



**Measuring
Health Equity
in Toronto
Central LHIN**

**A Quick Guide
to Evaluating
Data Quality**

TABLE OF CONTENTS

Contents

1. Understanding Data Quality	1
2. Why Assess Data Quality?	2
3. Data Quality Issues	3
4. Data Quality Indicators.....	4
4.1. INDICATOR: Participation Rate	4
4.2. INDICATOR: Item Response Rate	5
4.3. INDICATOR: Feedback	6
5. Ensuring High Quality Data	7
5.1. Planning Stages.....	7
5.2. Post- data collection.....	7
5.3. Additional Training.....	8
Appendix A: Data Collectors Focus Group Questions	9
Appendix B: Data Collectors Survey Questions	10
Appendix C: Questions on Data Collection Procedures	11

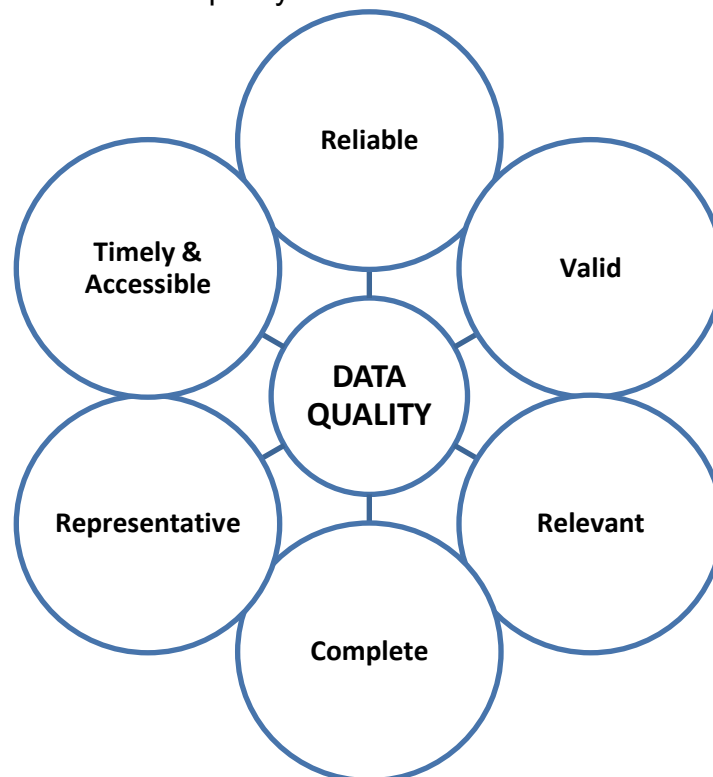
1. Understanding Data Quality

Broadly defined, data quality is “the totality of features and characteristics of a data set that bear on its ability to satisfy the needs that result from the intended use of the data”¹. In other words, data quality refers to the extent to which data allows users to meet data collection goals, answer questions, and draw conclusions. Just because a data set exists, doesn’t mean that it can be used. Poor data collection methods, poorly worded questions, weakly constructed response options, and non-interpretable data (e.g. difficulty in understanding patient/client responses under “Other”) are all common issues that render data sets unusable.

Applied to the “Measuring Health Equity in the TC LHIN” initiative, assessing data quality is about assessing whether the data can:

- Tell us who we serve
- Identify patient and client needs
- Identify existing health inequities
- Provide basis for developing services and health inequity interventions

There are a number of data quality dimensions to aim for:



¹ Arts, D. G. T., De Keiser, N. F., & Scheffer, G. (2002). Defining and improving data quality in medical registries. *Journal of the American Medical Informatics Association*, 9, 600-611.

A number of questions can help assess the quality of the data on multiple dimensions:

Dimension	Application to Measuring Health Equity in TC LHIN
Reliability	Do questions provide accurate and meaningful patient data?
Validity	Does the item measure what it's supposed to measure?
Completeness	What is the missing data (i.e. incomplete) rate?
Relevance	Can the data be used in way that meets goals for its collection?
Representativeness	Does the data reflect the actual patient population?
Accessibility	Is the data is available for use by those who want to advance equitable and quality healthcare?



Completeness is a crucial data quality dimension and tracing the source of missing data continues to be a major barrier facing the collection of high quality patient demographic data.

