



PUBLIC HEALTH
BIOINFORMATICS

CanCOGeN 



CanCOGeN VirusSeq

Comparison and Analysis of Canadian Public Health
SARS-CoV-2 Case Report Forms

Prepared By:

Rhiannon Cameron,
Sarah Savić Kallesøe,
and **Emma Griffiths**

Reviewed By:

William Hsiao

December 10th, 2020

CanCOGeN

Canadian Case Report Forms Analysis

December 10th, 2020

Rhiannon Cameron, Sarah Savić Kallesøe, and Emma Griffiths

CanCOGeN Metadata Harmonization Team

Executive Summary

A crucial element of Canada's response to COVID-19 is acquiring patient data, such as symptoms, demographic profile, suspected exposure, pre-existing conditions, and health outcomes. This information is required to construct a deeper understanding of the spread of SARS-CoV-2 and the efficacy of public health interventions. Patient data is acquired by means of **case report forms**, also known as **case collection forms**, initiated upon intake of a patient with COVID-19.

In Canada, there are at least eight principal variations of the case report form used at the provincial public health authority level. Seven provinces and territories utilize the Interim National Case Report Form developed by the Public Health Agency of Canada, and the remaining six provinces and territory have developed their own form or use multiple forms (see Table 1). The variety of forms used for the same purpose has created challenges and gaps in data harmonization across regions. The forms often ask different questions and implement different data structures, rendering data comparison and integration more laborious and prone to error.

The purpose of this report is to compare the common data elements of the various case report forms used across Canada in order to facilitate enhanced contextual data discussions within the Canadian COVID-19 Genomics Network (CanCOGEN). Identifying the common data types being collected will better inform researchers and health officials, enabling them to design surveillance and research questions that could be answered by analyses of available data. This report also investigates the specific data harmonization challenges currently faced in Canada, which have emerged from provinces and territories using different data collection instruments.

The implementation of data harmonization tools designed to standardize data formatting and data entry, such as CanCOGeN's **DataHarmonizer**, can help mediate many of these issues. Instead of requiring all provinces and territories to use the same case report, **the DataHarmonizer is a tool to help regions standardize the format of their data according to universal guidelines, no matter what case report form they use**. The DataHarmonizer is an open-source spreadsheet-based text editor application, designed to help data curators structure

their COVID-19 contextual patient data to international data standards^[1], and is freely available at github.com/Public-Health-Bioinformatics/DataHarmonizer.

Key findings of this report include:

Common field categories collected across all Canadian case report forms:

- Health status and outcome information;
- Signs and symptoms;
- Pre-existing conditions and risk factors;
- Complications and/or clinical evaluations;
- and exposure information (location, travel history, event, direct/indirect exposure, exposure setting and the role of the host in the exposure setting).

While these data field categories are broadly captured across all forms, they vary significantly at the field and value level, resulting in the following **data harmonization challenges** between the differing forms:

- **Variations in the categorization of information** (i.e. values matched to different fields and categories e.g. “cardiac disease” can be found under: pre-existing conditions, risk factors, assessment details, underlying conditions and comorbidity, etc.);
- **Variations in data structures and formats** (e.g. dates, gender values);
- **Variations in the granularity of information** (e.g. cough as compared to dry cough, productive cough, or new onset/exacerbation of chronic cough);
- **Variations in data types and questions being asked** (e.g. indigenous data).

Field by field comparisons across case report forms were difficult due to the wide range of variability in data structures, questions asked, information organization, terminology, and differences in the meanings of data fields. These variabilities contribute to the aforementioned challenges. In light of these challenges, **we recommend that agencies submitting data to the CanCOGeN national genomic surveillance database use the DataHarmonizer to standardize their contextual data prior to submission.** Harmonizing submissions to a standardized format will resolve inconsistencies, better enabling data integration for surveillance at international, national, and inter-regional scales.

Table of Contents

Executive Summary	2
Introduction	5
Case Report Form Analyses	7
Most Common Fields/Field Categories	7
General Case/Host Information	8
Host Subject ID	8
Gender	10
Host Health State & Host Health Outcome	10
Host Health Status Details	12
Host Origin Geo_loc Name (Country)	13
Clinical Diagnoses	14
Symptom Onset Date	15
Signs and Symptoms	16
Pre-Existing Conditions and Risk Factors	18
Complications	20
Exposure	21
Location Of Exposure Geo_loc Name (Country)	22
Destination Of Most Recent Travel (City)	23
Destination Of Most Recent Travel (State/Province/Territory)	23
Destination Of Most Recent Travel (Country)	23
Most Recent Travel Departure Date	23
Most Recent Travel Return Date	24
Travel History	24
Exposure Event	24
Direct/Indirect Exposure	25
Host Role	27
Exposure Setting	28
Case Form Data Not Currently Utilized by CanCOGeN VirusSeq	30
Indigenous Identification Data	30
Indigenous Reserve/Community	31
Data Harmonization Challenges	32
Variation in the Categorization of Information	32
Inconsistent Data Structures and Formats	33
Variations in Granularity of Information	34
Variations in Data Types and Questions Being Asked	35
Conclusion	35
References	36
Appendix	39

Introduction

This report summarizes the most common data elements being collected across the Canadian Public Health System via COVID-19 case collection forms, and highlights the data harmonization and integration challenges which arise due to the implementation of distinct data collection instruments across different jurisdictions.

A case collection form is a questionnaire used to capture epidemiological data about an ill individual, usually for surveillance and public health investigations. The case report forms reviewed in this report are the Canadian provincial and territorial forms that target confirmed or possible SARS-CoV-2 infected individuals (Table 1). The case report forms were downloaded between 2020-03-03 to 2020-04-28 and several curators independently mapped the common fields of the report forms. A final review process was carried out to verify the collection of fields presented. Manual field by field comparisons across case report forms were difficult due to the wide variety of data structures, questions asked, differences in the way information was organized, different terminology, and meanings of terms.

Table 1. Canadian provinces/territories and their associated COVID-19 case report forms.

Province/Territory	Form	Version #	Version Date (YYYY-MM-DD)
Alberta (AB) Newfoundland and Labrador (NF) Nova Scotia (NS) Nunavut (NU) Prince Edward Island (PEI) Saskatchewan (SK) Yukon (YK)	Interim National - PHSA Coronavirus Disease (COVID-19) Case Report Form ^[2]	2	2020-03-03
British Columbia (BC)	BC COVID-19 Case Report Form ^[3]		2020-04-20
Ontario (ON)	ON's Severe Acute Respiratory Infection Case Report Form ^[4]	7.0	2020-04-15
Québec (QC)	QC Coronavirus COVID-19 Déclaration Des Cas Confirmés Et Des Cas Cliniques De Covid-19 ^[5]		2020-04-28
	QC Coronavirus COVID-19 Questionnaire D'enquête Des Cas ^[6]		2020-04-02
Manitoba (MB)	MB Coronavirus Disease 2019 (COVID-19) Investigation Case Form ^[7]		2020-05-05
Northwest Territories (NWT)	NWT COVID-19 Report Form (Suspect Case/Person Under Investigation) - Part A ^[8]		2020-04-27

	NWT COVID-19 Report Form (For All Cases) - Part B ^[9]		2020-04-27
New Brunswick (NB)	NB COVID-19 Combined Referral and Lab Requisition Form ^[10]	5	2020-04-09

The Public Health Agency of Canada's Interim National Case Report Form^[2] was created to facilitate case reporting of confirmed and probable COVID-19 cases and for the identification of COVID-19 outbreaks^[11]. Seven out of thirteen provinces and territories (Table 1) were utilizing this case report form at the time of this analysis, and thus, when the Interim Case Report Form is referenced, it is given significantly greater weight owing to its wider usage. Some provinces and territories used multiple forms (e.g. NWT and QC); when these forms are referenced, it means a value was found on one or both of these forms, or on neither of them. These forms were downloaded during the first few months of the pandemic, and thus may not reflect changes made after June 2020.

Case Report Form Analyses

For the purposes of this report, fields will be labeled and defined as they are in the COVID-19 [DataHarmonizer](#)^[12], an application created by the Hsiao Public Health Bioinformatics Laboratory.

Most Common Fields/Field Categories

This section of the report outlines the most common data elements used across Canadian case report forms (**Table 2**).

Table 2. Most common fields/field categories used in all case report forms across Canada at-a-glance. The following table provides a high level overview of the data fields and field categories commonly found in the Canadian case report forms reviewed for this analysis. “Person Reporting” are data elements associated with the individual recording information on the form, while “Case information” includes data elements associated with the host/patient being observed/diagnosed/tested.

Case Report Form	Person Reporting	Case Information						
	Name	Name	DOB*	Phone Number	Gender	Symptom Onset Date*	Symptoms	Pre-existing Conditions and Risk Factors
Interim (AB, NS, PEI, NF, YK, NU, SK)								
BC								
MB								
NB								
NWT								
ON								
QC								

* Inconsistent date formats.

