Pan-Canadian Real-world Health Data Network (PRHDN) Glossary
Version 1.0
May 7th, 2015

Algorithm – A reusable process, ideally in the form of documented code, that implements a common approach or definition. Algorithms can be used for a variety of purposes, such as to define cases or to create derived variables.

Case or Case definition – Definition of characteristics that qualify individuals as having a particular condition (e.g., diabetes, hypertension) or experience of care (e.g., hospitalization).

Common analytic protocol – A series of technical and process documents detailing the steps for analyses designed to yield results that can be pooled or combined even though the studies are performed using non-harmonized data. Common analytic protocols may include simple statistical methods and/or more complex techniques such as propensity score matching.

Common data – A collection of variables that are comparable across time and/or jurisdictions through agreed upon content and terminological standards for data collection. Common data variables are a subset of harmonized common data.

Data – A collection of observations (records) on variables that are available for analysis.

Derived variable – A variable that does not exist in an original data set but can be created by combining or otherwise manipulating existing information. Examples include the number of chronic conditions, an indication of seeing a primary care provider within a set time period after hospital discharge and total cost of care.

Emergent properties – Properties that cannot be measured at a service encounter level; they emerge only by considering an individual's experience over time or by aggregating the experience of many individuals. Examples include continuity of care, accessibility and equity.

Federated data analysis – An analytic approach wherein analyses are run simultaneously on data that remain in separate locations. Federated data analysis is a way to run pooled analyses without requiring data to be in a central location

Harmonization: The process through which variables from different time periods, sources, or jurisdictions are made to be inferentially equivalent and thus can be considered common data. Harmonization can be prospective (i.e., by design) or retrospective (i.e., with rigorous methods and validation). Not all variables can be harmonized because there is a requirement for a minimal baseline level of concurrence

in the data collection.

Harmonized common data: A collection of variables that are comparable across jurisdictions, including data that are equivalent from the time that they are collected by the primary source and data that have been harmonized.

Metadata – Commonly referred to as "data about data", metadata provide information about the origin and structure of data sets and are intended to inform appropriate use of data. Metadata can include inclusion and exclusion criteria for data sets, definitions for any derived variables, and summary statistics for each field included in a data set.

Pooled data – Different data sets brought together for the purpose of analysis. This may or may not involve pooling of individual-level data.

Summary data meta-analysis – An analytic approach wherein individual analyses are performed on common data in their original locations. This is followed by pooling of non-confidential summary results and using statistical techniques to make inferences about the whole (pooled) population. Similar to meta-analysis that is often conducted in systematic reviews of the literature, but in this case it is part of the planned primary analysis of a study.

Variable – A data element that represents a measurement; for example sex, age, diagnosis. Sometimes referred to as a field or column.