2. Why Assess Data Quality?



- Data quality impacts usefulness of the data
- Data quality strengthens (or weakens) credibility
- Bad (poor quality) data leads to bad decisions

Bad data is costly in terms of time, money, resources:

	Time for training data collectors, collecting data, data entry, meetings
	Cost of materials, system upgrades, allocated hours for related activities

3. Data Quality Issues

Several issues can impact data quality during the collection of individual-level demographic data. The following table lists a series of data issues and identifies potential sources for those issues:

Data issue	Data Quality Dimension Affected	Potential source	
		Data Collection	Data Entry/Storage
Low rate of participation among patients/clients	<ul style="list-style-type: none"> Reliability Completeness Representativeness 	<ul style="list-style-type: none"> Patient is not asked the questions Patient doesn't want to participate 	Data not entered
Missing data, no response selected	<ul style="list-style-type: none"> Reliability Completeness Relevance 	<ul style="list-style-type: none"> Patient is not asked the question Patient does not understand the item Patient doesn't know the answer Patient prefers not to pick any of the answer options 	Data not entered
Answer/response fails to capture information about the patient	<ul style="list-style-type: none"> Reliability Validity 	<ul style="list-style-type: none"> Question is unclear Incorrect translation High rates of 'do not know' or 'prefer not to answer' 	Data incorrectly entered
Failure to reach all patients/clients in the target population	<ul style="list-style-type: none"> Reliability Representativeness 	<ul style="list-style-type: none"> Biased recruitment of patients/clients 	N/A
Information cannot be accessed for intended use	<ul style="list-style-type: none"> Accessibility 	<ul style="list-style-type: none"> Data is not collected at the right time Inability to link information to outcomes 	Data stored in inaccessible space



Poor data quality can be driven by multiple factors, making it essential that organizations build in the capacity to trace the source of poor data. E.g. develop a process to track whether staff are asking patients/clients, which can help identify whether that is a source of missing data.

4. Data Quality Indicators

A number of indicators can be used to examine data quality. Listed below are three indicators that are commonly captured and used to assess data quality:

4.1. INDICATOR: Participation Rate

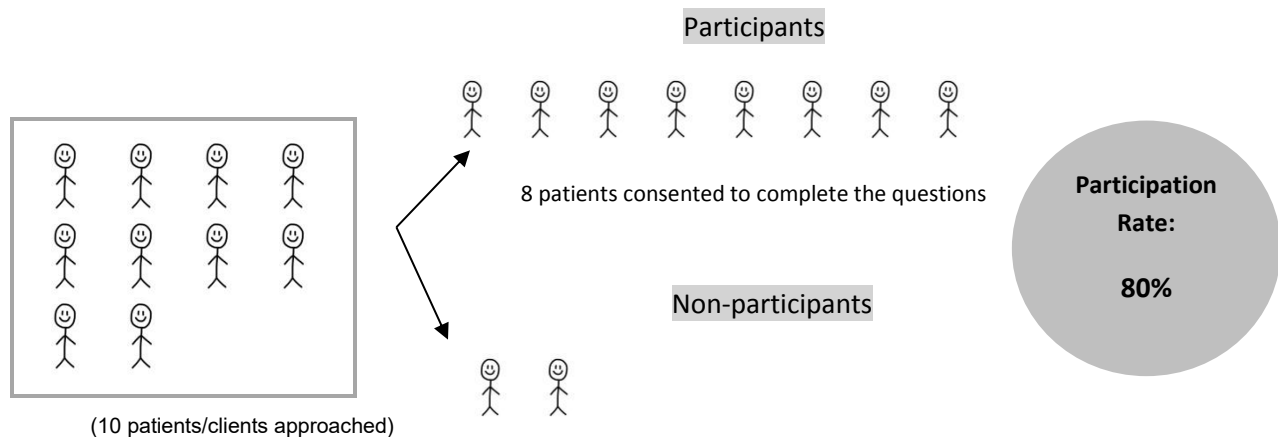
Purpose

Used as an estimate of the representativeness of the data - i.e., can we say that the information captured reflects an accurate picture of the patient population? If the majority of patients/clients have participated (high participation rate), then confidence in using the data to develop and implement programs/policies for all patients/clients increases

Definition

Participation rate refers to the percentage of participants (patients/clients) who consent to completing the questions.

For example: If 2 out of 10 patients/clients opt out of reading the questions → participation rate is 80%



Capturing participation rate

- Plan for periodic data audits: Audit Participation Rate once a month in order to create consistent follow-up with staff, develop benchmarks for improvements, and monitor overall progress.
- Develop protocols: Develop a method that allows the system to differentiate between 'patient/client not asked', 'missing data' (no response was provided), and 'prefer not to answer'.
- Set target rate: Start with a target participation rate, which can be used as a benchmark for evaluating improvements in participation and data quality. We recommend a minimum starting participation target of 80%.


4.2. INDICATOR: Item Response Rate

Purpose of 'Item Response Rate'










Item response rates capture the proportion of patients/clients responding with 'do not know', 'prefer not to answer', or filling out the 'Other' field. Ensuring high data quality means minimizing the number of responses for those categories. Reducing the proportion of those responses is important because 'do not know' + 'prefer not to answer' don't provide information on who patients and clients are, while open ended data from the 'other' field can be difficult to analyze and link to outcomes.