General Case/Host Information

This section of the report reviews data fields categories pertaining to “**General Case**”/”**Host Information**”; such as “**Host Subject ID**”, “**Gender**”, “**Host Health State**”, “**Host Health Outcome**”, “**Host Health Status Details**”, and “**Host Resident Information**”. This information is used to facilitate the linkage of VirusSeq data with other datasets (e.g. HostSeq), study relationships between host demographic information and disease outcomes, as well as aid in the public health surveillance of COVID-19. More detailed analyses of these fields can be found in the sections below and an overview of common “**General Case Information**” fields can be found in **Table 3**.

Table 3. Common General Case Information fields across Canadian case report forms.

The following table provides an overview of the “General Case Information” data fields commonly found in the Canadian case report forms reviewed for this analysis.










General Case Information	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Patient, Case, and Other Identifiers							
Personal Health Number	✗	✓	✓	✓	✓	✗	✓
Case and/or Other Identifiers	✓	✓	✓	✗	✗	✓	✓
Gender Field Values							
Female, Male	✓	✓	✓	✓	✓	✓	✓
Unknown	✓	✓	✓	✗	✗	✓	✗
Host Health State / Outcome							
Symptomatic, Deceased	✓	✓	✓	✓	✓	✓	✓
Asymptomatic	✓	✓	✓	✗	✓	✓	✓
Host Health Status Details							
Hospitalized	✓	✓	✓	✓	✓	✓	✓
ICU, ICU Start Date	✓	✓	✓	✗	✓	✓	✓
Date of Death / Disposition Date	✓	✓	✓	✗	✓	✓	✓
Host Resident Information							
City	✓	✓	✓	✓	✓	✓	✓
Address, Postal Code	✓	✓	✓	✓	✗	✓	✓

Host Subject ID

The DataHarmonizer collects each patient’s unique identifier under the field “**Host Subject ID**”. The “**Host Subject ID**” is a unique identifier that refers to the patient, usually the patient’s **Personal Health Number**.

Case report forms collect a unique identifier for each patient for the purposes of linking information. Different forms use different ways of assigning unique identifiers to patients. In the context of CanCOGeN, this information is important for facilitating linkage of HostSeq and VirusSeq data. Unique patient identifiers are documented to help link patient information collected in hospitals with patient information collected through diagnostic testing. Some case report forms use **Personal Health Numbers**, such as the forms used in BC, MB, NB, NWT, and QC. The Interim National Case Report Form and the forms used in QC and ON assign a unique case identifier instead of using personal health numbers. An overview of case report form identifiers can be found in **Table 4**. Case reports which do not include a unique identifier must to be followed up with on a case-by-case basis to ascertain the information needed to link patient information across hospital and diagnostic testing records.

Table 4. Patient, Case, and Other Identifiers across Canadian case report forms. The following table provides an overview of patient, case, and other identifier data fields found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Patient, Case, and Other Identifiers		
	Personal Health Number	Case Identifiers	Other Identifiers
Interim (AB, NS, PEI, NF, YK, NU, SK)		Local Case ID, P/T Case ID	
BC	Health Care Number (Personal Health Care Number)		Panorama Investigation ID*, PARIS Client ID**
MB	Health Number (Personal Health Identification Number)	Case Accession Number	Additional Accession Numbers, Investigation ID***
NB	Health Care Number (Medicare)		
NWT	Health Care Plan Number		
ON		iPHIS* Case ID	P/T Outbreak ID, Lab ID
QC	Health Insurance Card Number (fr. Numéro D'assurance-Maladie)		V10 Code** (fr. Saisie V10), Health Region Investigation Code (fr. Code RSS d'Enquête)

* <https://panoramacst.gov.bc.ca>

** Health Care Information System

*** Investigation ID of the contact's investigation (source: MB22000-Contact-Tracing, 2020-01-13)

Gender

The “**Gender**” field records gender of the host at the time of sample collection. This field is present in all forms. However, there is significant variation in the format of collecting gender information between them (Table 4). While “**F**” and “**Female**” or “**M**” and “**Male**” are recognizable as representing the same term, it can be difficult to disambiguate “**Intersex**”, “**Undifferentiated**”, and “**X**” without additional clarification that is not included on the forms themselves (Table 4). Also, not all forms contain an “**Other**” or “**Unknown**” equivalent, as is the case for NB and NWT. See **Table 5** for an overview of “**Gender**” field values across Canadian case report forms.

Table 5. Gender Field Values across Canadian case report forms. The following table provides an overview of the “Gender” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Gender Field Values						
	Male*	Female**	Intersex	Other	Undifferentiated	Unknown	X
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✓	✗	✓	✗	✓	✗
BC	✓	✓	✗	✗	✓	✓	✗
MB	✓	✓	✓	✗	✗	✓	✗
NB	✓	✓	✗	✗	✗	✗	✗
NWT	M	F	✗	✗	✗	✗	✓
ON	✓	✓	✗	✓	✗	✓	✗
QC	✓	✓	✗	✓	✗	✗	✗

Host Health State & Host Health Outcome

The “**Host Health State**” field records the host’s health status at the time of sample collection. “**Host Health Outcome**” records the host health status at the end of the disease; all collection forms except for NB record this information in some capacity. Examples of common host health states and outcomes are listed below and in **Table 6**. Host Health State/Outcome fields that are less common on Case Report Forms include “**Unknown**”, “**Other, specify**”, and “**Permanent disability**”.

- All except for NB are collecting “**Asymptomatic**” and “**Deceased**” (aka “**Fatal**”) host status data.
- The “**Symptomatic**” status can be inferred in all forms as a result of symptom selection and/or when “**Asymptomatic**” is not declared.
- “**Recovered**” is recorded by all except NB and NWT.
 - It’s important to note that BC doesn’t have “**Recovered**” as is, but rather two variants: “**Fully Recovered**” and “**Not yet recovered/recovering**”.
- “**Deteriorating**” and “**Stable**” are recorded by all except BC, MB, and NB.
- Less common data element include:
 - **Unknown** - BC, MB
 - **Other, specify** - BC
 - **Permanent disability** - BC

Table 6. Common Host Health State/Outcome data fields across Canadian case report forms. The following table provides an overview of the “Host Health State” and “Host Health Outcome” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Host Health State / Outcome					
	Asymptomatic	Symptomatic	Deceased	Deteriorating	Recovered	Stable
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✓*	✓	✓	✓	✓
BC	✓	✓*	✓	✗	✓	✗
MB	✓	✓	✓	✗	✓	✗
NB	✗	✓**	✗	✗	✗	✗
NWT	✓	✓*	✓	✓	✗	✓
ON	✓	✓*	✓	✓	✓	✓
QC	✓	✓*	✓	✓	✓	✓

* Can be inferred as a consequence of symptom selection and/or when “Asymptomatic” is not declared.

** Only explicit in relation to health care professionals, vulnerable populations, and/or staff working in a healthcare or other setting with vulnerable populations.

Host Health Status Details

The DataHarmonizer “**Host Health Status Details**” field is used to record further details regarding the health or disease status of the host known at the time of data collection. This information can be useful for inferring the severity of a case or determining what treatment measures may have had a correlation with specific “**Host Health Outcomes**”. All provinces will indicate whether the patient was hospitalized, often including details like whether the patient was in the ICU (including start and end dates), received mechanical ventilation, and whether the patient was isolated. Details are listed below and in **Table 7**. “**Isolation**” status information explicitly is recorded by all except BC and NB; however, forms varied on whether they specified isolation in hospital, in a facility, at home, or self-isolation at some other location. “**Negative-pressure**” was often given as an example but was not strictly declared.

- “**Hospitalized**” is recorded by all provinces/territories; however, in NWT it is implied based on patient setting selection rather than explicitly stated.
- “**ICU**” and “**ICU Start Date**” are recorded by all except for NB.
- “**ICU End Date**” is recorded by all except NWT and NB.
- “**Mechanical Ventilation**” (MV) is recorded by all except NWT and NB.
 - **MV Start and End Dates** are recorded by all except BC, NB, NWT, and ON.
- “**Isolation**” related status information is explicitly recorded by all except BC and NB;
 - “**Negative-pressure**” was often given as an example.
- If a patient was “**Deceased**”, all applicable* forms record the “**Date of Death**”/“**Disposition Date**” and all applicable* forms except for BC, MB, and NWT explicitly record whether the **death was attributed to a respiratory illness**.

