>For each variable of interest, give sources of data and details of methods of assessment (measurement).</p> <p>Describe comparability of assessment methods if there is more than one group</p>		8.a Describe key elements of data sources for each jurisdiction including time period covered, population covered, diagnosis code type and number of digits, and maximum number of diagnosis codes per claim. Describe data comparability across jurisdictions.
Bias	9	Describe any efforts to address potential sources of bias		
Study size	10	Explain how the study size was arrived at		

Continued

Table 1: Continued

	Item no.	STROBE items	RECORD* items	HDRN Canada's recommended new items
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) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses		12.a Describe and justify the methods selected to assess feasibility, such as face validity.
Data access and cleaning methods	..	..	RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	12.1.a Describe methods to address data sharing/confidentiality. 12.1.b Describe methods to align data sources/measurements/variables across jurisdictions (e.g., crosswalk algorithms).
Linkage	..	..	RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
<b>Results</b>				
Participants	13	(a) Report the numbers of individuals at each stage of the study (e.g., 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	RECORD 13.1: Describe in detail the selection of the persons included in the study (i.e., study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	13.1.a Describe selection of persons included in the study, stratified by jurisdiction.

Continued

Table 1: Continued

	Item no.	STROBE items	RECORD* items	HDRN Canada's recommended new items
Descriptive data	14	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) <i>Cohort study</i> - summarise follow-up time (e.g., average and total amount)		14.a Report descriptive data for study population, stratified by jurisdiction.
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> - Report numbers of outcome events or summary measures		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 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		16.a Describe algorithm applied to the study data. If the algorithm was developed or modified from an existing algorithm, provide a rationale for the decisions made that resulted in final algorithm. 16.b Report results from any face validity assessment of the algorithm by jurisdiction.
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses		
<b>Discussion</b>				
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	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	



Continued

Table 1: Continued

	Item no.	STROBE items	RECORD* items	HDRN Canada's recommended new items
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence <i>(Highlight known jurisdictional differences that could explain or provide insight on observed discrepancies in algorithm performance. State recommendations on future development, validation, or implementation, if applicable)</i>	<i>(Highlight the implication of using data from multiple jurisdictions and how differences in data availability may impact algorithm application and observed feasibility.)</i>	
Generalisability	21	Discuss the generalisability (external validity) of the study results <i>(Highlight challenges associated with feasibility over time and/or across geography or context, and how this may affect the long-term feasibility of the algorithm)</i>		
<b>Other Information</b>				
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		
Accessibility of protocol, raw data, and programming code	..	..	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	

\*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015 Oct 6;12(10):e1001885. <https://doi.org/10.1371/journal.pmed.1001885>.

STROBE: STrengthening the Reporting of OBservational studies in Epidemiology.

RECORD: REporting of studies Conducted using Observational Routinely collected health Data.

## Methods section: setting

### Additional context item 2

STROBE Item 5: Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.

### Explanation

The STROBE guidelines advise authors to “Describe the setting, locations, and relevant dates...”. A feasibility study spans multiple jurisdictions, and therefore authors should



highlight what jurisdictions are included in the study and why the jurisdictions were selected (i.e., data availability, funding, disease or condition of particular interest to those jurisdictions, etc.).

## Methods section: participants

### Additional context item 3

RECORD Item 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.

#### Explanation

The RECORD guidelines advise authors to report “the methods of study population selection (such as codes or algorithms used to identify subjects)”. Healthcare coverage is often a consideration when using electronic health data, therefore authors should highlight any healthcare coverage requirements for inclusion in the study population.

## Methods section: variables

### New item 3

Item 7.1.a: Describe: (1) algorithm implementation details across jurisdictions, including type of data source, fields used from each data source, timeframe/observation window, lookback window details; (2) any specifications such as diagnosis/medication/procedure code(s), number of contacts required, or other definition details; and (3) any other specifications, such as fee/payment codes for a service (delivered or received) or location codes specifying a service location. Where the algorithm includes multiple healthcare contacts or other longitudinal considerations, state how the onset date/classification date is defined (i.e., date of first contact or contact where criteria are fulfilled).