Definition of 'Item Response Rate'

Calculated for each question, this number is intended to indicate the total number of responses per category (or per grouped categories).

For example: 10 patients/clients were asked the question on sexual orientation and provided a range of responses (each individual response is represented by a )

What is your sexual orientation? Check ONE only

Bisexual Heterosexual ("straight")   Queer 
 Gay   Lesbian Two-Spirit
 Other (please specify): Asexual  Prefer not to answer  
 Do not know 

One possible summary of response rates:

- 20% response rate for 'Prefer not to answer'
- 10% response rate for 'Do not know'
- 10% response rate for: 'Other'
- 60% response rate for: All other responses ('Heterosexual' + 'Lesbian' + 'Gay')

The higher, the better

For data quality control purposes, response rates are used to examine whether we have usable data that can be analyzed to understand who our patients/clients are, detect inequities, etc. With those particular goals in mind, it's important to reduce rates for 'Do not know', 'Prefer not to answer', and 'Other'.

All responses, including 'Do not know', 'Prefer not to answer' and 'Other' can be a rich source of information. For example, an open-ended response under 'Other' can provide important information for service provision, although this sort of response may be more challenging to include in aggregate data sets. Furthermore, 'do not know' and 'prefer not to answer' can be used to explore patient attitudes toward questions and possible barriers to data collection.

Examining ‘Item Response Rate’

- Set target rate: Use a pre-set target for an item response rate as a benchmark for evaluating data quality. We recommend a minimum target of 85%.
- Examine response rate pattern: When calculating item response rates, you can differentiate between “prefer not to answer”, “do not know”, and blank, as they lead to different conclusions on the reasons underlying lower response rates
- Plan for periodical data audits: Perform monthly Item Response Rate audits and share results with staff, develop benchmarks for improvements, and monitor progress.

BEST PRACTICE

Periodic and consistent monitoring of Participation Rates and Item Response Rates are crucial for data quality because it allows you to:

- Develop benchmarks that will help you set targets for improvement
- Evaluate the demographic data collection process

4.3. INDICATOR: Feedback

We recommend collecting feedback from staff on a number of issues such as engagement, what works (or doesn’t), effectiveness of data collection methods, and the training model².

Focus Groups (Data Collectors): We recommend holding focus groups to get information on data collectors’ experiences with patients/clients and their thoughts on the methods, training, and ideas on improvement. Please refer to *Appendix A* for examples on focus group questions.

Survey (Data Collectors): We recommend supplementing information from focus groups with feedback using a survey that data collectors can fill out anonymously. Please refer to *Appendix B* for a sample feedback survey.

Patient/Client Feedback:

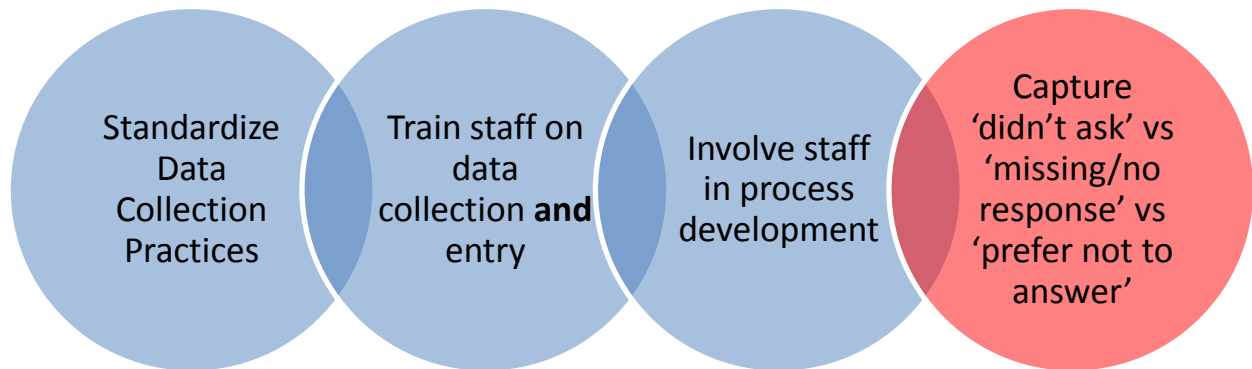
When possible, encourage staff to make a note of patient/client feedback on questions such as:

- Which questions were most difficult to answer?
- What made them difficult to answer?
- Which questions that made you uncomfortable?
- How clear were the questions?
- Were there any questions or concerns about confidentiality?