**Note: NB does not record “Deceased” nor associated data.*

Table 7. Common Host Health Status Details across Canadian case report forms. The following table provides an overview of the “Host Health Status Details” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Host Health Status Details	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Hospitalized	✓	✓	✓	✓	✓	✓	✓
ICU	✓	✓	✓	✗	✓	✓	✓
ICU Start Date	✓	✓	✓	✗	✓	✓	✓
ICU End Date	✓	✓	✓	✗	✗	✓	✓
Mechanical Ventilation (MV)	✓	✓	✓	✗	✗	✓	✓
MV Start Date	✓	✗	✓	✗	✗	✗	✓
MV End Date	✓	✗	✓	✗	✗	✗	✓
Isolation	✓	✗	✓	✗	✓	✓	✓
Date of Death / Disposition Date	✓	✓	✓	✗	✓	✓	✓
Death Attributed to a Respiratory Illness	✓	✗	✗	✗	✗	✓	✓

* Implied based on patient setting selection.

Host Origin Geo_loc Name (Country)

The DataHarmonizer field “**Host Origin Geo_loc Name (Country)**” refers to the host organism’s, in this case the patient’s, **country of residence**. This data supports public health surveillance for COVID-19 as this information, combined with other exposure data fields, informs whether the individual was infected out-of-country or within their country of residence.

There appears to be an assumption on most case report forms that all patients are residents of Canada. This poses an issue when collecting information from patients who have travelled to Canada and are not residents. Only BC and NWT ask for the patient’s country of residence to be specified. Other provinces/territories only ask for province/territory information, with the exception of Québec which does not ask for province/territory information; however, Québec does request postal code information which can be used to determine the patient’s residence province/territory within Canada. An overview of Host Resident Information fields across Canadian case report forms are listed in **Table 8**.

Table 8. Host Resident Information across Canadian case report forms. The following table provides an overview of the host’s residential information fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Host Resident Information				
	Country	Province/Territory	City	Address	Postal Code
Interim (AB, NS, PEI, NF, YK, NU, SK)					
BC					
MB					
NB					
NWT					
ON					
QC					

* Asks for “Community”.

Clinical Diagnoses

This section of the report reviews “**Clinical Diagnoses**” related data fields and categories. The “**Signs and Symptoms**”, “**Pre-existing Conditions and Risk Factors**”, and “**Complication**” categories are collected to help determine the manifestations of the disease, while “**Symptom Onset Date**” is used to for public health interventions and epidemiological inferences. More detailed analyses of these fields can be found in the sections below and an overview of common “**Clinical Diagnoses**” can be found in **Table 8**.

Table 8. Common Clinical Diagnoses fields across Canadian case report forms. The following table provides an overview of the “Clinical Diagnoses” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Clinical Diagnoses	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Symptom Onset Date*	✓	✓	✓	✓	✓	✓	✓
Signs and Symptoms							
Cough	✓	✓	✓	✓	✓	✓	✓
Fever*	✓	✓	✓	✓	✓	✓	✓
Headache	✓	✓	✓	✓	✓	✓	✓
Sore Throat	✓	✓	✓	✓	✓	✓	✓
Pre-Existing Conditions and Risk Factors							
Cardiac Disease	✓	✓	✓	✓	✓	✓	✓
Diabetes	✓	✓	✓	✓	✓	✓	✓
Pregnancy	✓	✓	✓	✓	✓	✓	✓
Respiratory Disease	✓	✓	✓	✓	✓	✓	✓
Complications							
Altered Mental Status	✓	✓	✓	✓	✓	✓	✓
Encephalitis	✓	✓	✓	✓	✓	✓	✓

* Significant variation in date format, recommended for this data field, across case report forms.

** Temperature that defines a fever has some variation between forms or is not defined.

Symptom Onset Date

All case report forms collect “**Symptom Onset Date**”, which is the date signs and symptoms are believed to have begun or were first noted (**Table 9**). However, there is variation in date formatting, the significance of which is outlined in the **Inconsistent Date Formats** section. These dates are important for directing public health interventions and are used to determine a variety of epidemiological inferences (e.g. timeframe for communicability period).

Table 9. Symptom Onset Date formats across Canadian case report forms. The following table provides an overview of the “Symptom Onset Date” formats found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Symptom Onset Date				
	DD/MM/YYYY	MM/DD/YYYY	YYYY/MM/DD	YYYY/MM/DD	Unspecified
Interim (AB, NS, PEI, NF, YK, NU, SK)	✗	mm/dd/yyyy	✗	✗	✗
BC	✗	✗	YYYY/MM/DD	✗	✗
MB	✗	✗	YYYY-MM-DD	✗	✗
NB	✗	✗	✗	✗	_____
NWT	✗	✗	✗	YYYY/MM/DD	✗
ON	dd/mm/yyyy	✗	✗	✗	✗
QC	✗	✗	YYYY-MM-DD	✗	✗

Signs and Symptoms

“**Signs and Symptoms**” are changes in function or sensation, experienced by the patient, that are indicative of disease. Signs are observed by a clinician while symptoms are subjective changes apparent to the patient. However, it is important to note that case report forms do not differentiate between clinically apparent signs and patient-apparent symptoms terms and the term “symptoms” may be used to describe both signs and symptoms. A “**Signs and Symptoms**” (or some variant label) section appears on all case report forms, however, there is variation between which signs and symptoms are listed. For the most common signs and symptoms listed on forms, refer to **Table 10**.

- “**Cough**”, “**Fever**”, “**Headache**”, and “**Sore Throat**” appear in all forms.
 - However, there can be significant variation in what defines a “**Fever**”. Some forms consider a fever to be greater than 38°C, while another may consider 38°C inclusively, other forms may not even specify a temperature or request the highest temperature recorded. There is an added layer of complication when/if this data is compared with other international data as a fever defined in another country, e.g. 37.5°C or higher in China^[13], may not meet the same temperature threshold.
- “**Diarrhea**”, “**Runny Nose**”, “**Dyspnea**”, and “**Vomiting**” appear in all forms except NB.

- However, diarrhea and vomiting are not always exclusive and undifferentiated, they can be listed as “**Diarrhea/vomiting**” or “**Nausea/vomiting**”. This causes issues for patients experiencing vomiting, as vomiting is included in two categories and not exclusively in one category.
- “**Other**” field, that allows the specification of an unlisted sign/symptom, is included in all forms except NB.
- Examples found uniquely under “**Signs and Symptoms**” on only one form:
 - **Anorexia/Decreased Appetite** - ON
 - **Coryza** (*nasal mucus membrane inflammation*) - NB
 - **Dizziness** - ON
 - **Nasal Congestion** - ON
 - **Nose Bleed** - ON
 - **Otitis** (*inflammation or infection of the middle ear^[14]*) - ON
 - **Rash** - ON
 - **Renal Failure** (*kidney failure*) - MB
 - **Septicemia or Sepsis*** (*systemic inflammatory response to infection^[14]*) - MB
 - **Sneezing** - ON
 - **Sputum Production** - ON
 - **Swollen Lymph Nodes** - ON

* Appears under at least one other form field category (e.g. Pre-Existing Conditions, Assessment Details, Underlying Conditions and Comorbidities, Clinical Evaluations, etc.).

Table 10. Common Signs and Symptoms across Canadian case report forms. The following table provides an overview of the “Signs and Symptoms” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Common Signs & Symptoms							
	Cough	Diarrhea*	Fever**	Headache	Runny Nose	Dyspnea	Sore Throat	Vomiting*
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✓	✓	✓	✓	✓	✓	✓
BC	✓	✓	✓	✓	✓	✓	✓	✓
MB	✓	✓	✓	✓	✓	✓	✓	✓
NB	✓	✗	✓	✓	✗	✗	✓	✗
NWT	✓	✓	✓	✓	✓	✓	✓	✓
ON	✓	✓	✓	✓	✓	✓	✓	✓
QC	✓	✓	✓	✓	✓	✓	✓	✓

* Sometimes combined with other signs or symptoms; e.g. Diarrhea/vomiting, Nausea/vomiting.

** Temperature limit that defines a fever has some variation between forms or is not defined.

Pre-Existing Conditions and Risk Factors

“**Pre-Existing Conditions**” are medical conditions that the patient exhibited prior to infection and “**Risk Factors**” are variables associated with an increased risk of disease or infection. “**Risk Factors**” can be internal (e.g. “Pre-Existing Conditions” such as “Diabetes”), external (e.g. “Exposures” to infected individuals or via travel, and “Socioeconomic Status” such as “Homelessness”), or even a combination of both (e.g. “Behavioural” risk such as “Smoking”). “**Risk Factors**” encompass both “**Pre-Existing Conditions**” and “**Exposures**”, and thus case report forms vary in how they implement the term, with some forms using it synonymously with “**Pre-Existing Conditions**” while others only use it to describe “**Exposures**”. However, this can be problematic as risk assessments are a critical public health metric and these differences can confound analyses of risk.

While MB lists patient information such as “**health care worker**”, “**experiencing homelessness**”, or “**lab worker**” under a “**Social Risk Factors**” category, the majority of forms list these types of variables under “**Exposures**”/“**Exposure Risks**” and thus they will be found under the Exposure related fields to follow within this report. For the most common pre-existing conditions and risk factors refer to **Table 11**.