#### Explanation

The RECORD guidelines advise “a complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers...”. As feasibility studies focus on applying the algorithm and assessing feasibility across multiple jurisdictions, it is recommended that all algorithm elements be provided for each jurisdiction. Elements of an algorithm to be reported (type of data source, observation window, diagnostic/medication/procedure code(s), and number of contacts for each data source) are based on previous publications on algorithm development [16]. In addition, authors should report details on lookback windows (i.e., retrospective period of time used to determine incidence or prevalence), as different lookback windows can influence misclassification rates and comparability of incidence and prevalence estimates [17]. In cases where the variables of interest are not specific to a disease or condition, but rather a health service used or determinant of health, authors should report other elements such as fee codes and location codes where applicable [18]. As fee and location codes can vary by jurisdiction, authors are encouraged to consider including a

supplementary table containing a comprehensive description of all codes used, stratified by jurisdiction, to help align the constructs being measured.

## Methods section: data sources/measurement

### New item 4

Item 8.a: Describe key elements of data sources for each jurisdiction including time period covered, population covered, diagnosis code type and number of digits, and maximum number of diagnosis codes per claim. Describe data comparability across jurisdictions.

#### Explanation

The quality of routinely collected data for research may be impacted by differences in healthcare systems across jurisdictions [19, 20]. Moreover, when certain populations are not included in the database, algorithm generalizability may be affected. Therefore, it is recommended that authors report the time period and populations covered by each database used across the different jurisdictions.

Variations in the data across jurisdictions can impact algorithms' accuracy and implementation [21, 22]. Authors should clearly describe the data comparability across jurisdictions. Creating common algorithms may require using fewer data elements, sources, or years than what are available in select jurisdictions to meet the 'lowest common denominator'. For algorithms developed for diseases or conditions, reporting the diagnosis code type and number of digits, and maximum number of diagnosis codes per claim across all databases provides information on if and where databases contain more elements than what was used in the applied common algorithm. For algorithms developed for health service use or determinants of health, details on data comparability, which may involve different data elements and coding systems across jurisdictions, should be included. Where there is variation in the data across jurisdictions, methods that may improve data consistency across jurisdictions should be fully described (*See New Item 7*).

## Methods section: statistical methods

### New Item 5

Item 12.a: Describe and justify the methods selected to assess feasibility, such as face validity.

#### Explanation

Feasibility studies aim to apply algorithms across multiple jurisdictions and assess their face validity. There is no single method or approach to assess face validity; authors should therefore provide justification for why the selected face validity method or approach is appropriate. Examples of methods or approaches to assess face validity can be found in the *Method Examples for Face Validity Assessment* section along with links to published feasibility studies that have used these methods. Other examples of feasibility studies can be found in HDRN Canada's Algorithms Inventory (<https://www.hdrn.ca/en/algorithm/>).

## Methods section: data access and cleaning methods

### New item 6

Item 12.1.a: Describe methods to address data sharing/confidentiality.

#### Explanation

Authors should report methods for addressing data sharing/confidentiality across the study jurisdictions. For example, the Canadian Chronic Disease Surveillance System (CCDSS) uses a distributed data analysis system, where algorithm implementation methods (i.e., statistical code) are developed centrally and distributed among participating jurisdictions to be applied locally. Aggregated results are then shared back to the CCDSS for reporting [6, 23]; no line-level data are shared by individual jurisdictions.

### New item 7

Item 12.1.b: Describe methods for aligning data sources/measurements/variables across jurisdictions (e.g., crosswalk algorithms).

#### Explanation

Applying common algorithms across multiple jurisdictions may require methods to standardize or align the data sources. Authors should describe any methods used, such as crosswalk algorithms (i.e., method to convert data from one coding type/version to another coding type/version) [24] or common data models [25]. Methods to address missing and invalid data across jurisdictions should also be described.

## Results section: participants

### New item 8

Item 13.1.a: Describe selection of persons included in the study, stratified by jurisdiction.

#### Explanation

Selection of persons may be different across jurisdictions. Authors should include the information required by the STROBE item 13 and RECORD item 13.1 stratified by jurisdiction.

## Results section: descriptive data

### New item 9

Item 14.a: Report descriptive data for study population, stratified by jurisdiction.

#### Explanation

Algorithm feasibility may be influenced by population characteristics across jurisdictions. Therefore, authors should include descriptive information required by STROBE stratified by jurisdiction.

## Results section: main results

### New item 10

Item 16.a: Describe algorithm applied. If the algorithm was developed or modified from an existing algorithm, provide rationale for the decisions made that resulted in the final algorithm.