² All feedback questions and forms used in this document were developed by the Tri-Hospital plus Toronto Public Health Demographic Data Collection Pilot Project

5. Ensuring High Quality Data

5.1. Planning Stages



Differentiating between Missing/Didn't Ask/PNA allows you to troubleshoot issues. For example:

- Allows you to determine why something was entered
- Helps you identify workflow issues (e.g. are only a proportion of patients/clients getting the questions?)
- Points to issues with data entry (e.g. Are all staff aware of codes for missing data?)

(Adapted from: Grasso, C (2016, November). *SOGI Data Integrity & Quality*. National LGBT Health Education Centre)

5.2. Post- data collection

Monitor data

Consistent monitoring can build accountability and provide benchmarks for tracking progress. Effective monitoring include:

- Data quality updates are periodical and consistent
- Data quality reports use a standard format and promote action (e.g. indicates participation rates and provides a performance target of 80%)
- Data collectors are part of ongoing discussions around improving data collection

Review data collection procedures

Improving data quality includes understanding how data collection methods and processes impact data quality along factors such as participation rates, missing data, and item response rates. To get involved in that aspect of data quality improvement, organizations can collect information such as “who provided the information”, “data collection method”, “language”, “location of data collection”, etc. Information on data collection procedures can then be used to understand factors impacting data quality in order to address them. For example, if collecting information from someone other than the patient/client (e.g. their partner) increases rates of ‘*prefer not to answer*’, ‘*do not know*’, or missing data, an organization may restrict who the information is collected from.

Please refer to *Appendix C* for sample questions that can be used to capture information on data collection procedures.

Engage department(s)/program(s)

Our environmental scan indicated that hospitals/CHCs already collecting individual-level demographic try to data share data findings with departmental leaders and staff, particularly around Participation Rates and Item Response Rates. These findings are also used to highlight success or identify a need for improving rates.

BEST PRACTICE

Update Reports outlining demographic data collection findings including participation and item response rates have been used in several hospitals/CHCs to (1) engage staff, (2) identify room for improvement

5.3. Additional Training

A significant number of data issues impacting data quality can be linked back to training and data collection methods. Therefore, we strongly advise organizations to consistently update training (e.g. through feedback) and provide follow-up training sessions where staff can refresh their knowledge and develop solutions to new and ongoing challenges in data collection.

Appendix A: Data Collectors Focus Group Questions

Training

1. What worked and what didn't work to prepare you to collect the patient/client demographic data?
Probes³
 - a. *In what ways did the training prepare you/not prepare you to administer the questions to patients/clients?*
 - b. *In what ways did the training prepare you/not prepare you to administer the questions to families?*
 - c. *In what ways did the training prepare you/not prepare you to administer the questions electronically (e.g. computer), verbally, or on paper?*
2. What are additional topics or areas that can be added to improve training?

Patient/Client Feedback

1. What types of questions and comments did you receive from patients/clients about the items and when?
2. What are questions patients/clients have difficulty answering?
 - a. *What do you think contributed to that difficulty?*
3. What questions were patients/clients most reluctant to answer?
 - a. *What do you think contributed to that reluctance?*

Data Collection Experience

1. What were some unexpected issues that arose?
2. What successful strategies did you employ to help patients/clients be more comfortable with the questions?
3. Are there any other concerns or suggestions?

³ Adapted from Sick Kids work on Focus Group Questions

Appendix B: Data Collectors Survey Questions

1. How well did the training prepare you to administer the questions?

- Not at all well Adequately well Very well Other (please specify): _____

How could we improve training?

2. On average, how much time did it take to administer the questions (in minutes)? _____

3. Did patients/clients need clarification of questions (Select all that apply)?

- What language would you feel most comfortable speaking in with your healthcare provider?
- Were you born in Canada?
- Which of the following best describes your racial or ethnic group?
- Do you have any of the following disabilities?
- What is your gender?
- What is your sexual orientation?
- What was your total family income before taxes last year?
- How many people does this income support?

4. Were patients/clients reluctant to answer any question (Select all that apply)?

- What language would you feel most comfortable speaking in with your healthcare provider?
- Were you born in Canada?
- Which of the following best describes your racial or ethnic group?
- Do you have any of the following disabilities?
- What is your gender?
- What is your sexual orientation?
- What was your total family income before taxes last year?
- How many people does this income support?

5. Do you have any other comments about the process of collecting this information?

Appendix C: Questions on Data Collection Procedures

Information provided by:

- Patient/Client
- Patient/client via interpreter (including family member, interpretation services, etc)
- Other person

Method for capturing data:

- Computer
- iPad/Tablet
- Kiosk
- Paper
- Other

Questions were administered by:

- Staff member (Admitting/Pre-Admitting/Registration area)
- Health care provider
- Self (patient/client)
- Other

Language used by patient/client:

- English
- Other

Location of data collection:

- Waiting room
- Bedside
- Emergency Department
- Admitting/Pre-Admitting/Registration area
- Other