- **“Cardiac Disease”, “Diabetes”, “Pregnancy”, and “Respiratory Disease”** occur in all forms.
- **“Chronic neurological or neuromuscular disorder”, “Immunodeficiency”, “Liver Disease”, “Renal disease”** occur in all forms except NB.
- **“Post-partum (≤6 weeks)” and “Trimester”/“Week Gestation”** occur in all forms except for MB and NB. Related Fields:
 - **Specify EDC** (Estimated Date of Conception) - MB
 - **Pregnancy Complications and unfavorable issues** - QC
- Selection of examples that appear uniquely on only one to two forms:
 - **Age 60+** - NB
 - Note: this could be inferred from “Date of Birth” on all forms.
 - **Alcohol abuse** - ON
 - **Asthma** - ON
 - **Convulsions** - QC
 - **Epilepsy** - ON
 - **Experiencing Homelessness/Homeless** - NWT, QC
 - **Hemoglobinopathy/Anemia** - ON
 - **Injection drug use** - ON
 - **Neonatal Complications** - QC
 - **Pregnancy Complications and unfavorable issues** - QC
 - **Smoking/Smoker current** - MB, ON
 - **Specify EDC** (Estimated Date of Conception) - MB
 - **Tuberculosis** - ON
 - **Vaping** - MB

Table 11. Common Pre-Existing Conditions across Canadian case report forms. The following table provides an overview of the “Pre-Existing Conditions” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Common Pre-Existing Conditions								
	Cardiac Disease	Chronic Neurological or Neuromuscular Disorder	Diabetes	Immuno-deficiency	Liver Disease	Post-partum (≤6 weeks)	Pregnancy	Renal Disease	Respiratory Disease
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✓	✓	✓	✓	✓	✓	✓	✓
BC	✓	✓	✓	✓	✓	✓	✓	✓	✓
MB	✓	✓	✓	✓	✓	✗	✓	✓	✓
NB	✓	✗	✓	✗	✗	✗	✓	✗	✓
NWT	✓	✓	✓	✓	✓	✓	✓	✓	✓
ON	✓	✓	✓	✓	✓	✓	✓	✓	✓
QC	✓	✓	✓	✓	✓	✓	✓	✓	✓

Complications

“**Complications**” are clinically evaluated and diagnosed medical conditions that are believed to have occurred in a patient as a result of the host disease. Often used synonymously or in combination with “**Clinical Evaluations**” and “**Diagnoses**” in case report forms. This information is important for studying the prevalence and incidence rate of chronic and acute disease complications associated with the disease in question. For the most common complications, refer to **Table 12**. Uncommon fields include but are not limited to: “**Arrhythmia**”, “**Fatigue**”, “**Malaise**”, and “**Meningismus/nuchal rigidity**”.

- “**Altered mental status**”, “**Diagnosed with acute respiratory distress syndrome**”, “**encephalitis**”, “**respiratory disease**”, “**renal failure**”, and “**sepsis**” occur on all forms except NB.
- “**Clinical or radiological evidence of pneumonia**” occurs on all forms except BC and NB.
- “**O2 saturation <95%**”, “**Hypotension**”, “**Tachypnea (accelerated respiratory rate)**”, and “**Other (specify)**” occur on all forms except for MB and NB.
- “**Abnormal lung auscultation**” occurs on all forms except ON, MB, and NB; however, it may appear under “Signs & Symptoms”, “Pre-existing Conditions & Risk Factors”,

“Clinical Evaluations”, “Complications”, or some variant of the aforementioned categories.

- Selection of examples that appear under “Complications”, “Clinical Evaluations” and/or “Diagnoses” on only one form:
 - **Arrhythmia** - ON
 - **Fatigue** - NWT
 - **Malaise** - NWT
 - **Meningismus/nuchal rigidity** - ON

Table 12. Common Complications or Clinical Evaluations across Canadian Case Report Forms. The following table provides an overview of the “Complications” (and synonymous “Clinical Evaluations”/“Diagnoses”) data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Common Complications / Clinical Evaluations				
	Altered Mental Status	Diagnosed with Acute Respiratory Distress Syndrome	Encephalitis	Renal Failure	Sepsis
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✓	✓	✓	✓
BC	✓	✓	✓	✓	✓
MB	✓	✓	✓	✓	✓
NB	✓	✗	✓	✗	✗
NWT	✓	✓	✓	✓	✓
ON	✓	✓	✓	✓	✓
QC	✓	✓	✓	✓	✓

Exposure

“**Exposure**” data is information pertaining to where and how the patient is believed to have been exposed to the COVID-19 causative infectious agent, SARS-CoV-2. This information is used to study the relationship between exposure to a disease and the host’s health/disease state and outcomes, as well as exposure monitoring and surveillance. “Country of Exposure” and travel

related exposure information, organized according to PHA4GE standards^[1], can be found below and summarized in an overview of “**Location of Exposure**” fields in **Table 13**.

Table 13. Location of Exposure fields across Canadian case report forms. The following table provides an overview of the “Location of Exposure” related data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Location of Exposure						
	Country	Most Recent Travel					Travel History
		Destination (City)	Destination (State/Province/Territory)	Destination (Country)	Departure Date	Return Date	
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✓	Inferred*	✓	✓	✓	✓ **
BC	Inferred*	Inferred*	Inferred*	✓	✗	✓	✓ **
MB	✓	✓	Inferred*	✓	✓	✓	✓
NB	Inferred*	Inferred*	Inferred*	✓	✗	✓	✓ **
NWT	Inferred*	Inferred*	Inferred*	✓	✓	✓	✓
ON	Inferred*	✓	Inferred*	✓	✓	✓	✓ **
QC	Inferred*	✓	Inferred*	✓	✓	✓	✓ **

* May be possible to infer from other form fields (e.g. if a patient hasn’t travelled it can be assumed exposure occurred within Canada, or Travel details).

** Within the 14 days prior to symptom onset.

Location Of Exposure Geo_loc Name (Country)

Country of exposure information is referred to in the DataHarmonizer as “**Location Of Exposure Geo_loc Name (Country)**” and contains the name of the country (geographical location) in which the host was exposed to SARS-CoV-2. This information is important to collect for national and global public health surveillance and outbreak investigations.

The interim national and MB forms explicitly ask if exposure occurred in Canada, but this information can be inferred from other forms’ fields (e.g. the patient didn’t travel in the past 14 or more days prior to illness onset) and/or provided that the respondent includes travel and exposure details.

Destination Of Most Recent Travel (City)

Recent travel information regarding the city of potential exposure is referred to in the DataHarmonizer as “**Destination Of Most Recent Travel (City)**” and contains the name of the city that was the destination of the host’s most recent travel. This information is important for public health surveillance and outbreak investigations. This information is explicitly requested in all forms except BC, NB, and NWT; however, this information could be provided on these forms under additional travel details.

Destination Of Most Recent Travel (State/Province/Territory)

Recent travel information regarding the state, province, or territory of potential exposure is referred to in the DataHarmonizer as “**Destination Of Most Recent Travel (State/Province/Territory)**” and contains the name of the state, province, or territory that was the destination of the host’s most recent travel. This information is important for public health surveillance and outbreak investigations. All forms did not explicitly request this information, though it may be included under additional travel history information. In the majority of cases it can be inferred from the country and city travel information. However, this can cause complications as city names may be duplicated within countries. For example, were one to record “Richmond” as the destination city within “Canada”, how would one distinguish between the cities named “Richmond” in BC, ON, and QC?

Destination Of Most Recent Travel (Country)

Recent travel information regarding the country of potential exposure is referred to in the DataHarmonizer as “**Destination Of Most Recent Travel (Country)**” and contains the name of the state, province, or territory that was the destination of the host’s most recent travel. It is listed in addition to “**Location Of Exposure Geo_loc Name (Country)**” as there is no guarantee that these fields will be the same. In a case where an individual was exposed in their country of residence, and travelled to another country, it would be important for contact tracing, public health surveillance, and outbreak investigations to differentiate this data. All provinces collect country information pertaining to recent travel history, often specifically within the 14 days prior to symptom onset.

Most Recent Travel Departure Date

The “**Most Recent Travel Departure Date**” DataHarmonizer field is the date of the patient’s most recent departure date from their primary residence to one or more travel destinations. This information is important for public health surveillance and outbreak investigations. All provinces except for BC and NB collect this information.

Most Recent Travel Return Date

The “**Most Recent Travel Return Date**” DataHarmonizer field is the date of the patient’s most recent return date from one or more travel destinations to their primary residence. This information is important for public health surveillance and outbreak investigations. All provinces and territories collect this information.

Travel History

The “**Most Recent Travel Return Date**” DataHarmonizer field is for collecting more granular travel location information, if known, that the patient has travelled to within the last six months. This information is important for public health surveillance and outbreak investigations. All provinces collect or provide opportunities to specify additional travel histories/information. However, only MB and NWT collect information regarding travel outside of the 14 days prior to symptom onset.