#### Explanation

Authors should provide a full description of the algorithm that was applied across jurisdictions in the feasibility study. This information should encompass information about the type of data source, timeframe/observation window, lookback window details, diagnostic/medication/procedure/fee/location code(s), and number of contacts required for each data source. This may be similar to the methods section that reported on algorithm implementation. As the main objective for feasibility studies is the implementation of common algorithms across jurisdictions, descriptions of the applied algorithms should be reported as a main result. Where multiple algorithms were considered, only the ones that were applied to the data should be reported as a result. In addition, authors should provide rationale or justification for the choices made during the algorithm development process, if applicable.

### New item 11

Item 16.b: Report results from any face validity assessment of the algorithm by jurisdiction.

#### Explanation

Authors should report results stratified by jurisdiction, in addition to overall results (if applicable). If multiple algorithms were assessed, results should be reported separately for each algorithm. Reporting stratified results allows readers to assess the application of the reported algorithms across different populations/databases and its variability and generalizability.

## Discussion section: limitations

### Additional context item 4

RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.

#### Explanation

RECORD guidelines advise authors to “discuss the implications of using data that were not created or collected to answer the specific research question(s)”. Authors should highlight the implication of using data from multiple jurisdictions and how differences in data availability may impact algorithm application and observed feasibility.

## Discussion section: interpretation

### Additional context item 5

STROBE Item 20: Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.

#### Explanation

STROBE guidelines advise authors to “give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence”. Authors should highlight known jurisdictional differences (e.g., difference in healthcare population coverage, different healthcare delivery models) that could explain or provide insight on observed discrepancies in algorithm performance across jurisdictions. Authors should also state any recommendations on future algorithm development, validation, or implementation, if applicable.

## Discussion section: generalisability

### Additional context item 6

STROBE Item 21: Discuss the generalisability (external validity) of the study results.

#### Explanation

STROBE guidelines advise authors to “Discuss the generalisability (external validity) of the study results”. When discussing generalisability, authors should highlight potential challenges associated with algorithm feasibility over time and/or across geography or context, and how this may affect the long-term feasibility of the algorithm. Routinely collected health data are susceptible to data drift [26], and thus feasibility of the algorithms may not be consistent over time.

## Method examples for face validity assessment

### Calculate rates (prevalence/incidence/mortality) over time

:

- Blais C, Dai S, Waters C, et al. Assessing the burden of hospitalized and community-care heart failure in Canada. *Can J Cardiol* 2014; 30: 352–358. <https://doi.org/10.1016/j.cjca.2013.12.013>
- Robitaille C, Bancej C, Dai S, et al. Surveillance of ischemic heart disease should include physician billing claims: population-based evidence from administrative health data across seven Canadian provinces. *BMC Cardiovasc Disord* 2013; 13: 88. <https://doi.org/10.1186/1471-2261-13-88>
- Marrie RA, Fisk JD, Tremlett H, et al. Differences in the burden of psychiatric comorbidity in MS vs the

general population. *Neurology* 2015; 85: 1972–1979. <https://doi.org/10.1212%2FWNL.0000000000002174>

- Guy P, Sheehan KJ, Morin SN, et al. Feasibility of using administrative data for identifying medical reasons to delay hip fracture surgery: a Canadian database study. *BMJ Open* 2017; 7: e017869. <https://doi.org/10.1136/bmjopen-2017-017869>
- Bernstein CN, Kuenzig ME, Coward S, et al. Increased Incidence of inflammatory bowel disease after Hirschsprung disease: a population-based cohort study. *J Pediatr* 2021; 233: 98-104.e2. <https://doi.org/10.1016/j.jpeds.2021.01.060>

### Calculate rates (prevalence/incidence/mortality) stratified by demographic characteristics such as age, sex, or residence location

- Blais C, Dai S, Waters C, et al. Assessing the burden of hospitalized and community-care heart failure in Canada. *Can J Cardiol* 2014; 30: 352–358. <https://doi.org/10.1016/j.cjca.2013.12.013>
- Shiff NJ, Lix LM, Oen K, et al. Chronic inflammatory arthritis prevalence estimates for children and adolescents in three Canadian provinces. *Rheumatol Int* 2015; 35: 345–350. <https://doi.org/10.1007/s00296-014-3085-0>
- Robitaille C, Bancej C, Dai S, et al. Surveillance of ischemic heart disease should include physician billing claims: population-based evidence from administrative health data across seven Canadian provinces. *BMC Cardiovasc Disord* 2013; 13: 88. <https://doi.org/10.1186/1471-2261-13-88>
- Marrie RA, Fisk JD, Tremlett H, et al. Differences in the burden of psychiatric comorbidity in MS vs the general population. *Neurology* 2015; 85: 1972–1979. <https://doi.org/10.1212%2FWNL.0000000000002174>
- Guy P, Sheehan KJ, Morin SN, et al. Feasibility of using administrative data for identifying medical reasons to delay hip fracture surgery: a Canadian database study. *BMJ Open* 2017; 7: e017869. <https://doi.org/10.1136/bmjopen-2017-017869>
- Godwin M, Williamson T, Khan S, et al. Prevalence and management of hypertension in primary care practices with electronic medical records: a report from the Canadian Primary Care Sentinel Surveillance Network. *CMAJ Open* 2015; 3: E76–E82. <https://doi.org/10.9778/cmajo.20140038>