Exposure Event

The “**Exposure Event**” DataHarmonizer field is the event that is believed to have led to the patient’s exposure to SARS-CoV-2. This information is important for public health surveillance and outbreak investigations; especially mass gathering events where a large population may have been exposed. While all forms have the opportunity for this information to be recorded in some capacity, in most cases “Exposure Events” are recorded under free text “additional information/details/comments” fields which may or may not be associated with an “Exposure” category. The lack of an explicit, controlled vocabulary decreases the likelihood the recorded exposure events may be compared. For an overview of how provinces record this information, refer to **Table 14**.

- A non-event specific “**Additional Information**” related data field available on the Interim Case Report, BC, NC, and ON forms. While this field provides opportunity to record exposure related information, the recorder may be less likely to record such information in this field than an “**Additional Information**” field directly associated with other exposure data.
- Exposure associated “**Additional Information**” fields are present on the Interim, BC, MB, NWT, and QC forms.
- The only explicit exposure events are “**Conference**” on BC and QC forms, and “**Agricultural Fair or event/petting zoo**” on the ON form. However, “**Conference**” on the QC form is only applicable for cases that have travelled in the 14 days prior to symptom onset, while the ON “**Agricultural Fair or event/petting zoo**” is only applicable for cases with direct or indirect animal contact.

Table 14. Opportunities to collect “Exposure Event” data across Canadian case report forms. The following table provides an overview of the data fields where “Exposure Event” information could be collected in the Canadian case report forms reviewed for this analysis.

Exposure Event	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Additional Information (Not Event Specific)*	✓	✓	✗	✓	✗	✓	✗
Exposures (General)							
Comments*	✓	✓	✗	✗	✗	✗	✗
Exposure History							
Details*	✗	✗	✗	✗	✓	✗	✓
Exposure Settings							
Acquisition Event*	✗	✗	✓	✗	✗	✗	✗
Agricultural Fair or event/petting zoo	✗	✗	✗	✗	✗	✓ **	✗
Conference	✗	✓	✓ ***	✗	✗	✗	✗
Has Opportunity to List an Exposure Event*	✓	✓	✓	✓	✓	✓	✓

* Vocabulary not controlled.

** Only in cases with direct or indirect animal contact.

*** Only completed for cases that have travelled in 14 days prior to symptom onset.

Direct/Indirect Exposure

The “**Direct/Indirect Exposure**” DataHarmonizer field indicates whether the event or setting in which the patient was exposed to SARS-CoV-2 involved direct or indirect contact, usually patient contact, with a SARS-CoV-2 infected person or organism. “**Direct Contact**” occurs via droplet spread or person-to-person contact involving skin-to-skin contact and/or fluid exchanged, and “**Indirect Contact**” being airborne, vehicle (carried via a substance, e.g. soil), and/or vector (carried via a living organism, e.g. mosquitoes) transmission. Understanding how disease transmission occurs provides opportunity for the implementation of control measures and interventions for preventing disease transmission in the future. Refer to **Table 15** for Close, Direct, and Indirect Contact Exposures across Canadian case report forms.

The Interim, NB, and ON forms explicitly specify whether the exposure involved “**Direct Contact**” with patients (for healthcare workers), while QC inquires as to whether the patient worked in direct contact with customers. ON is only one to explicitly use “**Indirect Contact**” and that was only in the context of exposure to animals/animal products. The Interim, ON, and MB (non-domestic) forms request contact information pertaining to animals such as exposure to: animal, animal waste, animal products, and location (home, work, travel, live animal market...).

All forms record “**Contact with Confirmed/Known Case**” in some capacity, usually only if contact occurred within 14 days of symptom onset, and using the broad “**Contact**” term or the

narrower “**Close Contact**” term. It is not always clear what is meant by “**Close Contact**” on a case report form, and how the term differs from “**Contact**” (i.e. synonymous with “**Direct Contact**” only or also “**Indirect Contact**”).

Definitions/clarifying information for “**Close Contact**” are provided on the Interim, BC, ON, and QC forms:

- **Interim, BC:** Close contact is defined as a person who provided care for the patient, including healthcare workers, family members or other caregivers, or who had other similar close physical contact OR who lived with or otherwise had close prolonged contact with a probable or confirmed case while the case was ill.
- **ON:** Cared for, lived with, spent significant time within enclosed quarters (e.g., co-worker), or had direct contact with respiratory secretions.
- **QC:** Within 2 meters or 6 feet. *Often paired with “Extended Contact”, meaning contact for more than 15 minutes total.*

This highlights the importance of providing definitions, or at least clarifying information, on the meaning of epidemiological terms on case report forms so that appropriate comparisons can be made across the resulting data.

Table 15. Close, Direct, and Indirect Contact Exposures across Canadian case report forms. The following table provides an overview of the “Close Contact”, “Direct Contact”, “Indirect Contact”, and “Contact” data fields commonly found in the Canadian case report forms reviewed for this analysis. See “Appendix - Table A1” for a larger view version of the table.

✓ = Present ✗ = Not Present ~ = Inferrable via Present Fields

Close, Direct, & Indirect Contact Exposures	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Contact with Confirmed/Known Case	~ **	~ **	✓ **	✓ **	✓	~ **	✓
Close Contact* with Confirmed/Known Case	✓ **	✓ **	✓ **	✗	✗	✓ **	✓ **
Contact with Probable Case	~ **	~ **	✓ **	✗	✗	~ **	✗
Close Contact* with Probable Case	✓ **	✓ **	✗	✗	✗	✓ **	✗
Close Contact*, Non-Household	✗	✗	✓	✗	✗	✗	✗
Contact with Someone with Similar Illness	~	✗	✓ **	~ **	✗	✓ **	✓ *****
Close Contact* with a person with a fever/or cough who has been to an affected area	✓ **	✗	✗	✗	✗	✗	✗
Close Contact* with a person with Acute Respiratory illness/group exposure	✗	✗	✗	✓ **	✗	✗	✗
Contact with Symptomatic People	✗	✗	✗	✗	✗	✗	✓ **
Direct Patient Contact, Healthcare Worker	✓	✗	✗	✓ ****	✗	✓	✗
Direct Patient Contact, Healthcare Volunteer	✓	✗	✗	✗	✗	✓	✗
Direct Customer Contact	✗	✗	✗	✗	✗	✗	✓
Contact with Animal or Animal Products	✓ **	✗	✓ ***	✗	✗	~ **	✗
Direct Animal Contact	✓ **	✗	✗	✗	✗	✓ **	✗
Indirect Animal Contact	✗	✗	✗	✗	✗	✓ **	✗
Direct Animal Product Contact	✓	✗	✗	✗	✗	✓ **	✗

* **Close contact** is defined as a person who provided care for the patient, including healthcare workers, family members or other caregivers, or who had other similar close physical contact OR who lived with or otherwise had close prolonged contact with a probable or confirmed case while the case was ill (Source: Interim National - PHSA Coronavirus Disease (COVID-19) Case Report Form). *Note: Cannot guarantee all Case Report Forms abide to this definition.*

** In the 14 days prior to symptom onset.

*** Does not include domestic pets.

**** Only if the case is symptomatic.

***** Frequented environments where exposure could have occurred.

Host Role

In the DataHarmonizer, the “**Host Role**” field contains the role of the host in the context of the exposure setting, or in general, for the purposes of contact tracing and transmission studies. E.g. Was the patient a student or a teacher in a school outbreak? Were they a patient, a worker, or a volunteer in a healthcare setting outbreak? There is a variety of host role information within the case report forms across Canada, a broad selection of which can be found in **Table 16**.

The level of granularity for this information can vary significantly between forms. Where the MB form asks if the host is a health care facility resident/patient or worker/volunteer, BC only asks if

the host is a healthcare worker, but then inquires further about their healthcare role (e.g. physician, nurse, housekeeping, administrative, and more).

Table 16. Collection of Common Host Roles from across Canadian case report forms. The following table provides an overview of the “Host Role” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Common Host Role Categories							
	Animal Handler	Correctional Center	Child/Day Care	Healthcare*	Laboratory	Long Term/ Personal Care Facility	School	Shelter
Interim (AB, NS, PEI, NF, YK, NU, SK)	Worker	✗	✗	Worker Volunteer	Worker	Resident	✗	✗
BC	✗	Inmate/ Resident Patient Worker	Attendee Worker	Student Worker Volunteer	Worker	Patient Resident Worker	Student Worker	Patient Resident Worker
MB	Worker	Resident	Attendee Work Volunteer	Patient Worker	Worker	Patient Resident Worker Volunteer	✗	Resident
NB	✗	Patient Resident	Worker	Patient Worker	Worker	Patient, Resident	✗	Patient Resident
NWT	✗	✗	Attendee Worker	Patient Worker	Worker	✗	Attendee Worker	✗
ON	Worker Visitor	Resident	✗	Patient Worker Volunteer	Worker	Resident	Attendee Worker	✗
QC	✗	Resident Worker	Worker	Worker	Worker	Resident Worker	Worker	✗

* Patient listed when explicit, but this information can be inferred from other form data.

Exposure Setting

The “**Exposure Setting**” DataHarmonizer field refers to the setting that led to exposure to the pathogen in question. This information is important for public health surveillance and outbreak investigations. Patient exposure setting information is collected in some capacity across all case report forms, though it may not always be under the label of “**Exposure Setting**”. Refer to **Figure 1** for an example of different case report forms approaching the same information category.

“**Healthcare Setting**”, “**Family Setting**”, and “**Workplace Setting**” information is collected on all forms except ON, NB, and NWT. All of the aforementioned forms also have an “**Other, specify**” field and some explicitly list other settings such as “**Correctional Facility**”, “**Long Term Care Facility**”, “**Household**”, etc. The most common fields include “**Workplace Setting**” (all except MB), “**Laboratory Setting**” (all except NB), and “**School/Daycare/Etc.**” (all except MB). The least common fields are “**Shelter**” (only on BC and MB), “**Conference**” (only BC), and “**Public**

Transport” (only QC). Many of these fields overlap with those seen in the **“Host Role”** section, but sometimes **“Exposure Settings”** have no specific role associated with them. For an overview of patient exposure settings fields and categories, refer to **Table 17**.

Patient Setting		If yes, setting type:*	
<input type="checkbox"/> Physician office/clinic	<input type="checkbox"/> Home visit	<input type="checkbox"/> Acute care facility	<input type="checkbox"/> Long term care facility
<input type="checkbox"/> ED (not admitted)	<input type="checkbox"/> Facility (LTC, Corrections)	<input type="checkbox"/> Group home (community living)	
<input type="checkbox"/> Inpatient (ward)	Admission date: YYYY/MMM/DD	<input type="checkbox"/> Correctional facility	<input type="checkbox"/> School or daycare
<input type="checkbox"/> Inpatient (ICU)	Admission date: YYYY/MMM/DD	<input type="checkbox"/> Assisted living	<input type="checkbox"/> Independent living
		<input type="checkbox"/> Other residential facility type, specify: _____	
		<input type="checkbox"/> Shelter	<input type="checkbox"/> Conference

Figure 1. Excerpts of **“Patient Setting”** information from **NWT (left)** and **BC (right)** case report forms.

Table 17. Exposure Settings across Canadian case report forms. The following table provides an overview of the **“Exposure Setting”** data fields commonly found in the Canadian case report forms reviewed for this analysis. See **“Appendix - Table A2”** for a larger view version of the table.

Exposure Settings	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Healthcare Setting	✓	✓	✓	✓	✓	✓	✓ *
Family Setting	✓	✗	✗	✗	✗	✗	✓
Workplace Setting	✓	✓	✗	✓	✓	✓	✓
Unknown	✓	✓	✗	✗	✗	✗	✗
Other, specify:	✓	✓	✓	✓	✓	✗	✗
Laboratory setting	✓ *	✓	✓ *	✗	✓	✓	✓
Home/Household	✓ **	✓	✗	✗	✗	✗	✗
During Travel	✓ *	✗	✗	✗	✗	✗	✗
Animal (E.g. Farm, Live Animal Market, Vet)	✓	✗	✓	✗	✗	✗	✗
Long Term Care Facility	✗	✓	✗	✗	✓	✓	✓
Other Residential Facility (E.g. Acute Care Facility, Assisted Living, Group Home, etc.)	✗	✓	✓	✗	✗	✗	✓
Correctional facility	✗	✓	✓	✗	✗	✗	✗
School/Daycare/Etc.	✗	✓	✗	✓	✓	✓	✓
Shelter	✗	✓	✓	✗	✗	✗	✗
Conference	✗	✓	✗	✗	✗	✗	✗
Public Transport	✗	✗	✗	✗	✗	✗	✓

* Inferrable from other data elements.

** Exposure in the context of **“animals”** only.

Case Form Data Not Currently Utilized by CanCOGeN VirusSeq

Indigenous Identification Data

Eleven of the 13 case report forms used across Canada collect information on whether the patient self-identifies as having Indigenous heritage (**Table 18**). Collectively, these case report forms further inquire whether the individual identifies as First Nations, First Nations Status, Métis, Inuit, or a combination of the fields, but no one form offers all these options.

The distinctive options of reporting as “First Nations” and “First Nations Status” are important, as the process of acquiring status is laborious and difficult. This results in only a portion of individuals who have First Nations heritage being legally recognized. The difference in status recognition is pertinent to acknowledge, as it provides access to health and government services not otherwise available, which may affect health outcomes. Only one case report form, MB, distinguishes between status and non-status First Nations.

Two case report forms fail to inquire about Indigenous identity at all, which poses challenges for evidenced-based public health initiatives. Indigenous Canadians carry a heavier burden of poor health compared to non-Indigenous Canadian across nearly every health outcome metric^{[15][16]}^[17]. To not collect data specific to Indigenous populations would be to ignore the health gap which exists between the Indigenous- and non-Indigenous Canadians. Currently, these data types are not represented in the DataHarmonizer due to the lack of appropriate and culturally sensitive data standards. The Metadata Harmonization team is working with the CanCOGeN Ethics and Governance work group to identify language appropriate for data capture.

Table 18. Indigenous Identification Data fields across Canadian case report forms. The following table provides an overview of Indigenous identification/heritage data fields commonly found in the Canadian case report forms reviewed for this analysis.

Case Report Form	Indigenous Identification Data					
	Identify as Indigenous	First Nations Status	First Nations	Métis	Inuit	Combination*
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✗	✓	✓	✓	✗
BC	✓	✓	✓	✓	✓	✓
MB	✓	✓	✓	✓	✓	✗
NB	✗	✗	✗	✗	✗	✗
NWT	✗	✗	✗	✗	✗	✗
ON	✓	✗	✓	✓	✓	✗
QC	✓	✗	✓	✗	✓	✗

* Explicit options for “First Nations and Inuit”, “First Nations and Métis”, “First Nations, Inuit and Métis”, or “Inuit and Métis”

Indigenous Reserve/Community

There is a significant gap in the collection of Indigenous reserve and/or community data on Canadian case report forms. It is important to clarify that while a reserve is or is a part of an Indigenous community, not all Indigenous communities are reserves - that is, have “reserve status”^[18]. Whether the patient resides on a reserve is collected on the Interim, ON, and QC forms. Whether the patient resides in an Indigenous community is collected on the NB and MB forms; however, NB only collects this information if the patient is symptomatic. The only province/territory to collect Indigenous organization information (e.g. “Nazko First Nation”) was BC. This information is important for identifying Indigenous nations and organizations in order to uphold First Nations Data Governance initiatives^{[19] [20]}.

Data Harmonization Challenges

Data harmonization challenges arise from the use of variable collection forms. Examples of harmonization challenges include:

- **Variations in the categorization of information** (i.e. values matched to different fields and categories e.g. “cardiac disease” can be found under: pre-existing conditions, risk factors, assessment details, underlying conditions and comorbidity, etc.);
- **Variations in data structures and formats** (e.g. dates, gender values);
- **Variations in the granularity of information** (e.g. cough as compared to dry cough, productive cough, or new onset/exacerbation of chronic cough);
- **Variations in data types and questions being asked** (e.g. Indigenous data).

Refer to **Figure 2** for case report form excerpts that demonstrate some of the aforementioned differences.

a

Laboratory	Specimen Collected
Specimen Collection Date: YYYY/MMM/DD	
<input type="checkbox"/> NP swab	<input type="checkbox"/> Upper respiratory (e.g., Nasopharyngeal or oropharyngeal swab)
<input type="checkbox"/> Throat swab	
<input type="checkbox"/> Sputum	<input type="checkbox"/> Lower respiratory (e.g., sputum, tracheal aspirate, BAL, pleural fluid)
<input type="checkbox"/> Other (e.g. BAL), specify: _____	

b

Patient Setting	If yes, setting type:*
<input type="checkbox"/> Physician office/clinic	<input type="checkbox"/> Acute care facility
<input type="checkbox"/> ED (not admitted)	<input type="checkbox"/> Long term care facility
<input type="checkbox"/> Inpatient (ward) Admission date: YYYY/MMM/DD	<input type="checkbox"/> Group home (community living)
<input type="checkbox"/> Inpatient (ICU) Admission date: YYYY/MMM/DD	<input type="checkbox"/> Correctional facility
<input type="checkbox"/> Home visit	<input type="checkbox"/> School or daycare
<input type="checkbox"/> Facility (LTC, Corrections)	<input type="checkbox"/> Assisted living
	<input type="checkbox"/> Independent living
	<input type="checkbox"/> Other residential facility type, specify: _____
	<input type="checkbox"/> Shelter
	<input type="checkbox"/> Conference

Figure 2. Differences in how information is collected across case report forms. (a) Excerpts of “Specimen Collection” information from NWT (left) and BC (right). In this example, the different forms use abbreviations and encode specimen information at different levels of granularity. **(b)** Excerpts of “Patient Setting” information from NWT (left) and BC (right). In this example, different questions are being asked using the same field.