### Calculate rate ratios over time

- Marrie RA, Fisk JD, Tremlett H, et al. Differences in the burden of psychiatric comorbidity in MS vs the general population. *Neurology* 2015; 85: 1972–1979. <https://doi.org/10.1212%2FWNL.0000000000002174>

## Evaluate associations between outcome variables derived using the algorithm and risk factors for the outcome

- Marrie RA, Garland A, Schaffer SA, et al. Traditional risk factors may not explain increased incidence of myocardial infarction in MS. *Neurology* 2019; 92: e1624–e1633. <https://doi.org/10.1212/wnl.00000000000007251>
- Bernstein CN, Kuenzig ME, Coward S, et al. Increased incidence of inflammatory bowel disease after Hirschsprung disease: a population-based cohort study. *J Pediatr* 2021; 233: 98–104.e2. <https://doi.org/10.1016/j.jpeds.2021.01.060>
- Godwin M, Williamson T, Khan S, et al. Prevalence and management of hypertension in primary care practices with electronic medical records: a report from the Canadian Primary Care Sentinel Surveillance Network. *CMAJ Open* 2015; 3: E76–E82. <https://doi.org/10.9778/cmajo.20140038>

## Describe characteristics of persons who meet algorithm eligibility criteria

- Guy P, Sheehan KJ, Morin SN, et al. Feasibility of using administrative data for identifying medical reasons to delay hip fracture surgery: a Canadian database study. *BMJ Open* 2017; 7: e017869. <https://doi.org/10.1136/bmjopen-2017-017869>
- Marrie RA, Garland A, Schaffer SA, et al. Traditional risk factors may not explain increased incidence of myocardial infarction in MS. *Neurology* 2019; 92: e1624–e1633. <https://doi.org/10.1212/wnl.00000000000007251>
- Bernstein CN, Kuenzig ME, Coward S, et al. Increased Incidence of inflammatory bowel disease after Hirschsprung disease: a population-based cohort study. *J Pediatr* 2021; 233: 98–104.e2. <https://doi.org/10.1016/j.jpeds.2021.01.060>
- Godwin M, Williamson T, Khan S, et al. Prevalence and management of hypertension in primary care practices with electronic medical records: a report from the Canadian Primary Care Sentinel Surveillance Network. *CMAJ Open* 2015; 3: E76–E82. <https://doi.org/10.9778/cmajo.20140038>

## Conclusion

Algorithm feasibility studies are intended to assess the application of one or more algorithms for routinely collected health data across multiple jurisdictions. Here, we provide recommendations from HDRN Canada's AHD-WG on the minimum elements for reporting algorithm feasibility studies. These minimum elements address many of the challenges associated with using data from multiple jurisdictions in observational studies and are intended to aid researchers in describing jurisdictional data variations

and efforts to standardize or harmonize data across jurisdictions. Recommendations are meant to augment RECORD and STROBE reporting guidelines and represent the minimum standard for producing reproducible and transparent information about algorithms. By adopting these recommended minimum elements for reporting, future feasibility studies will have improved transparency and better inform algorithm development and reuse. Moreover, providing clearer standards for reporting feasibility studies encourages both researchers and journals to publish algorithm feasibility work. Additional information that goes beyond the minimum requirements, such as open access to code production for algorithm implementation, should also be considered when working towards the requirements of open data and open science.

Strengths of this work are that it was undertaken by Canadian experts in algorithm development and implementation for routinely collected data. As well, this work was led by HDRN Canada, which has taken a leadership role in data harmonization and algorithm standardization across Canadian provinces and territories [27].