Variation in the Categorization of Information

Categories of information vary between case report forms, making them difficult to curate and correlate their underlying data fields. For example, depending on the form being utilized, “Cancer” may be located under “Pre-Existing Conditions”, “Risk Factors”, “Assessment Details”,

“Underlying Conditions and Comorbidity”, or a lexical variation of one of the previous labels. Additionally, what constitutes a “Risk Factor”? Some “Risk Factor” groups are used synonymously with “Pre-Existing Conditions”, while others are used to mean “Exposure” (e.g. Exposure to infected individuals, Travel Exposure, etc.), or possibly a combination “Pre-Existing Conditions” and “Exposures”.

Inconsistent Data Structures and Formats

A **Data Structure** is a collection of data values, the relationships between them, and the ways these data values can be manipulated or operated on. Data structures are generally designed for a specific purpose, that is, designed such that the desired interpolation/results can be appropriately inferred from them. For example, to represent a date, we structure it as three values, a day, a month, and a year with a specific hierarchical arrangement. The date is formatted in such a way to make known what the data values represent (e.g. “01” represents “January”) and how they relate to one another (e.g. a date belongs in a month, which belongs in a year). This has a particular format in order to uniquely represent it and remove ambiguity.

However, **date formats** are not consistent between, and sometimes even within, collection forms (Table 2), allowing for ambiguity and misinterpretations between day, month, and even year. This is especially problematic in cases where the form specifies no date format at all. With the example of the NB form, the individual entering data from the form would not have enough information to decipher if a date such as “03/04/2020” refers to the 4th of March or the 3rd of April, 2020. While Canada will accept dates in any format, to avoid misinterpretation of numeric dates the Government of Canada recommends and has declared the national standard to be the ISO 8601 international standard YYYY-MM-DD or YYYY-MM^[21] ^[22]. However, the common misinterpretation and undefined standard of date formats on collection forms causes major problems; this is especially so during the time-sensitive crisis of COVID-19 where errors in specifying the date of data has serious implications.

There is also the issue of differences in value types, for example, where one form may offer a boolean (True/False) value, another may simply ask for a free-text description. This results in data string variations that require different levels of process and may not be easily compared. For example, one form may inquire whether a patient’s fever was “≥38°C” with a check box (True/False), while another asks if the patient has a fever and asks the recorder to specify the highest temperature recorded in degree celsius. Furthermore, there is no standard as to whether the latter is a written string of characters or a number - which adds to the curation time of this data.

Table 19. Date Formats across Canadian case report forms. The following table provides an overview of date formats found in the Canadian case report forms reviewed for this analysis.

Case Report Form	All Case Report Form Date Formats*					
	DD/MM/YYYY	MM/DD/YYYY	YYYY/MM/DD	YYYY/Mo/DD	YYYY/MMM/DD	<i>Unspecified</i>
Interim (AB, NS, PEI, NF, YK, NU, SK)	✓	✓	✗	✗	✗	✗
BC	✗	✗	✓	✗	✗	✗
MB	✗	✗	✓	✗	✗	✗
NB	✗	✗	✗	✓	✗	✓
NWT	✗	✗	✗	✗	✓	✗
ON	✓	✗	✗	✗	✗	✗
QC	✗	✗	✓	✗	✗	✗

* Ignoring variation in uppercase/lowercase and separators. Were these variations to be included almost all of the case report forms would have different formats.

Variations in Granularity of Information

Variations in granularity is a prevalent complication in comparing data across case report forms. This variation can differentiate the data points such that it limits the depth of analysis to the point where a common term is difficult to match between case report forms. For example, terms may appear as “diarrhea/vomiting”, however this prohibits detailed analysis of specific pathologies as it lacks the granularity to distinguish between when the recorder means “diarrhea” and when they mean “vomiting”, or if they mean together. For example, the term “vomiting” may appear combined as “diarrhea/vomiting” or “nausea/vomiting”; in that case is it appropriate to correlate it with a standalone “vomiting” datapoint? What if the reporter meant diarrhea? We can no longer conclusively say this data point can be quantified as “vomiting” alone. While it sometimes is the case that the recorder is asked to circle the term they intended, this was not indicated in these cases. There is also a varying degree of granularity when it comes to collecting indigenous identity information as forms vary on whether they collect whether individuals identify as “First Nations”, “Inuit”, “Métis”, or a combination of the three.

Variations in Data Types and Questions Being Asked

Asking disparate questions makes it difficult to identify converging data points, and puts a limitation on the scale of data analysis that can be performed on said data. One form asks for Indigenous ethnicity information by inquiring whether the case is in the "symptomatic members of Indigenous communities" priority group, while other forms will inquire as to whether the patient identifies as aboriginal; a form may also inquire as to whether the patient has "First Nations Status" or resides on a reserve.

There are also significant differences in the values for symptoms and pre-existing conditions. The only time a province/territory asked whether a patient had the signs/symptoms of "coryza" (acute inflammation of the mucous membrane of the nose) was in NB case report form. While QC was the only province to ask if the patient experienced pregnancy complications as a consequence of the illness and whether the patient was a worker with direct customer contact. The meaning of words can differ across forms as well. If a form asks for "Isolation" without further explanation or explicit specification, do they mean "Self-Isolation", "Home Isolation", or "Hospital Isolation"? And if they ask for "Hospital Isolation" is this meant to be synonymous with "Negative Pressure" (i.e. hospital isolation in negative-pressure rooms) or do they just mean the patients were put in a private room without negative pressure precautions (e.g. a minimum number of air exchanges per hour)?

That is not to say that these are not valuable questions to ask, rather that they are devalued as a consequence of them only being asked by a small portion of care report forms, or by being ambiguous. While it's possible that some of the institutional specific questions are due to the unique circumstance of the jurisdiction, that does not appear to be the case in most if not all the variables we identified. Some coordination across the nation could significantly reduce the inconsistencies. While there is not necessarily a strict limitation on case report form size, form "real estate" is a valuable commodity. Too many fields increases the burden of data entry on healthcare workers, and increases the likelihood that some fields are ignored or missed. Therefore, requesting a large number of data elements may result in diminishing or negative returns in the total volume and quality of data being recorded.

Conclusion

This report presents the common data elements contained in federal, provincial, and territorial COVID-19 case collection forms. Identifying the common data fields being collected better informs researchers and health officials, enabling them to design surveillance and research questions that could be answered by analyses based on available data. While it is not clear whether all fields are being filled at the point of data collection, the commonalities identified in this report are a starting point for enhanced contextual data collection moving forward.

This report also reviews specific data harmonization challenges which have emerged from the use of different collection forms. The overarching theme is that provincial/territorial case report forms significantly vary in the questions they ask and the method in which data are structured. Employing the DataHarmonizer to standardize the data collected across Canada would help

improve comparability and interoperability. This work also highlights the need for a pan-Canadian framework for minimal data sharing for genomic surveillance during health emergencies. Such a framework would prescribe and define the data elements required, and the mechanisms for harmonization. We recommend that CanCOGeN work towards such a framework as a deliverable of the initiative to enable faster, more efficient data sharing and integration in the future.

References

- [1] Griffiths, E. J., Timme, R. E., Page, A. J., Alikhan, N. F., Fornika, D., Maguire, F., ... & Tyson, G. H. (2020). The PHA4GE SARS-CoV-2 contextual data specification for open genomic epidemiology.
- [2] PHSA - Public Health Services Authority. Coronavirus Disease (COVID-19) Case Report Form. [Online] 03 March 2020. [Accessed on 12 May 2020]
<https://www.canada.ca/content/dam/phac-aspc/documents/services/diseases/2019-novel-coronavirus-infection/health-professionals/2019-nCoV-case-report-form-en.pdf>
- [3] BCCDC - British Columbia Centre for Disease Control. COVID-19 Case Report Form. BC Centre for Disease Control. [Online] 20 April 2020. [Accessed on 12 May 2020]
http://www.bccdc.ca/Documents/COVID-19_Case_Report_Form.pdf
- [4] PHO - Public Health Ontario. Appendix 1: Ontario's Severe Acute Respiratory Infection Case Report Form. *In* Public health management of cases and contacts of COVID-19 in Ontario. [Online] 15 April 2020. [Accessed on 12 May 2020]
http://www.health.gov.on.ca/en/pro/programs/publichealth/coronavirus/docs/2019_PH_Mgmt_guidance.pdf
- [5] MSSS QC - Ministère de la Santé et des Services sociaux's du Québec. Déclaration Des Cas Confirmés Et Des Cas Cliniques De Covid-19. [Online] 28 April 2020. [Accessed on 12 May 2020]
<https://publications.msss.gouv.qc.ca/msss/fichiers/2020/20-210-103W.pdf>
- [6] MSSS QC - Ministère de la Santé et des Services sociaux's du Québec. Questionnaire D'enquête d'un cas de COVID-19. [Online] 02 April 2020. [Accessed on 17 July 2020]
<https://msss.gouv.qc.ca/professionnels/documents/coronavirus-2019-ncov/COVID-19-Cas-formulaire-d-enquete-QC.pdf>
- [7] Manitoba Health, Seniors and Active Living. MHSU-6683: Coronavirus Disease 2019 (COVID-19) Case Investigation Form. [Online] 05 May 2020. [Accessed on 12 May 2020]
https://www.gov.mb.ca/health/publichealth/surveillance/docs/mhsu_6683.pdf
- [8] GNWT. COVID-19 Report Form (Suspect Case/Person Under Investigation) - Part A: To be completed for all COVID-19 Suspect Cases/PUI. [Online] 27 April 2020. [Accessed on 12 May 2020]
<https://www.hss.gov.nt.ca/professionals/en/covid-19-report-form-suspect-cases-%E2%80%93-part>
- [9] GNWT - Government of the Northwest Territories. COVID-19 Report Form (Suspect Case/Person Under Investigation) - Part B: To be completed for all COVID-19 Confirmed, Epi-linked and Probable Cases. [Online] 27 April 2020. [Accessed on 12 May 2020]
<https://www.hss.gov.nt.ca/professionals/en/covid-19-report-form-confirmed-cases-%E2%80%93-part-b>

- [10] NB VHN - New Brunswick Vitalité Health Network. COVID-19 Combined Referral and Lab Requisition Form. [Online] 09 April 2020. [Accessed on 12 May 2020]
https://www.vitalitenb.ca/sites/default/files/documents/medecins/4_combined_referral-lab_requisition_form_2020-04-09.pdf
- [11] Government of Canada. Coronavirus disease (COVID-19): For health professionals. [Online] 08 November 2020. [Accessed on 08 November 2020]
<https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals.html>
- [12] Gill I.S., Griffiths E.J., Dooley D., Cameron R., Gosal G., Sehar A., Tindale L., and Hsiao W. [in preparation] 2020. COVID DataHarmonizer: harmonizing SARS-CoV-2 contextual information with a data-secure spreadsheet application for faster genomic sequence aggregation and analysis. Application Note, Public Health Bioinformatics Laboratory. [Accessed on 08 November 2020].
- [13] Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., ... & Du, B. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine*, 382(18), 1708-1720.
- [14] HPO - Human Phenotype Ontology. [Release] October 2020. [Accessed on 08 November 2020]
<http://www.obofoundry.org/ontology/hp.html> [hp.owl] via www.ebi.ac.uk/ols/ontologies/hp
- [15] Wilson, K., & Cardwell, N. (2012). Urban Aboriginal health: Examining inequalities between Aboriginal and non-Aboriginal populations in Canada. *The Canadian Geographer/Le Géographe Canadien*, 56(1), 98-116.
- [16] Bramley, D., Hebert, P., Jackson, R., & Chassin, M. (2004). Indigenous disparities in disease-specific mortality, a cross-country comparison: New Zealand, Australia, Canada, and the United States. *The New Zealand Medical Journal (Online)*, 117(1207).
- [17] Chen, L., Xiao, L., Auger, N., Torrie, J., McHugh, N. G. L., Zoungrana, H., & Luo, Z. C. (2015). Disparities and trends in birth outcomes, perinatal and infant mortality in Aboriginal vs. non-Aboriginal populations: a population-based study in Quebec, Canada 1996–2010. *PLoS One*, 10(9), e0138562.
- [18] Mccue, H.A. Reserves. Reserves. The Canadian Encyclopedia. [Online] 31 May 2011. [Accessed on 12 November 2020]. <https://www.thecanadianencyclopedia.ca/en/article/aboriginal-reserves>
- [19] First Nations Information Governance Centre. [Online] 2020. [Accessed on 12 November 2020].
<https://fnigc.ca/>
- [20] British Columbia First Nations' Data Governance Initiative (BCFNDGI). [Online] n.d.. [Accessed on 12 November 2020]. <https://www.bcfndgi.com/>

[21] TBITS 36: All-Numeric Representation of Dates and Times – Implementation Criteria. Treasury Board of Canada. [Online] 18 December 1997. [Accessed on 08 November 2020]

<https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=17284>

[22] Standards Council of Canada. CAN/CSA-Z234.4-89 (R2007): All-Numeric Dates and Times. [Online] 31 December 1989. [Accessed on 08 November 2020]

<https://www.scc.ca/en/standardsdb/standards/4449>

Appendix

Table A1. Close, Direct, and Indirect Contact Exposures across Canadian case report forms. The following table provides an overview of the “Close Contact”, “Direct Contact”, “Indirect Contact”, and “Contact” data fields commonly found in the Canadian case report forms reviewed for this analysis.

✓ = Present ✗ = Not Present ~ = Inferrable via Present Fields

Close, Direct, & Indirect Contact Exposures	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Contact with Confirmed/Known Case	~ **	~ **	✓ **	✓ **	✓	~ **	✓
Close Contact* with Confirmed/Known Case	✓ **	✓ **	✓ **	✗	✗	✓ **	✓ **
Contact with Probable Case	~ **	~ **	✓ **	✗	✗	~ **	✗
Close Contact* with Probable Case	✓ **	✓ **	✗	✗	✗	✓ **	✗
Close Contact*, Non-Household	✗	✗	✓	✗	✗	✗	✗
Contact with Someone with Similar Illness	~	✗	✓ **	~ **	✗	✓ **	✓ *****
Close Contact* with a person with a fever/or cough who has been to an affected area	✓ **	✗	✗	✗	✗	✗	✗
Close Contact* with a person with Acute Respiratory illness/group exposure	✗	✗	✗	✓ **	✗	✗	✗
Contact with Symptomatic People	✗	✗	✗	✗	✗	✗	✓ **
Direct Patient Contact, Healthcare Worker	✓	✗	✗	✓ ****	✗	✓	✗
Direct Patient Contact, Healthcare Volunteer	✓	✗	✗	✗	✗	✓	✗
Direct Customer Contact	✗	✗	✗	✗	✗	✗	✓
Contact with Animal or Animal Products	✓ **	✗	✓ ***	✗	✗	~ **	✗
Direct Animal Contact	✓ **	✗	✗	✗	✗	✓ **	✗
Indirect Animal Contact	✗	✗	✗	✗	✗	✓ **	✗
Direct Animal Product Contact	✓	✗	✗	✗	✗	✓ **	✗

* Close contact is defined as a person who provided care for the patient, including healthcare workers, family members or other caregivers, or who had other similar close physical contact OR who lived with or otherwise had close prolonged contact with a probable or confirmed case while the case was ill (Source: Interim National - PHSA Coronavirus Disease (COVID-19) Case Report Form). Note: Cannot guarantee all Case Report Forms abide to this definition.

** In the 14 days prior to symptom onset.

*** Does not include domestic pets.

**** Only if the case is symptomatic.

***** Frequented environments where exposure could have occurred.

Table A2. Exposure Settings across Canadian case report forms. The following table provides an overview of the “Exposure Setting” data fields commonly found in the Canadian case report forms reviewed for this analysis.

Exposure Settings	Case Report Forms						
	Interim (AB, NS, PEI, NF, YK, NU, SK)	BC	MB	NB	NWT	ON	QC
Healthcare Setting	✓	✓	✓	✓	✓	✓	✓ *
Family Setting	✓	✗	✗	✗	✗	✗	✓
Workplace Setting	✓	✓	✗	✓	✓	✓	✓
Unknown	✓	✓	✗	✗	✗	✗	✗
Other, specify:	✓	✓	✓	✓	✓	✗	✗
Laboratory setting	✓ *	✓	✓ *	✗	✓	✓	✓
Home/Household	✓ **	✓	✗	✗	✗	✗	✗
During Travel	✓ *	✗	✗	✗	✗	✗	✗
Animal (E.g. Farm, Live Animal Market, Vet)	✓	✗	✓	✗	✗	✗	✗
Long Term Care Facility	✗	✓	✗	✗	✓	✓	✓
Other Residential Facility (E.g. Acute Care Facility, Assisted Living, Group Home, etc.)	✗	✓	✓	✗	✗	✗	✓
Correctional facility	✗	✓	✓	✗	✗	✗	✗
School/Daycare/Etc.	✗	✓	✗	✓	✓	✓	✓
Shelter	✗	✓	✓	✗	✗	✗	✗
Conference	✗	✓	✗	✗	✗	✗	✗
Public Transport	✗	✗	✗	✗	✗	✗	✓

* Inferred from other data elements.

** Exposure in the context of “animals” only.