There are some limitations to this work. Articles screened to develop the recommended minimum elements for reporting feasibility studies were limited to health domain articles within HDRN Canada's Algorithms inventory. The inventory is based on previously published studies, which may inherently reflect specific regions or populations that have been more extensively studied (e.g., provinces like British Columbia and Ontario are more frequently represented). Additionally, studies included in the review were in English only, which could introduce language bias and may have a disproportionate focus on high-income regions. This limits the generalizability of the findings to non-health domain studies, lower-income regions, and non-English language studies. In addition, content expertise of the subcommittee may not have been comprehensive, as there were only four members. To mitigate this, the minimum elements, explanations, and references were reviewed by all members of HDRN Canada's AHD-WG, which at the time of this work was comprised of more than 14 organizations.

In summary, transparent reporting and publication of feasibility studies is critical for ensuring the scientific rigor of algorithm development and facilitating algorithm reuse. By recommending minimum elements for transparent reporting, HDRN Canada provides a framework for transparent reporting and publication of feasibility studies. HDRN Canada is a nation-wide effort to enable multi-jurisdictional research through harmonized electronic datasets and national data linkages [27]. Work towards this effort includes providing a single portal for data access, support for federated and distributed data analyses, and developing and standardizing algorithms for priority disease. The methods used here are a reflection of HDRN Canada's role and expertise in algorithms development and data harmonization and may not generalize to other types of studies.

## Acknowledgements

The authors would like to sincerely thank the HDRN AHD-WG for the insights and input provided on the recommended minimum elements as well as their explanations.

## Conflict of interest

LML receives research funding from the Canadian Institutes of Health Research. The other authors have no conflicts of interest to declare.

## Ethics

This work did not require ethical approval as it did not involve primary data collection, secondary data analysis, or non-public record review.

## Data availability statement

Data for this work was obtained through review of published articles contained within Health Data Research Network Canada's Algorithms Inventory, at <https://www.hdrn.ca/en/dash/algorithms/algorithms-for-multi-regional-research/>

## References

1. Robitaille C, Bancej C, Dai S, Tu K, Rasali D, Blais C, et al. Surveillance of ischemic heart disease should include physician billing claims: Population-based evidence from administrative health data across seven Canadian provinces. *BMC Cardiovasc Disord* 2013;13:88. <https://doi.org/10.1186/1471-2261-13-88>
2. O'Donnell S, Canadian Chronic Disease Surveillance System (CCDSS) Osteoporosis Working Group. Use of administrative data for national surveillance of osteoporosis and related fractures in Canada: Results from a feasibility study. *Arch Osteoporos* 2013;8:143. <https://doi.org/10.1007/s11657-013-0143-2>
3. Simard M, Sirois C, Candas B. Validation of the Combined Comorbidity Index of Charlson and Elixhauser to predict 30-Day mortality across ICD-9 and ICD-10. *Med Care* 2018;56:441–7. <https://doi.org/10.1097/MLR.0000000000000905>
4. Havard A, Jorm LR, Lujic S. Risk adjustment for smoking identified through tobacco use diagnoses in hospital data: A validation study. *PloS One* 2014;9:e95029. <https://doi.org/10.1371/journal.pone.0095029>
5. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–9. <https://doi.org/10.1016/j.jclinepi.2010.10.006>
6. Lix LM, Ayles J, Bartholomew S, Cooke CA, Ellison J, Emond V, et al. The Canadian Chronic Disease Surveillance System: A model for collaborative surveillance. *Int J Popul Data Sci* n.d.;3:433. <https://doi.org/10.23889/ijpds.v3i3.433>
7. Feely A, Lix LM, Reimer K. Estimating multimorbidity prevalence with the Canadian Chronic Disease Surveillance System. *Health Promot Chronic Dis Prev Can Res Policy Pract* 2017;37:215–22. <https://doi.org/10.24095/hpcdp.37.7.02>
8. Haven T, Gopalakrishna G, Tjldink J, van der Schot D, Bouter L. Promoting trust in research and researchers: How open science and research integrity are intertwined. *BMC Res Notes* 2022;15:302. <https://doi.org/10.1186/s13104-022-06169-y>
9. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Med* 2015;12:e1001885. <https://doi.org/10.1371/journal.pmed.1001885>
10. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *BMJ* 2007;335:806–8. <https://doi.org/10.1136/bmj.39335.541782.AD>
11. McGrail K, Diverty B, Lix L. Introducing Health Data Research Network Canada (HDRN Canada): A new organization to advance Canadian and international population data Science. *Int J Popul Data Sci* 2020;5. <https://doi.org/10.23889/ijpds.v5i5.1493>
12. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016;38:655–62. <https://doi.org/10.1007/s11096-016-0257-x>
13. Lix LM, Smith M, Wu J, Al-Azazi S, Dahl L, Poppel A, et al. Canadian Data Platform: Developing an algorithm inventory for health and social measures. *Int J Popul Data Sci* 2020;5. <https://doi.org/10.23889/ijpds.v5i5.1499>
14. Lix L, Vasylykiv V, Ayilara O, Dahl L, Poppel A, Al-Azazi S. A Synthesis of algorithms for multi-jurisdiction research in Canada. *Int J Popul Data Sci* 2022;7. <https://doi.org/10.23889/ijpds.v7i3.1911>
15. Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med Peer-Rev Indep Open-Access J* 2012;6:e134–140.
16. Lix, L, Yogendran, M, Burchill, Charles, Metge, C, McKeen, N, Moore, D, et al. Defining and validating chronic diseases: An administrative data approach. Winnipeg, MB: Manitoba Centre for Health Policy; 2006.
17. Nanditha NGA, Dong X, McLinden T, Sereda P, Kopec J, Hogg RS, et al. The impact of lookback windows on the prevalence and incidence of chronic diseases among people living with HIV: An exploration in administrative health data in Canada. *BMC Med Res Methodol* 2022;22:1. <https://doi.org/10.1186/s12874-021-01448-x>
18. Lavergne MR, Rudoler D, Peterson S, Stock D, Taylor C, Wilton AS, et al. Declining comprehensiveness of services

- delivered by Canadian family physicians is not driven by early-career physicians. *Ann Fam Med* 2023;21:151–6. <https://doi.org/10.1370/afm.2945>
19. Lavergne MR, Law MR, Peterson S, Garrison S, Hurley J, Cheng L, et al. Effect of incentive payments on chronic disease management and health services use in British Columbia, Canada: Interrupted time series analysis. *Health Policy Amst Neth* 2018;122:157–64. <https://doi.org/10.1016/j.healthpol.2017.11.001>
  20. Hamm NC, Pelletier L, Ellison J, Tennenhouse L, Reimer K, Paterson JM, et al. Trends in chronic disease incidence rates from the Canadian Chronic Disease Surveillance System. *Health Promot Chronic Dis Prev Can Res Policy Pract* 2019;39:216–24. <https://doi.org/10.24095/hpcdp.39.6/7.02>
  21. Kisely S, Lin E, Lesage A, Gilbert C, Smith M, Campbell LA, et al. Use of administrative data for the surveillance of mental disorders in 5 provinces. *Can J Psychiatry Rev Can Psychiatr* 2009;54:571–5. <https://doi.org/10.1177/070674370905400810>
  22. Doyle CM, Lix LM, Hemmelgarn BR, Paterson JM, Renoux C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf* 2020;29:68–77. <https://doi.org/10.1002/pds.4889>
  23. Public Health Agency of Canada. The Canadian Chronic Disease Surveillance System – An Overview. Government of Canada; 2018.
  24. Hamad AF, Vasylyk V, Yan L, Sanusi R, Ayilara O, Delaney JA, et al. Mapping three versions of the International Classification of Diseases to categories of chronic conditions. *Int J Popul Data Sci* 2021;6. <https://doi.org/10.23889/ijpds.v6i1.1406>
  25. Henke E, Zoch M, Peng Y, Reinecke I, Sedlmayr M, Bathelt F. Conceptual design of a generic data harmonization process for OMOP common data model. *BMC Med Inform Decis Mak* 2024;24:58. <https://doi.org/10.1186/s12911-024-02458-7>
  26. Rahmani K, Thapa R, Tsou P, Casie Chetty S, Barnes G, Lam C, et al. Assessing the effects of data drift on the performance of machine learning models used in clinical sepsis prediction. *Int J Med Inf* 2023;173:104930. <https://doi.org/10.1016/j.ijmedinf.2022.104930>
  27. Guttman A. The SPOR Canadian Data Platform: Opportunity for multi-provincial research. *CMAJ* 2019;191:E1091–2. <https://doi.org/10.1503/cmaj.191040>

## Abbreviations

- AHD-WG: Algorithms and Harmonized Data Working Group
- CCDSS: Canadian Chronic Disease Surveillance System
- CNODES: Canadian Network for Observational Drug Effect Studies
- FAIR: Findable, Accessible, Interoperable, and Reusable
- HDRN: Health Data Research Network Canada
- PHAC: Public Health Agency of Canada
- RECORD: REporting of studies Conducted using Observational Routinely collected health Data
- STROBE: STrengthening the Reporting of OBServational studies in Epidemiology

